Computational Linguistic Models for Understanding Attitude and Behavior

by

MeiXing Dong

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science and Engineering)
in The University of Michigan
2022

Doctoral Committee:

Professor Rada Mihalcea, Chair
Associate Professor Munmun De Choudhury
Assistant Professor David Jurgens
Assistant Professor Lu Wang

MeiXing Dong

meixingd@umich.edu

ORCID iD: 0000-0003-4332-5289

ACKNOWLEDGEMENTS

It's often said that "it takes a village" to accomplish large goals. The many people I encountered along my Ph.D. journey have undoubtedly influenced me and helped me reach where I am today.

First and foremost, I thank Rada Mihalcea, my research advisor. She has supported my intellectual curiosity throughout the years and adapted to the various iterations of me as I developed as a researcher and as a person. I'm also grateful for the other members of my thesis committee, David Jurgens, Lu Wang, and Munmun De Choudhury, for their insightful feedback and questions.

Throughout my time at the University of Michigan, I've had the privilege of calling many wonderful people as my colleagues and friends. I learned much from my time with the members of the LIT Lab: Carmen, Ian, Steve, Artem, Veronica, Mohamed, Shibu, Charlie, Mahmoud, Aparna, Yiqun, Laura, Paul, Oana, Jonathan, Santiago, Allie, Laura, Ash, Yiting, Harry, Zhijing, Andrew, Do June, Siqi, and many others who visited the lab and the CSE department. Thank you to my collaborators, Sophia, Xueming, and Yiwei; your presence not only helped me bring my ideas to life, but also made the research process less of a solitary pursuit.

Life doesn't stop while the Ph.D. happens, and I'm grateful for my friends, Preeti, Jeeheh, Varshini, and many others, who helped keep me sane and brought joy into my days. I thank my family for being there from the beginning.

I'm also thankful for my violin teacher, Elbert, who not only helped to rebuild my violin playing from scratch, but also taught me how to pursue excellence sustainably,

without beating myself up for every mistake. He has been an invaluable mentor in music and in life.

Thanks a bunch to my partner, Richard, who never ceases to show his support and love for my endeavors, for who I am as a person, and for who I'm working to become.

The work in this dissertation was supported by the Rackham Merit Fellowship Program, the Michigan Institute for Data Science, the National Science Foundation (grant #1815291), and by the John Templeton Foundation (grant #61156).

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Attitudes are often expressed in what people say and write, as well as the content they choose to interact with. With the proliferation of social media and other online content, we are able to understand how people express their attitudes through large-scale linguistic analyses. Further, people's attitudes and behaviors are often intertwined: attitude signals can be predictive of future behaviors, and conversely behavioral patterns can reveal underlying attitudes. This thesis explores the development of computational linguistic models to understand attitudes and behaviors. We surface the attitudes that people hold with respect to social roles (e.g., "professor," "mother") and compare them across different cultures using corpus-statistics models and dependency-based embedding models. Next, we look at how personal traits are predictive of behavior. To this end, we explore how we can incorporate implicit world knowledge into language models by predicting attitudes towards charitable giving. In this same direction, we examine traits, as expressed on social media, that are indicative of people likely to persist in pursuing self-improvement. We leverage linguistic characteristics such as expressed affect, writing style, and latent topics. Finally, we gain insight into how attitude and behavior give insight to each other by predicting attitudes towards philanthropic causes based on engagement behavior with newsletters and personal background information, using text-aware graph representation models. We also show how behavioral traits present in online communities are predictive of resilient attitude during the COVID-19 pandemic.

CHAPTER I

Introduction

People experience the world in largely subjective ways. The way that people perceive their surroundings, or their attitudes, is often expressed through language, as can be seen in the proliferation of social media content. People's attitudes can also have implications in how they behave in related contexts. In this thesis, we aim to understand ,*through language*, (1) how people express their attitudes; (2) how people indicate intended behavior; (3) how attitudes or other personal characteristics manifest in behavior; as well as (4) how behavior can reveal underlying attitude.

Through our work, we can expand existing social science theories about attitude and behavior to unseen domains and scales of magnitude. It can also serve as a starting point to develop theories to explain novel phenomena unique to our modern world, with its ever-increasing complexity in quantity and types of human interaction facilitated by technology. Data-driven insights enabled by our work can inform the design of new technology platforms that improve our experience of the world and with each other.

1.1 NLP for Computational Social Science and Social Computing

The recent explosion of digital content affords unprecedented opportunities to not only study human behavior, but to use the insights to enhance people's lives using technological tools. People go to social media platforms to share information, participate in communities, seek support, and to express themselves in countless other ways. We stand to learn how people's attributes, such as personality, attitudes, and values, are tied to behavior in both online and offline settings. It is difficult to manually read and analyze the vast volume of available data from online sources. The use of computational methods, informed by work from social science, has allowed researchers to examine greater magnitudes of information. However, much of this available data is in the form of raw natural language text. Text is often unstructured and complex, necessitating the development of Natural Language Processing (NLP) techniques to be able to fully utilize the richness of the data.

The fields of computational social science embody the intersection of social science and computational science, using computational methods to tackle questions that are centered around people. Such work has wide-reaching influence, ranging from the structure of online platforms [1–3] to government policy considerations [4–6].

While the primary focus of this thesis is on the use of language, other signals of user behavior are also useful to consider. For instance, someone's social connections and online engagement patterns can tell us about their personal traits and behavior [7, 8]; such data is often complementary to linguistic data [9]. Therefore, we also incorporate features such as the types of online communities in which people (Chapter IV) and user interaction patterns within online communities (Chapter VI).

Our work contributes to this growing interdisciplinary field by showing how to leverage and extend NLP methods to gain a deeper understanding of people's atti-

tudes and behaviors, as well as how they connect to one another.

1.2 Attitudes and Behaviors

We can learn much about people from the language they produce. Prior work has explored deriving demographic information, such as age, gender, education level, and political orientation [10–12]. Further, accounting for demographic factors can improve performance in NLP tasks [13–15]. However, this is just the surface of what we can learn; we can gain insight into personality [16], values [17], attitudes [18], political orientation [19], mental health [20], and more. In gaining this more nuanced knowledge, we can model users, understand human behavior, and deliver improved and personalized digital services.

From psychology, a person's attitude is a way of thinking or feeling about an aspect of that person's world, such as another person, a group of people, a physical object, or a behavior, and is typically reflected in that person's behavior.

Beliefs are a basis for attitudes. Through experiences, people form beliefs about an entity by associating the entity with various characteristics, qualities, and attributes. More formally, belief is the subjective probability that an object has a certain attribute [21]. Such beliefs can be greatly influenced by one's surroundings, such as one's geographic location or cultural context.

In our work, we consider attitude as an aggregate of one's underlying beliefs. Specifically, we show that we can automatically extract attitudes towards social roles, composed of the attributes that people associate with social roles. Further, we present evidence that these attitudes reflect their cultural context and conduct a cross-cultural analysis of these attitudes.

There has been prior computational work that detects attitudes from text. For instance, others have used product reviews to identify associated attributes of objects, such as food, movies, or other products [22]; analyzed attitudes towards news topics in

media [23]; and detected subgroups in ideological online discussions based on attitudes [24]. Such work on detecting attitude has broad applications in core NLP tasks such as opinion mining [25, 26], and question answering [27, 28]. Most prior work has focused on attitudes that are explicitly and readily expressed in text, such as "The sushi was great, but pricey," or "His claims are so ignorant." However, many attitudes are implicit and not stated, and are therefore more difficult to detect.

Modeling implicit attitudes in text yields opportunities for building improved NLP models that capture and reveal more nuanced information from text than standard language models. Linguistic models that incorporate characteristics about the authors, the audience, or other related people can see improved performance in numerous tasks such as sentiment analysis [29] and dependency parsing [30].

Further, attitudes and behaviors are intertwined with each other. Attitude can be predictive of behavior since it influences how people react to situations and stimuli [31]. By the same reasoning, behavior is then indicative of underlying attitude. Further, our definition of behavior is not limited to physical actions. What people choose to say and write can also be considered behavior, and therefore expressed linguistic characteristics are implicit ties between behavior and attitude.

To explore how attitudes inform behavior, we use signals of people's attitude towards charitable causes to predict whether their future donation behavior will be directed towards similar causes.

We also explore how historical behavior can be indicative of attitude. We use features derived from the normal aggregate behavior of communities, such as how people engage with each other and what they talk about, to predict how these same communities will cope during the COVID-19 pandemic; these patterns implicitly reflect the communities' attitudes towards their negative life circumstances.

We use the wide availability of behavioral data to **gain insight into attitudes and computationally conduct experiments at a scale difficult to achieve**

through the traditional social sciences. In parallel, **we develop enhanced natural language processing models that better capture implicit human characteristics**. We aim to understand how attitudes are expressed in different contexts and how attitudes manifest in linguistic differences and behavior.

1.3 NLP Techniques

Many techniques from NLP have been leveraged to examine computational social science problems. Lexicons, such as LIWC [32] and the NRC Emotion Lexicon [33] have seen wide use for gaining insight into personality [34–36], sentiment [37], and more. Though lexicons are useful because of their interpretability and ease of use, many were traditionally built using manual annotation [38, 39] which is often expensive both financially and with respect to human effort. To address this, there have been many efforts towards automating the process of building lexicons [40, 41].

Recent word embedding models have become ubiquitous due to their ability to capture latent semantic information. Models such as word2vec [42, 43], GloVe [44], and BERT [45] have exhibited strong performance on a wide range of NLP tasks. Beyond these more generalized methods, researchers are able to adapt or craft linguistic features, such as readability [46, 47] and dependency parse information [48, 49], as desired for the task at hand.

1.4 Research Questions

In this thesis, we address three main research questions centered on computational linguistic models of attitudes, behaviors, and their relation to each other.

RQ 1: How can we computationally model the attitudes that people hold towards entities in their world? We investigate several ways to build linguistic models to extract the implicit attitudes that people hold with respect to

groups of people.

By explicitly extracting the underlying associations from language, we not only quantify the underpinnings of people's attitudes, but build models that yield attitude explanations that are easily understandable by people. Such work provides social scientists with additional computational tools to analyze culture.

In the work detailed in Chapter 2, we show how language can be used to identify and understand the implicit attitudes people hold in regards to social roles across different cultures and societies. Attitudes people hold about the world often manifest themselves in the way we use language. Understanding what people say or write can help us gain insight into their worldview, beliefs, and the way they are primed to interact with the surrounding world.

Such analyses of language can also lead to new insights into cultural differences. Groups of people sharing certain characteristics – e.g., nationality, region, state, gender, or religion – would often have a shared understanding of the world, which in turn is reflected in their use of language.

RQ 2: How can we predict the behaviors that people are likely to exhibit in a given context based on their personal characteristics?

Personal characteristics can be indicative of intended or future behavior. We study the connections between expressed traits and behavior in two lines of research.

In work described in Chapter 3, we predict attitude towards charitable giving. Data-driven learning has made great strides over the past three decades. While many recent learning strategies assume the availability of a large amount of data, there are still many applications that only benefit from limited amounts of data. We can enrich sparse textual content inside categorical datasets, to bring into the learning framework additional information that is implied by the text but not explicitly stated. To this end, we conduct experiments in the context of a donation prediction problem, where we use a dataset consisting of the profiles of university alumni who have previously

donated, as well as alumni who did not make any donations, and attempt to predict whether a previously unseen person is likely to donate or not. We demonstrate how sparse text can be enhanced using external information and use our models to better predict whether someone is likely to be a donor.

The act of donation is not straightforward; many factors are involved, such as a person's willingness to give, interest in the funding target, and level of wealth. Our work shows that implicit information about people may be present in their associated categorical information, such as major, degree, and profession.

In our second line of work, described in Chapter 4, we explore how traits are indicative of behavior and focus on people who are pursuing self-improvement. Many people aim for personal change at different points in their lives. People's levels of perceived self-efficacy, risk perceptions, and outcome expectancies can be predictive of eventual behavior change success in a wide number of contexts. Such attitudes can appear in linguistic patterns, such as expressed affect and writing readability. We seek to understand the characteristics of people who are in the early, motivation phase of behavior change and how this reflects in whether someone maintains persistent interest in self-improvement.

We leverage linguistic characteristics such as expressed affect, writing style, and latent topics to automatically distinguish people who sustained their intent for self-improvement from those who did not continue. These features provide human-understandable rationale for how these people behave; in social science applications, model explainability is often as, if not more important, than pure model performance.

RQ 3: How do attitude and behavior give insight into each other? Knowledge about people's attitudes can have implications in how they behave in contexts related to those attitudes. Towards the goal of understanding the relationship between attitude and behavior, we conduct work in two research directions.

First, in Chapter 5, we model attitude towards philanthropic causes based on en-

gagement with emails and personal background information, and use this to predict donation behavior. We build graph representation models from prior user donations and article clicks, and further enhance these graphs with additional edges derived from textual similarity relations among donations and clicks. This context is promising because we see explicit behavior (donations) following from underlying attitudes towards subjects related to the behavior.

Second, in Chapter 6, we analyze subjective wellbeing in US cities in response to COVID-19, and characterize how community behavior prior to the pandemic is predictive of recovery patterns and resilient attitude during the pandemic.

1.5 Outline

The thesis is organized as follows. In Chapter 2, we address our first main research question by tackling the task of extracting implicit attitudes of social roles from social media.

We shift our focus to predicting behavior in Chapter 3. We detail our work in predicting donation behavior based on personal background and our developments in enhancing sparse text with rich information derived from related corpora. We continue in Chapter 4 with understanding persistent intent to change based on social media data.

In Chapter 5, we begin addressing our third research question. We build upon our work in Chapter 3 and we predict attitude towards potential donation interests, as expressed through actual donations, based on related behavior such as interacting with emails. We extend the alumni donation data with recorded engagement with alumni newsletter emails and model user behavior using text-aware graph representations.

We continue in Chapter 6 by studying the trajectories of subjective wellbeing of cities across the US during the COVID-19 pandemic, and the community attitude characteristics that correlate with resilient behavior.

Finally, we close with a summary of our contributions, conclusions, and future work in Chapter 7.

CHAPTER II

Attitudes Towards Social Roles Across Cultures

2.1 Introduction

In this chapter[1] we present approaches to computationally understand the attitudes that people hold with respect to groups of people in society. Attitudes we hold about the world often manifest themselves in the way we use language. Understanding what people say or write can help us gain insight into their worldview, beliefs, and the way they are primed to interact with the surrounding world. Such analyses of language can also lead to new insights into cultural differences. Groups of people sharing certain characteristics – e.g., nationality, region, state, gender, or religion – would often have a shared understanding of the world, which in turn is reflected in their use of language.

While the connections between language and culture have traditionally been the purview of cultural psychology [50], more recent work in computational linguistics has also started to address these connections, resulting in models that can uncover the different use of words across cultures [51, 52], the various distribution of topics in different cultures [53], or the word associations that people with different demographics tend to make [54, 55].

[1]The work in this chapter benefited from input from Carmen Banea, David Jurgens, and Rada Mihalcea. This work was published in the Proceedings of the 2019 International Conference on Social Informatics.

The hypothesis driving our work is that we can use language to identify and understand the implicit perceptions and expectations that people hold with regards to social roles in our society. For instance, the frequent use of the descriptor *kind* or the action *help* in connection to the role *friend* can be an indication that friends are usually regarded as people who are kind and provide help. Moreover, we also hypothesize that there may be cultural differences in these social role perceptions, and that different groups of people may correspondingly use different descriptors or actions when they refer to the same social role.

This chapter makes four main contributions. First, we examine what constitutes a social role, and we propose the use of descriptors (adjectives) and actions (verbs) as a way to understand the implicit perception of social roles as reflected in language. Second, we introduce a new data set, consisting of 49 frequent social roles (e.g., *mother*, *friend*, *lawyer*) and the associated descriptors and actions, as contributed by over 400 human judges from two different cultures (United States and India). Third, we perform several analyses to uncover cross-cultural variations in social role perception, and we identify roles with high, medium, and low variations. Finally, we propose two computational models that can predict the most likely social role based on a descriptor or an action. One model is based on statistics collected over a large syntactically annotated collection of texts authored by people from two cultures, while the second one relies on neural models that are aware of the syntactic relations between words.

Our main findings show that there are indeed differences in the perceptions associated with the roles between the two cultures, and that the degree of cultural similarity varies across the roles. The computational models show that it is possible to predict roles from the attributes that people associate with them. Furthermore, our models exhibit higher performance when the train and test set cultures match, indicating that our models encode cultural differences.

2.2 Related Work

The concept of "roles" is frequently considered by those in the social sciences as a way to analyze social structures and behaviors [56–60]. Roles can be characterized by the norms and expectations that society places on people of particular social or functional positions [59]. Such norms greatly influence how people act and interact with others [61], especially when one is acting as a member of a role [62]. The perceptions of others are important; depending on whether one acts according to role expectations, there exist rewards or punishments doled out by society [58]. By asking members of a group about the behaviors that a role is likely to participate in, one can analyze the differences in perceptions of roles between cultural groups, such as Hispanics versus the general US population [63].

We take inspiration from previous work that models latent character types, or personas (such as the "love interest" or "best friend") and their typical characteristics in films [64]. To extract character aspects, the authors look at a subset of the syntactic dependencies that involve the personas. We extract aspects in a similar way and focus on predicting a role based on its expected characteristics, in contrast to Bamman et al. that focus on partitioning types of roles. Additionally, films tend to create stereotypical personas with strong associations to their characteristics. Social roles, however, are constructed from societal expectations in aggregate and can be much more nuanced.

Another related line of research has considered the prediction of words that are most likely to be associated with a stimulus word [54]. Our task differs in that we go beyond free-form associations and instead hone in on specific aspect types, namely actions and descriptors as they relate to a given social role.

To use natural language, we must build word representations. A straightforward approach is to treat words as discrete symbols, leading to many bag-of-words methods for representing text [65, 66]. While useful for many tasks, this representation does

not encode relations between words or semantics. Many recent word representation methods model words as continuous, dense vectors derived from neural networks [67, 68] or word co-occurrence information [69], also known as word embeddings. These have been shown to perform well across numerous tasks [70]. Additionally, [54, 71–74] have sought to encode additional sources of information to be captured in word embedding vectors.

One of our models is derived from dependency-based embedding models [75], where dependency links are used to form the contexts in a skip-gram model. The resulting embeddings encode functional similarity rather than topical similarity. For instance, *rapping*, *busking*, and *breakdancing* are among the most similar words for "dancing" when using dependency-based embeddings, as opposed to topically related words surfaced by regular linear context embeddings, such as "dancer", "dance", and "dances." We adapt the former model to focus on specific types of dependencies that encode aspects, distinguishing between the different functional uses of a word. For example, we can find roles that are most relevant to a given aspect, rather than the words that are generally related either by domain or by function.

2.3 Collecting a Cross-Cultural Data Set of Social Roles

The perception of a social role can be characterized by the descriptors or actions that people associate with it. We created a data set by surveying a large and demographically diverse audience on Amazon Mechanical Turk (AMT) about the aspects they associate with different roles. Our survey task is similar to that of gathering word associations, where survey participants are provided with a list of *stimulus words* and are asked to provide the first word that comes to their mind [54, 76, 77]. However, rather than asking for free-form associations, as done before, we added structure to our prompts to induce responses that correspond to descriptive aspects. Specifically, we asked survey participants to provide actions and descriptors for each stimulus role,

given prompts such as *What is a friend like?* and *What does a friend do?*

Selecting Social Roles. Language abounds with the names of the many social roles that people partake in, from common names (like mother or teacher) to less common ones (like debtor or occultist). Here, we aim to curate a set of social roles for annotation that meet three criteria: (1) occur with high frequency in text, (2) appear in daily life, and (3) have relatively unambiguous words associated with them. We detail the selection process next.

A large set of candidate social roles were selected using WordNet [78], a large lexical database for English. WordNet provides an ontological organization of a word's meanings and contains a semantic network of how these meanings (i.e., *senses*) relate to one another. In particular, WordNet specifies the hyponymy relationship between senses that allows us to identify more specific meanings of people; for example, *mother* and *father* are both hyponyms of *parent*. To get all potential social roles, we collected the 8,654 words that are children of *person* in the hyponymy tree.

As WordNet contains many infrequent words, we extracted frequency counts for each role from a large collection of blog data from India and the US, described in detail in §2.5.1. We tagged each blog sentence with part-of-speech information, and then counted the frequency of each candidate role occurring as a noun.

Finally, we analyzed the most frequently occurring candidate roles and identified roles that occurred in blogs from both countries, that are generally unambiguous, and are likely to be encountered in day-to-day life. For instance, we did not include *queen* because most people are unlikely to interact with queens, and therefore descriptors and actions are unlikely to reflect personal experiences. We also excluded ambiguous roles such as *official* or *director*, since their attributes can change depending on the context. Ultimately, the selection process resulted in a set of 49 social roles.

Crowdsourcing Setup. The descriptors and actions for each social role were col-

lected through AMT English surveys[2], targeted to individuals in India and the US. We chose countries that were likely to differ in terms of cultural and societal norms, but still have many English speakers to bypass translation issues. Each participant was presented with five social roles and asked to provide three actions and three descriptors for each role. Participants were also asked to indicate how often they interact with the role and how positively they view those interactions. A demographic questionnaire was included at the end of the each survey containing questions about the respondent's gender, age, level of education, ethnicity, and nationality. Responses were collected from 200 participants from each country for each role. This resulted in 600 actions and 600 descriptors collected for each social role, for each country.

To ensure answer quality, we included a spam-check question that asked for the answer to an earlier question. This filtered out participants that responded without reading the prompts. Built-in form restrictions prevented the submission of answers that were given as examples, or empty answers. As a final check, we manually spot-checked responses before accepting them, to make sure participants did not fill in random words. We lemmatized all of the responses and for each given social role we kept those responses that occurred five times or more as culturally-salient aspects of the role.

Previous studies [79, 80] have shown that while Turkers tend to be younger and more educated, it is possible for the data they supply to reflect aspects of the population at large, such as ideology. The data we gathered serves as an additional resource to complement existing cross-cultural resources, providing insight into cultural differences pertaining to how social roles are perceived. Despite the potential skew in demographics, we still find differences between the two countries, as detailed in later sections.

Table 2.1 shows the top responses for a sample set of social roles.

[2]English is one of the official languages of India and the second most-spoken language behind Hindi.

Table 2.1: Top survey responses for societal role words.

	Actions		Descriptors	
Role Word	US	India	US	India
mother	care, love, cook	care, love, cook	loving, caring, nurturing	caring, lovable, loving
baby	cry, sleep, eat	cry, play, smile	loving, sweet, kind	cute, innocent, chubby
doctor	diagnose, prescribe, examine	treat, care, cure	smart, intelligent, helpful	caring, god, helpful
policeman	protect, arrest, serve	arrest, protect, help	strong, brave, helpful	strict, brave, strong
student	study, learn, read, write, work	study, play, learn, read, write	studious, smart, young	obedient, intelligent, studious
politician	lie, campaign, speak, talk, cheat	speak, vote, lead, promise, rule	dishonest, greedy, corrupt	powerful, honest, influential

2.4 Demographic Variations in Social Roles

The characteristics associated with social roles in different countries can reveal cultural similarities and differences. Many aspects are associated with a role regardless of the underlying culture, such as a *mother* being *caring* and a *policeman* being *brave*. On the other hand, *doctors* are more associated with preliminary actions in the treatment process such as *examine*, *diagnose* and *prescribe* in the US, while in India they are more associated with treatment results, such as *treat* and *cure*. Also, Indian descriptors show a stronger perception of doctors as being *caring*, versus *smart* and *intelligent* in the US. Additionally, US participants associate many negative aspects with *politician*, reflecting the current political climate, while in India, the actions are mostly associated with positive aspects.

Intra-group and Inter-group Similarities. We measure the agreement between respondents within and across cultural groups. Given the set of response words for a social role from a single held-out respondent, we determine whether any of these responses match the most frequent response or any of the top 25 responses of the

Table 2.2: Left: intra-group similarities (higher similarity indicates a more cohesive group). Right: inter-group similarities (higher similarity indicates a less distinct group).

	Intra-group similarity			Inter-group similarity		
	Demographic	Primary	Top 25	Demographic	Primary	Top 25
Descriptors	US-US	0.33	0.89	US-IN	0.19	0.78
	IN-IN	0.24	0.76	IN-US	0.15	0.61
Actions	US-US	0.40	0.93	US-IN	0.35	0.90
	IN-IN	0.40	0.89	IN-US	0.32	0.85

remaining respondents in the group. If so, then we consider this respondent in agreement with the group. We define the agreement score as the ratio of participants whose responses are in agreement with the group. Similarly, we measure the agreement between each survey respondent in one group with the most frequent or top 25 most frequent responses from the other group.

The intra-group and inter-group analyses are shown in Table 2.2. From the intra-group similarities, we can see that there is high agreement among both the top and top 25 responses given by participants from the same country, with the US having higher agreement in general than India. Overall, action responses are more cohesive across the two countries compared to descriptor answers; we noted earlier that there is more variation and subjectivity in regards to descriptors.

When we look at how much participants from one country agree with participants from *the other country*, we find a much lower agreement for descriptors, both in terms of primary response and the top 25 responses. For example, the similarity drops by 0.08 (from 0.40 to 0.32) between India-India and India-US for the most frequent response for actions. We see the agreement drop in all cases when comparing intra- versus inter-group similarity. We conclude that the agreement for actions between the countries is comparable to the agreement within countries, implying that the actions attributed to roles are more objective and universal.

Levels of Social Role Similarity Across Cultures. We closely examine how

various roles are perceived differently across countries. To measure how similar a role is between India and the US, we compute the cosine similarity between the frequencies pertaining to the set of aspects resulting from the union of the responses for that role for each country. Table 2.3 shows a sample of roles that display various levels of similarity ranging from high to low in regards to their associated actions or descriptors across the countries in question. We notice that *soldier* exhibits the highest similarity level both for actions (*fight, protect*) and descriptors (*brave, strong*). *Actor*, on the other hand, showcases a medium action-based similarity, as actors in India regularly engage in dancing, unlike their US counterparts. Interestingly, *friend*, despite its ubiquitousness as a social role, displays among the lowest scoring action-based similarity. We note that in the US, *friend* is more associated with communication-focused actions such as *listen*, *talk*, *laugh*, while in India, given the more collectivist culture, people primarily think of friends in the context of being *helpful* and *caring*.

Levels of Aspect Similarity Across Cultures. We further analyze the frequency of aspect usage across roles to identify how predictive a given aspect is of a social role. Table 2.4 aggregates the responses at the aspect level. We see that some actions are highly predictive of a role. For instance, *arrest* occurs with *police* and *vote* appears with *citizen, politician* roles in both countries. However, *sacrifice* occurs in the action-focused answers for *soldier* in the US, while in India, *mother* and *father* also trigger this response. *Counsel* also displays a divergent usage, in the US being associated with *lawyer*, while in India, with *priest*. Similarly, descriptors also show variations in their associations with roles. These range from a high similarity of 1 for *religious* (which always appears in the context of *priest*), to mid-range (0.46) for *obedient* (which in the US caries a stronger meaning of loyal, and applies to a hierarchical organization, e.g. army for *soldier* or country for *citizen*, while in India it is more indicative of filial piety and the need to listen to one's elders, whether as

Table 2.3: Social roles exhibiting different levels of similarity (H(igh), M(edium), L(ow)) between the US and India based on the differences between the top 20 responses.

Similarity	Role	Score	Top US aspects	Top IN aspects
			Actions	
H	soldier	0.94	fight, protect, defend	fight, protect, shoot
	professor	0.91	teach, grade, lecture	teach, guide, educate
	mother	0.89	care, love, cook	care, love, cook
M	girlfriend	0.77	love, kiss, listen	love, care, help
	policeman	0.77	protect, arrest, serve	arrest, protect, help
	actor	0.71	act, perform, pretend	act, dance, perform
L	doctor	0.66	diagnose, prescribe, examine	treat, care, cure
	politician	0.59	lie, campaign, speak	speak, vote, lead
	friend	0.58	listen, talk, laugh	help, care, play
			Descriptors	
H	soldier	0.89	brave, strong, loyal	brave, strong, patriotic
	writer	0.88	creative, imaginative, smart	creative, imaginative, good
	mother	0.72	caring, loving, nurturing	caring, lovable, kind
M	researcher	0.62	smart, intelligent, curious	intelligent, knowledgeable, brilliant
	prisoner	0.60	angry, sad, guilty	bad, criminal, guilty
	friend	0.59	fun, loyal, funny	helpful, caring, honest
L	farmer	0.44	strong, hardworking, diligent	hardworking, poor, helpless
	judge	0.38	fair, powerful, smart	honest, intelligent, knowledgeable
	politician	0.17	dishonest, greedy, corrupt	powerful, honest, influential

a *student*, *son*, or *daughter*), to low (0) for *committed* (which in the US occurs in prompts for *wife* and *husband*, while in India it appears in prompts for *farmer*).

Sentiment and Emotion in Social Role Perceptions. Social roles can evoke a variety of emotional responses, such as feelings of authority, love, or even fear. Viewed in aggregate, the aspects used to describe roles can potentially reveal which emotional aspects of social roles are most important to a culture. Therefore, we

Table 2.4: Aspects exhibiting different levels of similarity (H(igh), M(edium), L(ow)) between the US and India based on the differences between the top 20 responses.

Similarity	Aspect	Score	Top US roles	Top IN roles
			Actions	
H	arrest	1.0	police	police
	vote	0.99	citizen, politician	citizen, politician
	kiss	0.93	girlfriend, boyfriend, mother	girlfriend, boyfriend, husband
M	medicate	0.71	nurse	doctor, nurse
	sacrifice	0.61	soldier	father, mother, soldier
	forgive	0.51	priest	friend, priest, mother
L	invent	0.15	engineer, chef, writer	scientist, researcher, engineer
	meditate	0.0	—	priest
	counsel	0.0	lawyer	priest
			Descriptors	
H	religious	1.0	priest	priest
	loving	0.96	mother, husband, wife	mother, husband, sister
	curious	0.91	tourist, researcher, journalist	journalist, tourist, researcher
M	wise	0.71	father, priest, professor	professor, teacher, judge
	loyal	0.65	friend, husband, wife	friend, citizen, soldier
	obedient	0.46	son, soldier, citizen	student, son, daughter
L	faithful	0.15	wife, husband	priest, secretary, chef
	jovial	0.0	—	politician
	committed	0.0	wife, husband	farmer

perform two analyses where we convert the actions and descriptions for each role into their sentiment and emotion associations. For sentiment, we map each word to its mean score in SentiWordNet [81] and then average across all the words for each aspect of a role for its estimated sentiment score. For emotion, we repeat a similar process with the NRC Emotion Lexicon [82]. This lexicon maps individual words to a binary indicator of whether they have an association with each of the eight Plutchik emotions [83]. Here, we compute the probability that an aspect word for a role has an association with each emotion. We then average the sentiment and

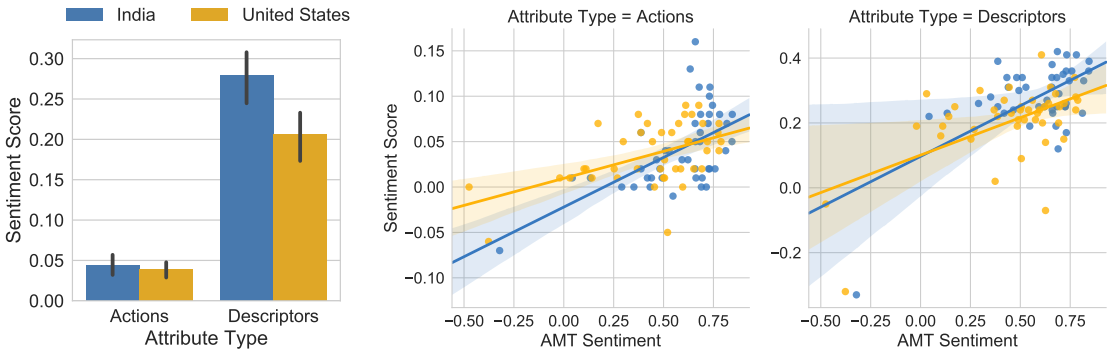

Figure 2.1: India and the US differ significantly in the sentiments of roles attributes (left); indeed, AMT workers' explicit sentiment ratings for each role were highly correlated with inferred sentiments of their descriptors and actions (right). Bars and shaded regions show 95% confidence intervals.

emotion-association probabilities across all roles.

For sentiment, Figure 2.1 shows clear differences between India and the US responses, with AMT workers from India using significantly more positive descriptors about roles. No significant difference is seen for actions, though this is expected, as adjectives (descriptors) typically carry more sentiment than verbs (actions); for example, common sentiment lexicons like SentiWordNet [81] and OpinionFinder [84] contain more adjectives than verbs, and adjectives have been shown to outperform verbs as features in sentiment classification [85]. Examining AMT workers' explicit sentiment ratings for roles, we see that their ratings have high correlation with the inferred sentiment, with Pearson's r ranging from 0.51 to 0.61. This result suggests that the inferred ratings are capturing representative attitudes but, crucially, that roles' aspect words convey more than the workers' sentiment about the role.

The emotion trends, shown in Figure 2.2, reveal a more complex picture with Indian respondents being more likely to use emotionally-associated language than their US counterparts. This heightened emotionality occurs both for positive emotions like trust, surprise, and anticipation, as well as negative emotions like disgust and sadness. However, US respondents are more likely to evoke anger or fear; yet, these emotions are the two least-frequently used in our data. While cross-cultural studies

of emotion have shown differences between India and the US [86–88], these studies have typically looked at specific settings such as childhood development, rather than general attitudes; our data set provides a valuable new source of comparison.

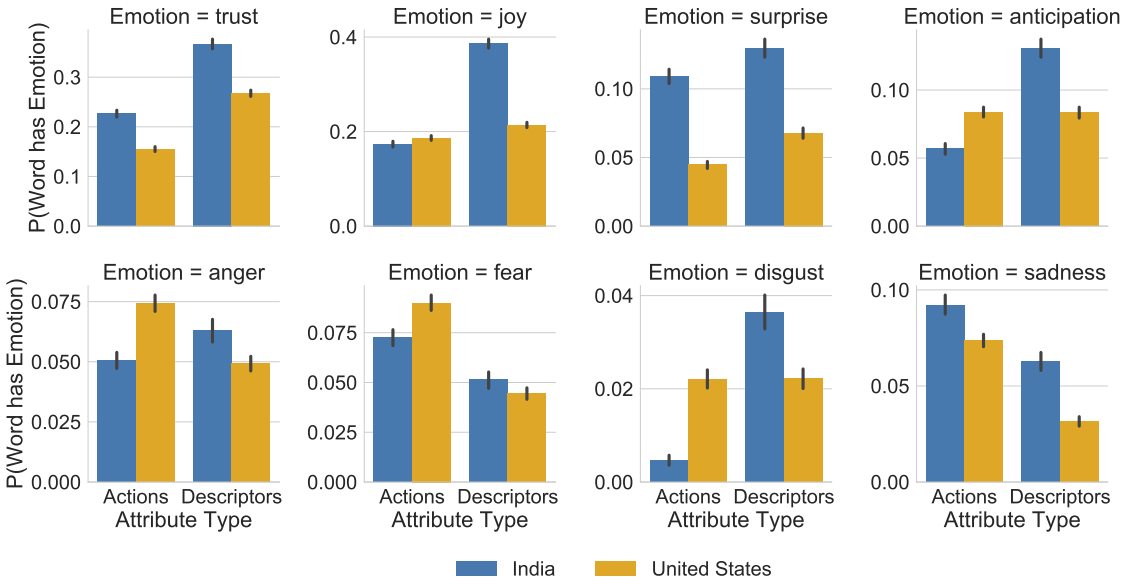

Figure 2.2: AMT answers' emotions. The emotionality of actions and descriptors across social roles shows clear cultural differences in how each is conceived; plots show the probability of an action or descriptor using a word associated with each emotion in the NRC Emotion Lexicon, with bars showing 95% confidence intervals of mean probability.

2.5 Pilot Evaluation

We conduct two initial experiments to gauge how demographic-aware roles can be predicted from descriptors and actions using textual data. We evaluate our models on the task of predicting the most likely social roles for a given aspect. For example, if we think of *lovely,* we want to identify the roles in each culture that are most strongly associated with this descriptor. This enables us to underscore the particularities of each culture, where some roles are associated with softer traits, while others with stronger traits.

We take the top 20 aspects for each role from the survey responses of both countries and combine them into a set of descriptors and a set of actions. Table 2.5 shows

statistics pertaining to these aspects. For each aspect, the set of expected roles (i.e., ground truth) are the ones for which the aspect appears in the top 20 responses. Evaluations are conducted on each aspect type separately. Our models rank the roles for each aspect, which we compare with the expected roles. We report the precision and recall at 5, averaged over the aspects. We also report Pearson correlation (which is typical in word similarity tasks [89, 90]), as this gauges how accurate the model is in arranging the roles in order of association with the given aspect.

2.5.1 Blogger Data Set

To train our models, we need to employ text written by authors whose demographic location is known. For that, we use a set of blogs written between 1999 and 2016 collected from Google Blogger [51]. We select those blogs that also contain location information and only consider those with authors in India or the US. This allows us to analyze the cultural differences between India and the US that may appear as differences in the meaning and usage of roles. We filter out sentences with more than 150 words[3] or with more than 25% non-English words[4]. The remaining sentences are cleaned from HTML tags and truncated to the first 50 words. We then use the Stanford CoreNLP library [91] to obtain dependency parses for the sentences in our data set, as well as part-of-speech tags and lemmatized versions of the tokens. Only lemmas that occur 5 or more times are considered.

Because the US blog data is roughly twice the size of the Indian blog data, we balance the data by downsampling the US data to match the number of sentences in the Indian data. Table 2.6 provides statistics of the resulting data set.

[3]Normal sentences are rarely this long, and upon manual inspection we found that these tend to be malformed sentences.

[4]`https://github.com/rfk/pyenchant`

Table 2.5: Statistics for unique aspect words given by survey responses.

Type	US-only	India-only	Both	**All**
Action	126	76	199	**401**
Descriptor	154	136	156	**446**

Table 2.6: Blog data statistics.

	# Sentences	# Tokens
US	17,476,527	348,479,631
US (balanced)	7,394,484	146,347,629
India	7,426,583	148,710,411

2.5.2 Computational Models to Predict Social Roles

We propose two computational models to predict social roles. The first model focuses on corpus statistics using dependency link counts, while the second uses neural-network dependency-based word embeddings.

Dependency Link Count (DLC) We first look for the actions and descriptors that engage in a syntactic relation with a role in a sentence as a way of modeling the way people associate roles and their aspects. In order to extract role-aspect relations, we leverage dependency parsing information.

Let us consider the following example: *"The attentive policeman arrested the perpetrators."* The dependency parse results in the following relations (the relations in which *policeman* appears in are italicized): *det* (policeman-3, The-1), *amod* (policeman-3, attentive-2), *nsubj* (arrested-4, policeman-3), root (ROOT-0, arrested-4), det (perpetrators-6, the-5), dobj (arrested-4, perpetrators-6).

The first three relations showcase scenarios where words appear with our target role *policeman*. Since we are interested in finding descriptors for the target role, we focus on the *AMOD* relationship (or adjectival modifier) where the role appears in the source position in the dependency relation; to identify associated actions for the target role, we utilize the *NSUBJ* relations (or nominal subject) where the role appears in the target position in the dependency relation. Consequently, *attentive* is marked as

participating in an AMOD relation, while arrested participates in a NSUBJ relation, corresponding to descriptors and actions, respectively.

Word co-occurrence is used extensively to model relationships between words [92–94]. Therefore, we rank the roles for an aspect according to how frequently they co-occur in the link type corresponding to the aspect type.

Dependency Aspect Embedding (DAE) Neural word embeddings have proven useful for a large variety of tasks [95–97]. Here, we make use of their representation power, but aim to capture demographic-focused embeddings for social role aspects in particular. Previous work has shown that dependency-based word embeddings induce different word similarities [75], yielding more functional similarities, rather than topical similarities.

Rather than training our embedding models on all dependency links, we consider only the links that correspond to descriptors and actions. We train separate models for the two different link types for each of the two countries. This yields four models for: actions in the US, descriptors in the US, actions in India, and descriptors in India. We use the Python *Gensim* Word2Vec library and use 300 latent dimensions with negative sampling.

For a given aspect (a descriptor or action), we compute the cosine similarity between the embedding pertaining to the aspect and the embedding pertaining to each of the social roles. The roles are then ranked according to their cosine similarity, where higher values imply a greater likelihood that the aspect is associated with the role.

2.5.3 Evaluations & Discussions

Our experiments analyze models that are widely assumed to capture social information [98] and we test the degree to which they are able to do so on a data set designed with this information in mind. The results for the role prediction task are

provided in Table 2.7[5]. The columns represent the source countries of the survey responses, used as the gold standard data for evaluation, while the rows indicate the country of the blogs on which the models were trained. Bold values represent the best performance, when comparing between model countries, for a given combination of model type and gold standard evaluation.

The DAE model is able to achieve a higher recall for actions than the DLC model, but otherwise the two perform comparably. Notably, these two models achieve equal or better performance when the country of the gold standard responses matches the one on which the model was trained. This implies that these models are picking up the distinctive cultural features of the countries.

The gap between identical models trained on different countries is more pronounced when evaluating on the gold standard US responses than on the Indian responses. As English is not the primary language used by Indians, online users may implicitly be conforming to Western societal norms.

We also noticed that implicit or common sense aspect assumptions, while appearing in primary positions in AMT responses, were less likely to appear in the blog data, and sometime did not occur at all. For instance, *faithful* is a top AMT descriptor for *wife* and *husband*, but occurs very infrequently in the blog text. We also see this for *educated* with *professor* and *creative* with *musician*. Blog data often contained aspects that were actually antonyms of the actions and descriptors provided as answers by the respondents. For example, *corrupt* is among the most frequent descriptors linked to *policeman*, as is *estranged* with *wife* and *unwed* with *mother*. This shows that commonsense knowledge is often not expressed in text, as humans tend not to state the obvious. Consequently, in the blog genre, one tends to express anomalous behavior as it pertains to roles.

[5]Results for word association tasks are traditionally low, and our results are within the same range as previous word association research [54].

Table 2.7: Role prediction results for *actions* (above) and *descriptors* (below). Metrics: precision at 5, recall at 5 and Pearson correlation.

			US			India		
	Model		P@5	R@5	Corr.	P@5	R@5	Corr.
Actions	DLC	US	**0.13**	**0.26**	**0.15**	**0.10**	**0.19**	**0.11**
		India	0.12	0.24	0.14	**0.10**	0.18	**0.11**
	DAE	US	**0.12**	**0.28**	**0.14**	**0.11**	0.24	0.12
		India	0.11	0.25	0.12	**0.11**	**0.25**	**0.13**
Descriptors	DLC	US	**0.12**	**0.20**	**0.13**	0.09	0.14	0.07
		India	0.11	**0.20**	0.11	**0.10**	**0.16**	**0.09**
	DAE	US	**0.09**	**0.19**	**0.09**	0.08	**0.18**	0.08
		India	**0.09**	0.17	**0.09**	**0.09**	**0.18**	**0.09**

2.6 Conclusion

In this chapter, we introduced a new data set of social roles and the associations they trigger in terms of actions and descriptors in two cultures (US and India). We showed that there are differences in the perceptions associated with the roles, with actions showcasing less variability and descriptors exhibiting a wider variation. Furthermore we analyzed the way roles are associated with various sentiment and emotional dimensions. We further used the data set we collected to conduct pilot evaluations focused on predicting social roles. Both our corpus-statistics and embedding dependency-based models show a stronger predictive ability when the train and test set culture match, indicating that there are indeed cultural differences that can be automatically accounted for in our models. The dataset introduced in this chapter is publicly available at `http://lit.eecs.umich.edu/downloads.html`

We have shown that it is possible to extract and predict people's attitudes towards social roles using data that is *not* focused on social roles. This implies that people's attitudes are exhibited through their language even in seemingly unrelated contexts. Further, straightforward applications of natural language processing techniques were unable to yield the desired information, showing that there is room for improvement in computationally modeling people's implicit characteristics, such as their attitudes.

CHAPTER III

Predicting Donation Behavior By Extending Sparse Text

3.1 Introduction

In this chapter[1], we further explore how we can infuse computational language models with implicit information from the world at large. We shift our focus to predicting behaviors and address the problem of expanding sparse textual content to increase the accuracy of data-driven prediction tasks. We evaluate the use of word embeddings and lexicons within the context of a donation behavior prediction task, where we classify potential donors as either likely or unlikely to donate. We perform several comparative experiments and analyses, and show that our methods to automatically enhance sparse textual data significantly improve the predictive performance on this task.

Over the past three decades, data-driven learning has made great strides and brought significant progress across many disciplines, ranging from computer science and information sciences, to psychology, astronomy, economics, and many other science or humanities fields. While many of the most recent learning strategies assume the availability of a large amount of data, there are still many applications that only

[1]This work was published in the Computer Speech and Language journal in 2020 and done with guidance from Rada Mihalcea and Dragomir Radev.

benefit from limited amounts of data. Among these, we often deal with datasets that include only small amounts of textual information that, because of their size and limited vocabulary, end up not contributing as much as they could to the overall learning process.

In this chapter, we explore the question of whether we can enrich sparse textual content inside categorical datasets, to bring into the learning framework additional information that is implied by the text but not explicitly stated. As an example, consider a dataset that includes a text field whose value for one of the instances is the word "computer." Typically, such categorical features are used "as is" and are weighted and used alongside other features, depending on the learning framework. However, aside from being a string of characters, the word "computer" implies "an electronic device for storing and processing data," has associations with other words such as "data," "hardware," "software," and so forth. In this chapter, we present several methods for automatically enriching categorical fields in a dataset where the categorical elements can also be treated as text. Our goal is to improve data-driven predictions, so we perform comparative evaluations that allow us to learn what text expansion techniques work best.

Specifically, we primarily ask our questions in the context of a donation prediction problem, where we use a dataset consisting of the profiles of university alumni who have previously donated, as well as alumni who did not make any donations, and attempt to predict for a new instance whether they are likely to donate or not. We also consider the task of gender prediction on a dataset of blog profiles to determine to what extent our methods can be applied to other datasets.

The amount of textual data available in both datasets is limited in terms of both quantity and variety; each piece of text is a few words at most, and the category definitions restrict the vocabulary. Yet, it can still be quite useful. For instance, a "CEO" is more likely to donate than a "clerk", or a "senior" employee is more likely

to donate than a "recent graduate."

We explore four different strategies for extending sparse text, including two lexicon generation methods, and two embedding methods that are influenced by domain knowledge. Using features obtained from these methods, we build models that predict whether someone is likely to donate, and compare their performance with baseline models that do not make use of such additional features.

The chapter makes two main research contributions. First, we address the question of whether we can effectively augment text fields in a dataset by leveraging information specific to the target domain, and show that with such textual expansion strategies we can significantly improve over a baseline that does not make use of this additional information. Second, we compare several different models for extending sparse text in datasets, including methods that rely on information drawn from (a) the database itself; or (b) external resources, and gain new insights into what methods lead to the highest performance improvements. We seek to answer these questions using the donation prediction task, where we rely on a dataset that has information on previous donors including limited free-form text, and show the role played by different text expansion strategies to improve the effectiveness of our predictive model. We also show that these methods can apply to other cases by evaluating on a second task and dataset.

3.2 Related Work

Our task is related to the classification of short texts, which is challenging because the text is typically sparse and do not provide much word co-occurrence information. In contrast to standard free-form short-text datasets, such as tweets from Twitter, our categorical text is not only short but also restricted in content. For instance, the set of academic majors available at a particular university only contains text from the names of the majors.

Unfortunately, the bulk of recent machine learning methods assume the availability of large amounts of varied data, but there exist many ways of tackling machine learning without this. Hand-built lexical resources have been used extensively in natural language processing tasks like word-sense disambiguation ([99]), sentiment analysis ([100]), and short text classification ([101]). Text embedding methods allow models trained on one domain to be adapted to new domains that have little data.

We focus on lexical resources and embedding methods as they are two of the most straightforward and commonly used methods for text classification tasks. In this section, we overview the work that has been previously done on these related directions.

3.2.1 Lexical Resources

Lexicons have been used extensively in sentiment analysis tasks ([102]). There are many manually created sentiment lexicons such as the NRC Emotion Lexicon ([103]), MPQA Lexicon ([104]), and Bing Liu Lexicon ([105]). General lexical resources have been adapted to the domain of sentiment analysis as well. For instance, SentiWordNet ([106]) extends WordNet ([107]) such that each group of synonyms in WordNet, a manually-created lexical database, is tagged with three sentiment scores: positivity, negativity, objectivity. These lexical resources are very useful but manual efforts to create them are costly and time-consuming ([103]), requiring experts or crowdsourced annotators. This has inspired great interest in automatically inducing sentiment lexicons.

Much work has been focused on Twitter, a microblogging website with hundreds of millions of users from around the world. User-generated text is always short, as tweets are limited to 280 characters. Mohammed et al. ([100]) construct a sentiment lexicon for Twitter based on calculating how closely a word is associated with positive or negative sentiment. A word's association score is calculated using the pointwise

mutual information (PMI) between the word and a seed set of hashtags, such as *#good* and *#bad*.

Many other lexicon induction methods use label propagation to build sentiment lexicons from a seed set of words ([108, 109]). Typically, a lexical graph is built, where each word or phrase is a node and edges represent the similarity between two nodes. Then, propagation methods are used to determine the sentiment of each node, given the sentiment of an initial set of nodes.

Most of these lexicons are built for large, general domains like Twitter. However, the sentiment of a word depends on the specific domain in which it is used. Recent work builds domain-specific sentiment lexicons using label propagation methods and domain-specific corpora ([110]).

Lexicons are also used for many tasks outside of sentiment analysis. For instance, LIWC, a general lexicon, is used to quantitatively analyze content in tasks ranging from personality prediction ([111, 112]) to deception detection ([113]).

3.2.2 Text Representations

There are numerous ways of representing text for computational processing, most of which transform text into a numerical vector. These vectors ideally embed important characteristics of the text, such as the semantics.

Classical representations of text include bag-of-words (BOW), where a body of text is represented as the set of words that compose it, and latent semantic analysis (LSA) ([94]), where the representation is derived from the factorization of a term-document occurrence matrix.

Recent text embedding methods such as Word2Vec ([114]) and GloVe ([69]) are able to capture semantic relationships such as "man is to woman as brother is to sister." A particular type of the Word2Vec model, skip-gram with negative sampling, has been shown to be implicitly factorizing a word-context matrix ([115, 116]). There

have been many extensions of these methods that embed larger bodies of text such as sentences, paragraphs and entire documents ([114, 117]). A downside of neural embedding models like Word2Vec is the prerequisite of large amounts of training data. For instance, the pre-trained Word2Vec vectors released by Google were trained on part of the Google News dataset, containing about 100 billion words.

Representations for sets of words such as phrases and sentences can be constructed by linearly averaging the embeddings of the constituent words. This has remained a strong feature or baseline across many tasks ([118–121]).

3.3 Predicting Alumni Donations

We conduct our exploration in the context of a donation prediction task, in which we attempt to determine the likelihood of an alumnus/alumna to donate, based on the limited background data available for that person. This is not a straightforward task. Previous studies on alumni donations ([122–124]) found that there are many different contributing factors to alumni giving, including having the capacity to give, extracurricular involvement during the time at the university, and the prestige of the university.

We use the dataset described in this section. The ground truth is extracted from the alumni donation history, where those who have donated $10,000 or more to a single fund are designated as having donated, and those who have not donated anything to any fund are designated as not having donated. The resulting set of alumni has a much greater number of non-donors than donors. There are 31,780 non-donors, as compared to 655 donors, which allow models to achieve 98% donor classification accuracy by simply classifying all samples as the majority class. Sampling methods to balance classes are commonly used when working with imbalanced data. We therefore create a balanced dataset by including all of the 655 alumni who donated more than $10,000 and randomly sampling an equal number of those who donated nothing.

Name	Educational	Professional
Amanda Alamns	MSE in Electrical Engineering - 2000	Electrical Engineer, Senior Project Engineer, Principal Systems Engineer
Bob Beustton	BS in Economics - 2000	Financial Analyst Trainee
Claire Carshter	BS/Teaching Certificate in Elementary Education - 2000	Elementary School Teacher, CEO of EduStartup

Table 3.1: Fictitious Alumni Examples

In all of our experiments, we use 10-fold cross validation, resulting in training and test set sizes of 1179 and 131 respectively for each split. We use a logistic regression model with L2 penalties and a regularization parameter C of 1.0 in all cases.[2]

3.3.1 What Makes a Donor?

We want to be able to predict whether a person will donate from her personal and professional attributes. Let us consider the fictitious alumni in Table 3.1 (real examples could not be used due to privacy agreements). Amanda Alamns graduated with a graduate degree in engineering and has steadily climbed the ranks in her professional career. From her position in her career, we can infer that she has the means to donate. Bob Beustton, on the other hand, has somehow remained a trainee for over a decade. It is unlikely that he will make any donations for the time being. Lastly, we have Claire Carshter. If we look solely at her educational history and first job, it appears unlikely that she would donate; the teaching profession is not known for its lucrative opportunities. However, we see that she then proceeded to start her own company. She appears to be a successful individual and is probably more likely to donate because she has the means to do so. Additionally, perhaps her experience at the university helped inspire her to pursue entrepreneurship.

[2]We also obtained results using an SVM classifier, but obtained results and trends similar to those obtained from a logistic regression model. We therefore show results only for the regression model.

3.3.2 Data Description

The work in this chapter is based on a database of alumni information maintained by a large, public Midwest university. We call this dataset Donor Information (DI). In addition, we also have a dataset of public LinkedIn profiles for a subset of the alumni who are in DI. The DI dataset contains each alumna's donation history along with her educational history while at this particular university.

An alumna's educational history contains her major, graduation year, degree level (e.g. Bachelor's level, Master's level, Doctoral Level), and degree type (e.g. BS, MD, PhD). Every record in the LinkedIn dataset contains all job titles and companies listed on the corresponding LinkedIn profile. In our experiments, we only consider the most recent three job titles and companies. We consider the degree level, degree type, degree major, and the most recent three job titles and companies as text fields that are used both as categorical features and as input for the textual feature methods.

There are 56,259 people who appear in both the DI and LinkedIn datasets; we focus on this subset of alumni. Of this set, approximately half have donated some amount. However, many donations are on the order of a few dollars. Therefore, we further hone in on those alumni who have donated more than $10,000 to a single fund.

To represent a person, we extract categorical features such as major, recent job titles, gender, and age, among others. To focus our results on the effects of textual enhancement, we use only the categorical features that can also serve as textual features. Each instance in our dataset is then represented as a feature vector that encodes all of the categorical features by concatenating one-hot embeddings of each feature. Table 3.2 lists all the features that are available in the dataset.

Source	Features
DI	age, gender, graduation year, degree level^T, degree type^T, degree major^T
LinkedIn	city, state, country, most recent three job titles^T, most recent three companies^T, NAICS number

Table 3.2: Dataset features (text fields are marked with ^{T})

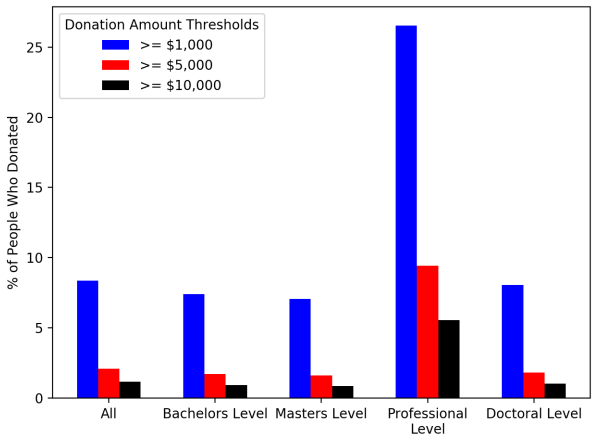

Figure 3.1: Percentage of population who donated for obtained degree levels at several donation amount thresholds.

3.3.3 Qualitative Analysis

To gain further insight into the data, we conduct several qualitative analyses of the backgrounds of donors. We first look at the percentage of people who donate at different degree levels, shown in Figure 3.1. Of the different degree levels, a much higher percentage of those with professional level degrees are donors. This is consistent across different donation amount thresholds. The donor statistics of the other degree levels are consistent with the overall statistics, across the entire population.

We further look at different types of professional level degrees, which are comprised of various medical and law degrees. The five professional degree types with the highest percentages of donors are shown in Figure 3.2. We see that Juris Doctor degrees (J.D.) and Doctor of Medicine degrees (M.D.) are among the top five, which is consistent with the correlation lexicons that we automatically generate, as described in the next

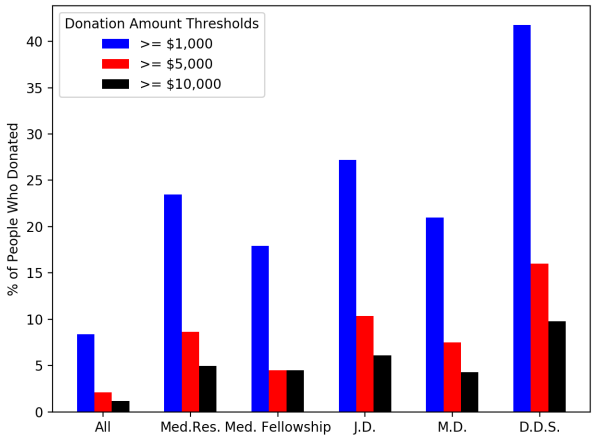

Figure 3.2: Percentage of population who donated for obtained professional degree types at several donation amount thresholds.

section.

Medical residencies (Med. Res.) and medical fellowships (Med. Fellowship) occur much less than J.D.s and M.D.s in our dataset, which could have contributed to their lack of representation in the lexicons.

We also look at the number of popular majors across different departments. We see that those who studied law consistently donated more than the others across the different donation thresholds. We also see that education majors have a higher percentage of donors than other popular majors shown in Figure 3.3. This could be because those who choose to pursue education are more philanthropic by nature, wanting to teach and help others without the promise of a high salary.

Finally, we analyze the industries that donors work in. We obtain the high level industry sectors by only using the first two numbers of the NAICS code in each profile. We then manually selected a few sectors to show in Figure 3.4. The health care sector, with a NAICS code of 62, includes doctors, dentists, and others who provide health care and social assistance. People from this sector have the highest percentage of donors among those shown across donation thresholds, which is in line with what we have seen in the analysis so far. The technical and scientific field also seems to have

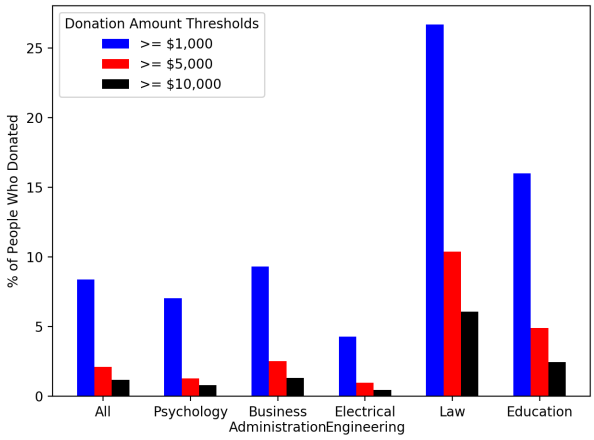

Figure 3.3: Percentage of population who donated for obtained majors of study at several donation amount thresholds.

Figure 3.4: Percentage of population who donated for industry sectors at several donation amount thresholds. The finance sector includes jobs in banking and insurance. The entertainment sector includes jobs in the arts and leisure activities. Scientific jobs entail professional, scientific, and technical services. Jobs in the education sector include teachers at all levels of education, such as elementary and secondary schools and colleges. The health care sector includes physicians, dentists, and others who provide health care or social assistance.

a higher percentage of donors, which could stem from the fact that they have the means to give.

3.4 Text Expansion using Domain-Specific Knowledge

A core hypothesis of our work is that the sparse text that is available in many sources of data, such as our alumni dataset, can still hold much useful information. To make the sparse text useful, we can augment the text with additional information by using natural language processing methods that leverage knowledge about the target domain drawn from within or outside the dataset.

We explore four main methods, described in detail below: (1) word embeddings obtained from a domain-specific corpus; (2) correlation lexicons that aim to identify

from within the dataset additional words that are indicative of donations; (3) lexicons induced starting with a few seeds and using external corpora and graph propagation; and (4) domain-specific distance representations, reflecting the semantic similarity between the textual features and a set of domain-specific seeds.

All of these methods are illustrated, and later evaluated, using the donation prediction task and associated dataset described above.

3.4.1 Domain-specific Embeddings

Unsupervised methods for learning word embeddings represent one of the most recent successes in word representations ([69, 125]). As a first method to expand the text fields we thus use word embeddings.

We construct a corpus of articles that discuss philanthropy-related topics from the New York Times that we will refer to as the NYT Philanthropy News corpus. We use their API[3] and collect 8,525 articles dated from January 1981 to March 2017. The final corpus includes 57 million words, with a vocabulary of 94,623 words. Of those, only the words that occur five times or more are considered during the training of the GloVe model; 32,324 such words exist in the corpus.

We create a set of word embeddings using the GloVe embedding model ([69]) trained on this philanthropy-focused news corpus. We chose to use GloVe as it was shown to have better performance on several word representation and word similarity tasks ([69, 110]). We use 300 dimensions for the embeddings, as is standard practice[4]. For each text field in the dataset, we take the constituent words. The embeddings for all of the words from every text field are then averaged to form a feature vector.

[3] `https://developer.nytimes.com/`

[4] We use the author-provided code for GloVe at `https://github.com/stanfordnlp/GloVe`. All parameters are left as default other than the embedding size.

3.4.2 Correlation Lexicons

Previous work has shown that domain-specific lexicons can be effectively used to induce features for prediction tasks. Specifically, our method is inspired from previous work on sentiment analysis, where a lexicon of positive and negative words generated specifically for Twitter was found to bring significant improvements ([100]). We adapt their method to our task, and generate a lexicon of words that are specific to the task of donation.

Using pointwise mutual information (PMI), as done in ([100]), we measure the strength of association between each word in the dataset and the labels of donation/no-donation. The words are drawn from all the textual fields, consisting of the degree levels, degree types, degree majors, job titles, and job companies. Note that the correlations are calculated only from the training data. Specifically, given a word W, we calculate its PMI score as:

$$\begin{aligned} PMIScore(W) = & PMI(W, donated) \\ & - PMI(W, nondonated) \end{aligned} \quad [3.1]$$

where the $PMI(W, class)$ for any of the two classes is calculated as:

$$PMI(W, class) = \log \frac{p(W, class)}{p(W)p(class)}$$

To create the lexicon, we first calculate the $PMIScore$ for each of the words included in the text fields in the dataset, as described in Section 3.3.2. We then rank the words in decreasing order of their score, and select the top 30 with the assumption that the words that have the highest score are most strongly correlated with the class of donation. Table 3.3 shows the top 10 words from a generated lexicon.

Using the PMI lexicon, we generate 30 binary features, one for each entry in the lexicon. We set the value of each feature to 1 (0), reflecting the presence (absence) of

	Sample words
Top 10 (donation)	educational, partner, m.d., j.d., ceo, professional, board, law, owner, managing

Table 3.3: Sample words from the PMI lexicon

the feature in any of the text fields.

3.4.3 Seed-Induced Lexicons

The third method we consider is to generate a lexicon starting with a few seed words and expanding the set of words using a label propagation algorithm on a lexical graph. We use the SentProp method introduced in ([110]), which was originally proposed for the task of building a lexicon for sentiment analysis.

We first manually build two sets of seed words, associated with philanthropic tendencies and the lack thereof, respectively. Table 3.4 shows these seed words.

	Seed words
Donation	donation, endowment, investment, charity, generosity, benefaction, giver, grantor, donor, donator, benefactor, benefactress, endow, sponsor, backer
Non-donation	miserly, stingy, uncharitable, ungenerous, frugal, selfish, skimping, scrimping, tightfisted, closefisted, parsimonious, inhospitable, greedy, cheap

Table 3.4: Seed words used to generate the SentProp lexicon

We then build a weighted lexical graph using the words from the text fields in the dataset, as well as all of the seed words. Each word is connected to its nearest *10* neighbors by using a measure of cosine similarity applied on word embedding

representations for each word that is present in both the dataset vocabulary and the trained word embeddings. We use GloVe embeddings, following the original SentProp implementation.

The donation and non-donation labels are then propagated through the graph using a random walk method. Finally, a word's donation score is calculated as the probability of a random walk from the corresponding seed set hitting that word. In our experiments, we try lexicon generation using both generic pre-trained GloVe embeddings[5] as well as GloVe embeddings that we train on the NYT Philanthropy News corpus. They perform comparably; we only show results using the latter embeddings.

To create the final lexicon, we take only the words that have a donation association score higher than 0.7. We chose this threshold heuristically; lower thresholds introduced noisy words and higher thresholds excluded many words that appear in the dataset. Sample words from the resulting lexicon are shown in Table 3.5. As with the PMI lexicon, we create a feature for each of the lexicon words, and set its value as 1 (0) depending on whether the feature is present (absent) among the words in the text fields.

	Sample words
NYT GloVe based	contributor, giving, investor, management, banking, mutual, venture, institutional, profit, corporate, philanthropy, market, cash, asset, hedge, managed

Table 3.5: Seed-induced lexicon entries using label propagation on graphs

[5] `https://nlp.stanford.edu/projects/glove/`

Source	Features
BASELINES	
DI	degree level, degree type, degree major
LinkedIn	most recent three job titles, most recent three companies
TEXT EXPANSION FEATURES	
DomainEmbed	300-dimension GloVe embeddings trained on the donation corpus, averaged over all the words in the text fields
CorrelLex	30-word correlation lexicon generated from training data (text fields in both DI and LinkedIn); one feature for each lexicon word, reflecting presence/absence among words from text fields
SeedProp	Seed-induced donation lexicon using label propagation on a lexical graph formed by using pretrained GloVe embeddings; one feature for each lexicon word, reflecting presence/absence among words from text fields
SeedSim	Semantic similarity between the text fields and the 15 donation seeds, using cosine similarity between pre-trained GloVe embeddings; one feature for each of the 15 donation seeds

Table 3.6: Summary of features

3.4.4 Seed-Similarity Embeddings

Finally, as an alternative to the previous seed-induced lexicon method, we also consider a method that measures the semantic distance between the words in the text fields and the donation seed words. The hypothesis behind this method is that we can circumvent the need for a domain-specific corpus by measuring the distance between a small set of domain words and the text fields.

We use the same seed set as listed in Table 3.4 (row Donation). We use the pre-trained GloVe embeddings with 300 dimensions. For each seed word, we find the maximum cosine similarity score between that word's embedding and each of the word embeddings from the text fields. The result is a feature vector that reflects these maximum similarity scores, and is the same length as the seed set.

3.5 Results and Discussion

We evaluate the performance of the donation prediction task described in Section 3.3 using the original features available in the dataset, as well as expanded feature sets obtained with the four text expansion methods described above. Table 3.6 summarizes the features we use, described in the previous sections.

The top part of Table 4.6 shows the results obtained with the two baselines (DI features, and DI combined with LinkedIn features), while the bottom part of the table shows the results obtained when augmenting the top performing baseline with the various text-expansion features. We combine features by concatenating their feature vectors. This combination method has been shown to work well in many applications ([114, 126, 127]).

Statistical significance over the DI+LinkedIn baseline is calculated using the McNemar two-tailed test. We used an alpha value of 0.05.

Source	Accuracy
BASELINES	
DI	68.8%
DI+LinkedIn	76.8%
TEXT EXPANSION FEATURES	
DI+LinkedIn+DomainEmbed	81.3%*
DI+LinkedIn+CorrelLex	80.1%*
DI+LinkedIn+SeedProp	78.5%*
DI+LinkedIn+SeedSim	77.2%

Table 3.7: Donation prediction results using text expansion methods. Results with * are statistically significant compared to the DI+LinkedIn baseline system.

Among the four text expansion methods, the correlation lexicons, domain-specific embeddings, and seed-induced lexicon result in significant improvements over the baselines as seen from Table 4.6. The seed-similarity embedding features also bring small improvements, but they are not found to be significant.

To gain further insight into the performance of these models, we perform several

	DI+LinkedIn	+SeedSim	+SeedProp	+DomainEmbed	+CorrelLex
DI+LinkedIn	1.0	0.90	0.83	0.56	0.63
+SeedSim		1.0	0.83	0.57	0.63
+SeedProp			1.0	0.56	0.66
+DomainEmbed				1.0	0.68
+CorrelLex					1.0

Table 3.8: Pearson correlation coefficients among the output of the four models and the baseline. Each model includes the DI and LinkedIn categorical features with the specified additional single feature type.

additional analyses and evaluations, which we describe next.

3.5.1 Model Correlation

First, we measure the correlation between the output produced by the top baseline model (DI+LinkedIn) and by the four different methods considered. Table 3.8 shows the Pearson correlation between all the pairs of two models. As seen in this table, most of the models are medium correlated, which indicates there is some overlap between the predictions they make. The models that are most divergent from the baseline are the correlated lexicons (CorrelLex) and the domain specific embeddings (DomainEmbed), which is also reflected in the higher performance of these models (see Table 4.6). The highest correlation is found between the model that measures the similarity with the seed set (SeedSim) and the model that performs label propagation on a lexical graph starting with the seed set (SeedProp); their high correlation is likely a reflection of the dependence of these two models on the same seed set.

3.5.2 Classification with Less Distinguishable Classes

We also perform an evaluation for a classification task where the division between donors and non-donors is less clear. Specifically, we again consider all the alumni who donated $10,000 and above as donors, but now we consider alumni who donated any amount below $10,000 as a non-donor. The non-donors are randomly sampled from the instances corresponding to people who donated less than $10,000. Table 3.9 shows

Source	Accuracy
DI	66.5%
DI+LinkedIn	72.6%
DI+LinkedIn+DomainEmbed	75.3%*
DI+LinkedIn+CorrelLex	75.6%*
DI+LinkedIn+SeedProp	73.3%
DI+LinkedIn+SeedSim	73.9%

Table 3.9: Classification results when non-donors include alumni who donated any amount below $10,000. Results with * are statistically significant compared to the DI+LinkedIn baseline.

the results obtained during these evaluations. As expected, all the results are lower than the ones obtained during the earlier evaluations. In this more difficult setup, the use of correlation lexicons and domain embeddings continues to bring consistent improvements over the baseline.

3.5.3 Influence of Seeds on Sentprop Lexicon

We analyze the effects of changing the seed words used for SentProp by looking at the overlap between our original lexicon and the new lexicons generated using different seed words. To highlight the effects of different donation words, we change the donation words to be entirely different from those used in our experiments, but maintain the topic of donation among the words. For each of the three different donation word sets, the non-donation words remain the same as those used in our experiments. To understand the influence of the non-donation words, we also choose a set of random words as non-donation words. For this, we retain the same set of donation words as used in our experiments. The chosen words are shown in Table 3.10.

For these different sets of seed words, we generate lexicons with SentProp on the NYT Donation corpus at two different association score thresholds. We measure the overlap between the new lexicons and the original one used in our experiments by calculating their Jaccard similarity and overlap coefficient. For two sets of words, X

and Y, we have

$$Jaccard(X, Y) = |X \cap Y|/|X \cup Y|$$

and

$$Overlap(X, Y) = |X \cap Y|/min(|X|, |Y|).$$

The new lexicons corresponding to altered donation words (Set 1, Set 2, Set 3), generated at an association score threshold of 0.7, contain many more words that are not related to philanthropy. The large size of the lexicon is indicative of this. This could be a result of the seed words not being as unambiguously tied in topic as the original set of seeds. However, the overlap coefficients are close to 1, showing that the original lexicon words are present in the new lexicons. This implies that the donation topic was still captured, but with much more noise.

We raise the association score threshold to filter out the less relevant words. Overall, the lexicons resulting from Set 1, Set 2, and Set 3 still maintain much overlap with the original lexicon. Set 2's lexicon has a much lower overlap coefficient than Set 1 or Set 3. This is likely because Set 2's donation seeds contain words like "kind" and "charitable" that have more ambiguous meanings.

Interestingly, having random non-donation words does not greatly perturb the captured topic of the lexicon. All words generated also appear in the original lexicon. Additionally, the number of words is actually smaller than the original set. This may be because having random non-donation words encourages SentProp to choose words that are unambiguously related to the donation words. There is a separation of the donation topic from effectively all others, rather than from just the non-donation topic. This is desirable in applications where we are primarily interested in generating a lexicon related to one theme, rather than two polar themes, as is the case here.

Seed Words	Donation	Non-donation
Set1	contribution, gift, funding, foundation	miserly, stingy, uncharitable, ungenerous, frugal, selfish, skimping, scrimping, tightfisted, closefisted, parsimonious, inhospitable, greedy, cheap
Set2	kind, supporter, charitable, patron	
Set3	contribution, gift, funding, foundation, kind, supporter, charitable, patron, compassionate	
NegRand	donation, endowment, investment, charity, generosity, benefaction, giver, grantor, donor, donation, benefactor, benefactress, endow, sponsor, backer	cattle, evanescent, vague, jittery, trade, grade, excited, signify, clear, toad

Table 3.10: Different sets of seed words used for SentProp. Sets 1-3 retain the same set of non-donation words as used in the experiments, but with different donation words. NegRand retains the same set of experiment donation words, but with random non-donation words.

Seed Words (Threshold)	Lexicon Size	Jaccard Sim.	Overlap Coef.
Set1 (0.7)	897	0.14	0.95
Set2 (0.7)	1667	0.08	1.00
Set3 (0.7)	929	0.13	0.95
NegRand (0.7)	49	0.37	1.00
Set1 (0.8)	37	0.17	0.65
Set2 (0.8)	126	0.21	0.35
Set3 (0.8)	48	0.21	0.65
NegRand (0.8)	0	0.00	0.00

Table 3.11: Number of words, Jaccard similarity, and overlap coefficients for different sets of seed words at different association score thresholds. Jaccard similarity and overlap coefficient are calculated with respect to the generated lexicon used in the experiments. The original generated lexicon has 132 words.

3.5.4 Error Analyses

To better understand the performance of our methods, and where and when they fail, we perform several error analyses. Specifically, since the job related information from LinkedIn was the most varied, and therefore the area that could benefit the most from our text expansion methods, we mainly focus our analyses on how well our features understood LinkedIn information. We have anonymized the examples below by excluding names and modifying job titles to be generic.

Categorical features were not able to understand complex or non-standard job titles such as "Director of Major, Planned, and Special Gifts", "Senior Director of Major Gifts", or "CEO of A Philanthropic Trust". These particular job titles are highly indicative of philanthropic tendencies, but the categorical-only DI+LinkedIn model classified these individuals as non-donors. The categorical model also was not able to correctly detect people working in known high-pay fields because of non-standard titles. For instance, one donor is a "Pulmonary Specialist", which is a type of doctor. From our data (Figure 3.4), we can see that health care professionals are the most charitable individuals. However, the categorical model was unable to make the association between "Pulmonary Specialist" and the health care profession.

The embedding features helped find such associations. The "Pulmonary Specialist" was found to be a donor by the model that incorporated embedding features. It was also much better at detecting individuals with advanced career positions such as "Senior Vice President", "Strategic Advisor", and "Executive Director". While these titles may seem obvious, there exist many variations on advanced titles, such as "Managing Director", "Principal Advisor", and "Creative Director". Embedding features implicitly help the model understand that positions like these are indicative of donors, without explicitly having a list of such titles. However, the embeddings were not good at distinguishing between those who had a single position indicative of a donor and those who had a history of such positions.

The correlation lexicon features focused on finding individuals that held multiple indicative positions. For example, some of the donors that were correctly identified only by including CorrelLex each had at least three advanced career positions. One was a "Senior Counselor", "CEO", and "Founder"; another was a "Senior Development Officer", "Consultant", and "President"; and yet another was a "Senior Manager", "Chief Operating Officer", and "Senior Clinical Manager". These results follow the fact that the generated correlation lexicons from Table 3.3 seem to mainly focus on high income or advanced titles. However, this misses people who are philanthropic but do not necessarily hold traditional advanced positions.

Some of the donors that were correctly identified only by using SeedProp had titles such as "Head of Police Board", "Workplace Learning Specialist", "Program Director/Scholarship Manager", or worked at foundations. These careers involve public service, interacting with people, and being in environments that are geared towards philanthropy. SeedSim produced similar results, though the detected associations were limited to very explicit indicators, such as someone being an "Evangelist".

3.5.5 Evaluation on Other Datasets

We also want to determine to what extent our methods can be applied to other datasets. Although there are public donation records available at crowdfunding sites such as Kickstarter.com and DonorsChoose.org, there is usually little information revealed about the donors themselves beyond what they have donated to.

We therefore evaluate our proposed text expansion methods on a different task: gender classification on a dataset of blog profiles collected from Blogger.com. Previous work has shown that it is possible to detect demographic information such as gender from writings and social media content [128–131].

Bloggers can choose to fill in information, such as gender, occupation, and interests, on their profile page. We use a set of 76,971 profiles that have both gender and

Source	Features
Blogs	gender, interests,^{T}, occupation^T, city, state, country, introduction^T, movies^T, music^T, books^T

Table 3.12: Blog dataset features (text fields are marked with ^{T})

Source	Accuracy
Blogs	70.6%
Blogs+DomainEmbed	83.3%*
Blogs+CorrelLex	75.6%*
Blogs+SeedProp	72.0%*
Blogs+SeedSim	74.5%*

Table 3.13: Gender prediction results on blog profiles. Results with * are statistically significant compared to the Blog baseline.

interests listed and are in the USA. The full set of features is listed in Table 3.12. We classify each blogger as male or female based on the information available on their profile.

Our text expansion features are replicated for gender in this setting. The domain embeddings are trained on the blog dataset. The correlation lexicon is generated from the training set of blog data. Gender-based seed words are used for SeedProp and SeedSim. The results are shown in Table 3.13. All of the text expansion methods improve significantly over the baseline categorical method, with the domain embeddings (DomainEmbed) yielding the highest performance. These results demonstrate that our methods can be successfully applied to other datasets.

3.6 Conclusions

In this chapter, we explored whether we can enhance sparse textual content to improve data-driven predictions using the task of alumni donation behavior prediction.

We introduced a dataset of alumni donations, and we qualitatively analyzed the

donations and the backgrounds of the donors to highlight the differences between the backgrounds of donors and non-donors as well as the patterns of donations attracted by different academic departments.

We used four different methods of expanding sparse text, including lexicon generation methods and text embedding methods. We evaluated these methods on the task of predicting whether someone is likely to donate, and compared with baseline models that do not make use of any textual features.

We showed that we can classify alumni who exhibit large donation behavior from non-donation behavior with an accuracy of up to 80%. We also showed that the enrichment of sparse text through the extraction and use of textual features does benefit model performance when predicting donation prediction. Our domain-specific embeddings and correlation-based lexicon consistently improved over the baseline models that only use categorical features. We also showed that our methods can be successfully applied to other sparse-text datasets.

CHAPTER IV

Understanding Personal Change Behavior Using Social Media

4.1 Introduction

Many people aim for change, but not everyone succeeds. While there are a number of social psychology theories that propose motivation-related characteristics of those who persist with change, few computational studies have explored the motivational stage of personal change. In this chapter[1], we investigate a new dataset consisting of the writings of people who manifest intention to change, some of whom persist while others do not. Using a variety of linguistic analysis techniques, we first examine the writing patterns that distinguish the two groups of people, looking at what topics they discuss, their writing style, and the emotions they express. Drawing on these characteristics, we then build a classifier to identify the people more likely to persist, based on their language. Our experiments provide new insights into the motivation-related behavior of people who persist with their intention to change.

People aim for personal change at different points in their lives [132]. A glance at a list of top-selling books readily yields self-help manuals whose content ranges from

[1]The work in this chapter was aided by input from Rada Mihalcea and help from two undergraduate students, Xueming Xu and Yiwei Zhang. This work was presented as a poster at the 2020 International Conference on Social Informatics.

implicitly motivating ("Seven Habits of Highly Effective People" [133]) to explicitly calling for action ("Lean In" [134]). However, simply wanting change is not sufficient to achieve change. Persistence through the process of pursuing personal change is important for actual change to happen, and changes rarely happen overnight. Often, research on behavior change focuses on understanding what makes people committed to regular or increased action, such as exercise [135], or refraining from certain actions such as not overeating [136] or not smoking [137]. An ever-growing number of technological tools, such as food diary apps and wearable activity trackers, have emerged to help monitor and motivate healthy behavior [138, 139]. Regardless of the tools that they use, if someone is not ready for change yet, the intervention is likely to fail [140].

Stage-based models of intentional behavior change posit that people progress through a sequence of two stages [140, 141]: *motivation* and *volition*. In the initial *motivation* stage, a person develops an intention or goal to act. A person's intention to adopt better behavior depends on factors such as: *risk perceptions*, or the belief that one is at risk of a negative outcome (e.g. "If I keep procrastinating, I'll fail all my classes."); *outcome expectancies*, or the belief that behavioral change would improve the outcome (e.g. "If I can have a more consistent daily routine, I will be more successful at work."); and *perceived self-efficacy*, or the belief that one is capable of doing the desired actions.

In this chapter, we seek to understand the characteristics of people who are in the *motivation* stage of behavior change, and how they talk about behavior change. Traditional behavior change tactics focus on convincing people to take action without consideration for what happens during the lead up period [140]. Insight into how people act during these earlier stages can help us better understand their needs and inform interventions, such as recommending social media content that exemplifies healthy approaches to self-improvement. They can also help predict later behavior

and persistence using early signals.

4.1.1 Research Questions

We explore how we can computationally model change-seeking behavior and distinguish between those who maintain persistent interest in personal change during the *motivation* phase and those who do not. People often turn to social media to express their thoughts and emotions, which provides a rich data source for studying their perceptions and thoughts [142].

We address our research questions using a dataset consisting of the writings of 536 people from an online community focused on self-improvement (the Reddit community r/getdisciplined). In this dataset, we identify those who post frequently and those who post infrequently to identify persistent and non-persistent commitment to change. We analyze the topics, linguistic style, and expressed emotions of the posts authored by the persistent and non-persistent groups of people. Specifically, in this chapter, we address three main research questions:

1. What are the aspects of life that people want to improve?

2. What linguistic style do people use to signal their persistent interest in self-improvement?

3. How does persistent interest in self-improvement reflect in the emotions that authors express?

Using the features tested in the three separate analyses, we are able to classify persistently and non-persistently active authors with over 60% accuracy, even when using the posts that authors write prior to joining r/getdisciplined. Considering both the descriptive and predictive analyses, our findings indicate that persistent interest in change can be signalled by early changes in behavior in online discussions.

4.2 Related Work

Behavior Change. Personal and behavioral change have a long history in the field of psychology [143]. Improving health behaviors motivated much work in areas like smoking cessation and increasing physical activity. However, work on understanding how to encourage positive change has expanded to cover countless areas, like decreasing crime [144], increasing environmentally friendly behavior [145], and enhancing overall well-being [146]. This previous work has shown that many factors can influence an intervention's efficacy, a person's willingness to change, and which strategy to choose for a given person [147]. Further, an intervention's efficacy may change based on where a person is in their process of change. Different stages in the process can be correlated with different levels of attitudes, such as risk perception or self-efficacy [148]. Such attitudes capture a person's estimate of their ability to perform and succeed in challenging situations and are often reflected in the actions that people choose to take or not to take later in later stages [149–151]. Several theories of behavioral change delineate stages of change and advocate for interventions tailored to each stage [140, 148].

Self-Improvement in Online Communities. In recent years, many have turned towards online communities and platforms, such as Reddit and Facebook, to help them make positive personal changes. The anonymity available in online discussions helps combat fears of stigma or lack of understanding [152]. This relative freedom of expression enables researchers to analyze how people seek help through online channels and what they seek [153]. People join online communities to obtain support from those with similar experiences [154], to ask for guidance and resources [155], and to seek accountability [156]. Such support can lead to higher perceived self-efficacy [157].

However, as noted by prior work in behavior change, the type of help needed can be highly dependent on one's personal characteristics and situation. In our work, we

seek to better understand this using Reddit. There has been considerable effort spent on learning about people's demographic attributes from social media posts [158]. Work has also targeted internal attributes, such as personality and value, which can be more difficult to extract but can provide richer features for downstream tasks [159]. However, few have studied general intentional personal change efforts based on social media posts. We tackle uncovering the underlying linguistic characteristics of those who maintain persistent interest in self-improvement.

4.3 Data

We focus on a Reddit community called r/getdisciplined, where people seek and give advice about how to achieve life goals and build better habits. This community boasts over 768,000 members as of March 2021 and is one of the largest self-improvement subreddits on Reddit. Whereas most self-improvement groups target specific behaviors or goals, such as exercising, losing weight, dieting, or improving mental health, this subreddit targets improving general mental habits. For instance, people ask questions such as "How do I relearn doing things just for fun?" and "How do I stop caring about people and craving their attention?" as opposed to questions that are more specific to activities like "Tips for increasing strength in arms?" or "How do I eat properly?"

Each submission, or original post that is not a comment, must designate the intent of the post using a set of specific tags. One can seek advice ([NeedAdvice], [Question]), give advice ([Advice], [Method]), facilitate discussion and accountability ([Discussion], [Plan]), or talk about r/getdisciplined overall ([Meta]). Most submissions seek advice from the community and tend to discuss fundamental issues such as procrastination, lack of motivation, and time management. A sample of submissions are shown in Table 4.1.

From the submissions, we can see clear distinctions between people who seek help.

Post
I am a chronic procrastinator without any hope... do you know any drastic measures that might help me turn my life around? I have been procrastinating intensely for pretty much my whole life. It just seems to be a part of my personality at this point. I tried many things but I could never handle it. I have been mildly depressed for a long time now and have no belief in myself whatsoever.
How do you balance Parkinson's Law with producing quality work? I often find myself spending a lot of time on tasks, and I recently read about Parkinson's Law from Tim Ferriss' 4 Hour Workweek. The law states that a project or task will expand to fill the time you have allotted to it. It obviously takes a lot of time and hard work to produce something of quality, whether it be music, writing, etc. How do you stave off Parkinson's Law while still producing something of quality?

Table 4.1: Sample [NeedAdvice] posts from the r/getdisciplined subreddit.

In the first submission, the author expresses that they think a negative trait, procrastination, is probably a set part of their personality and that they do not believe in themselves, resulting in expression of negative emotions ("mildly depressed"). On the other hand, the second submission seeking advice does not make any self-deprecating statements and asks only for productivity tips ("producing quality work"). This implies that they believe in their ability to change their habits with guidance. Across all submissions, it is clear that the writers have made concerted efforts to understand their own behavior.

We focus on people who join r/getdisciplined and then become active during a period of five months, from 2017/1 to 2017/5. These are people who had an initial intent to change which turned into continued engagement and persistent intent.[2] We categorize people as persistently active in the subreddit, or *persistent*, if they have posted at least four or more times in the given five months.[3] Only people who have posted in three unique months before and after the target period, respectively, are considered. This pre-processing ensures that there is sufficient data for analysis before, during, and after each person's participation in r/getdisciplined. We then randomly sample an equal number of *non-persistent* people, or people who have

[2]Data collected using `http://pushshift.io`

[3]This number of posts is the 90th percentile among people who posted during this time.

Data Summary	
Total Number of Users	536
Posts from r/getdisciplined	6010
Posts from other subreddits	336455

Table 4.2: Summary statistics about the dataset, such as the number of users and posts.

posted only once in the 5 months, with the same requirement for posts before and after. Table 4.2 shows the number of users and posts in our dataset. The total number of users, including both persistent users and a random sample of non-persistent users, is 536.

4.4 Characteristics of Persistent Interest in Change

We address the study's questions about persistence in personal change by analyzing the discussed topics, the linguistic style, and the expressed emotions in Reddit posts. We analyze both their general behavior on Reddit *prior* to joining r/getdisciplined as well as their *initial behavior* within r/getdisciplined. Investigating how people act before joining r/getdisciplined helps us learn about the mental or behavioral patterns that indicate a higher likelihood of their intent to change their behavior. As a complement to prior behavior, an individual's first post indicates how they are approaching behavior change.

4.4.1 What Are the Aspects of Life that People Want to Improve?

We uncover the particular areas of life that people seek to improve and their prevalence in discussion. We use topic modeling techniques to uncover the areas of interest that people discuss in their online posts, both within and outside of the context of personal change.

Participation in Subreddits. The subreddits, or Reddit communities, in which a person posts shows the general topics with which they engage. We therefore calcu-

User type	Top Subreddits
Persistent	Advice, DotA2, EliteDangerous, **Fitness**, GameStop, GlobalOffensiveTrade, **LifeProTips**, MakeupRehab, MarvelPuzzleQuest, RWBY, argentina, aww, conspiracy, cowboys, explainlikeimfive, fantasyfootball, hearthstone, me_irl, **personalfinance**, photography, relationships, summonerschool, wow
Non-persistent	BigBrother, CFB, CringeAnarchy, DeadBedrooms, HelloInternet, IAmA, NoMansSkyTheGame, NoStupidQuestions, OutreachHPG, Roadcam, SquaredCircle, SubredditDrama, WTF, Warframe, baseball, bjj, cars, casualiama, nottheonion, skyrimmods, skyrimrequiem, slatestarcodex, smashbros
Both	AdviceAnimals, AskMen, AskReddit, Jokes, MMA, Overwatch, Showerthoughts, The_Donald, funny, gaming, gifs, leagueoflegends, mildlyinteresting, movies, nba, news, nfl, pcmasterrace, pics, pokemon, pokemongo, politics, soccer, television, todayilearned, videos, worldnews

Table 4.3: Top 50 subreddits prior to joining r/getdisciplined for persistent and non-persistent users respectively, divided into those that correspond to only one group and both groups. Subreddits relevant to self-improvement are bolded.

late how frequently each user posts in every subreddit, considering only the subreddits that receive 10 posts in aggregate by users that we observe. We consider only the posts made by the users before their first post in r/getdisciplined.

We show the top 50 subreddits for persistent and non-persistent users prior to joining r/getdisciplined in Table 4.3. We can see that persistent individuals are active in a number of topic-specific self-improvement subreddits, such as **Fitness**, **LifeProTips**, and **personalfinance**. Non-persistent individuals participate in many more gaming subreddits, i.e. related to leisure rather than self-improvement. Both groups post in popular subreddits like **AskMen**, **AskReddit**, and **funny**; the prevalence of "ask-X" related subreddits suggests a level of open-mindedness to change that one would expect of people potentially committed to change.

Topics of General Discourse. To gain further insight into the topics that motivated people engage with, we turn to topic modeling. Latent topics can group concepts that overlap between subreddits and ones that differentiate posts in the same subreddit. We use the Latent Dirichlet Allocation (LDA) model [160] to discover

topics in our dataset. LDA takes a set of documents, D, which each contain a sequence of words, and outputs a set of latent topics that make up the documents. We treat each post as a document d and consider all posts made by our target users in the six months prior to joining r/getdisciplined.

To choose the number of topics for the LDA model, we train models on the general posts made prior to r/getdisciplined with $k = 5, 10, 15, 20, 30$ and then manually inspect the resulting topics and their constituent words to evaluate intra-topic coherence and inter-topic separation. To do this, for each value of k we look for resulting topics whose words seemed to primarily be related to one topic, as well as having a lower number of overlapping words between topics. We intentionally keep to a smaller number of topics since we qualitatively found that increasing the number of topics past 30 led to much lower coherence. We choose the 30-topic LDA model for our analysis and later classification experiments. Using the resulting model, we examine the content of user posts pre-r/getdisciplined.

In Table 4.4, we show a subset of topics and label them through a manual inspection of the top words associated with the topic from the LDA model (e.g. "school", "college", and "classes" correspond to the topic labeled "Education"). We note the topics that differ significantly across posts made by persistent and non-persistent users before joining r/getdisciplined. We see that persistent users talk more about education, indicating pre-existing interest in a common area of self-improvement. On the other hand, non-persistent users discuss music, politics, and Reddit more, which are general or leisure interests that may be less related to one's personal life.

Topics of Interest in Self-improvement. The topics that people discuss in general on Reddit differ greatly from those that are discussed in a focused subreddit. To hone in on the content specific to r/getdisciplined, we train another 30-topic LDA model using all the posts made in r/getdisciplined between 2016/1 to 2020/2.

We represent each initial post with the distribution of topics that it contains,

Feature	P	NP	P-NP
1st post			
Studying	0.072	0.037	0.036*
Routines	0.114	0.085	0.028
Productivity	0.062	0.073	-0.011**
Mental Health	0.102	0.105	-0.002
Time	0.165	0.118	0.047*
Goals	0.086	0.071	0.015
Encouragement	0.021	0.049	-0.028*
Habits	0.129	0.083	0.046
Conversation	0.046	0.102	-0.056*
Work	0.130	0.125	0.005
Prior six months			
Music	0.092	0.093	-0.002**
Relationships	0.213	0.180	0.033
News	0.147	0.148	-0.001*
Finance	0.172	0.186	-0.014*
Politics	0.133	0.180	-0.047**
Gaming	0.164	0.188	-0.024
Education	0.228	0.189	0.039**
Reddit	0.102	0.122	-0.019**
Automobiles	0.112	0.133	-0.021*
Family	0.314	0.300	0.014

$* - p < 0.05, * * - p < 0.01, * * * - p < 0.001$

Table 4.4: Mean distributions of topics among posts for persistent (P) and non-persistent (NP) users, as well as the differences between them (P-NP). Statistical significance is determined using a two-sided T-test, with the Benjamini-Hochberg Procedure applied to control for multiple hypotheses testing.

Feature	**1st post** P	NP	P-NP	**Prior 6 mon.** P	NP	P-NP
Linguistic Features						
Readability	-9.800	43.002	-52.802***	49.163	52.971	-3.807
Post Length	96.276	47.522	48.754***	40.572	34.548	6.024**
Emotions						
Anticipation	0.124	0.108	0.016	0.115	0.115	0.001
Disgust	0.031	0.024	0.007	0.044	0.044	-0.001
Sadness	0.045	0.042	0.003	0.067	0.066	0.002
Trust	0.109	0.090	0.019	0.135	0.133	0.003
Surprise	0.032	0.033	-0.000	0.054	0.054	-0.000
Anger	0.038	0.029	0.009	0.059	0.066	-0.007*
Negative	0.116	0.103	0.014	0.131	0.138	-0.007
Joy	0.060	0.062	-0.002	0.098	0.095	0.003
Fear	0.059	0.047	0.013	0.070	0.073	-0.003
Positive	0.210	0.199	0.011	0.226	0.216	0.010

$* - p < 0.05, * * - p < 0.01, * * * - p < 0.001$

Table 4.5: Mean feature values of linguistic and emotion features in posts from persistent (P) and non-persistent (NP) users, as well as the differences between them (P-NP). Note that the differences for different measures are on different scales. Statistical significance is determined using a two-sided T-test, with the Benjamini-Hochberg Procedure applied to control for multiple hypotheses testing.

according to this LDA model. In Table 4.4, we again show a subset of topics and note those that differ significantly between the two groups of users. Persistent users discuss studying and academics more than non-persistent users, as well as time and time management, showing interest in longer-term shifts in how to go about their life. Non-persistent users engage in more words of encouragement and conversation, perhaps trying to establish connection with the community to increase the likelihood of helpful responses. They also speak about productivity more than persistent users, which is indicative of asking for straightforward productivity tips to solve immediate problems (e.g. "What apps can I use to help with work?"), rather than tackling longer-term change (e.g. "I really want to gain some discipline and self control. I would appreciate advice!").

4.4.2 What Linguistic Style Do People Use to Signal their Persistent Interest in Self-Improvement?

Patterns in how people express themselves through language can potentially tell us about how they think. Linguistic style has been shown to reflect numerous behavioral characteristics such as personality [161], and intent [162]. We look at the length of each post, taking the number of words contained in the post as a feature. We also consider each post's readability as defined by its Flesch Reading Ease score [163]: higher scores indicate longer average word length and sentence length, which implies more difficulty in reading. We compute these two scores for each post and use these two values as features in our predictive models. As before, we analyze the posts of persistent and non-persistent users both prior to posting in r/getdisciplined and in their first post in the subreddit.

General Linguistic Style. We show the average post lengths and Flesch Reading Ease scores for the prior posts of persistent and non-persistent users in Table 4.5. Persistent users tend to have longer posts than non-persistent users, which could

indicate a more committed writing style (e.g., explaining all necessary details of a situation when posting). In contrast, the two groups' posts do not differ much in readability.

Self-Improvement Linguistic Style. Next, we look at the average post lengths and readability scores of initial posts in r/getdisciplined (Tab. 4.5). In contrast to the pre-joining posts, persistent users write significantly longer posts and lower readability, indicating more complex posts. Initial posts that ask for help without self-deprecation, such as the second post in Table 4.1 can include many details about the situation at hand so that others can offer pertinent advice.

4.4.3 How Does Persistent Interest in Self-improvement Reflect in the Emotions that Authors Express?

The third research question considers trends in emotional expression among people seeking motivation for change. Emotions can signal attitude towards one's intended behavior change. For instance, someone who believes that success is based on innate ability or who expects that they will fail at difficult tasks will probably shy away from goals that require large effort [164]. On the other hand, those who believe success results from hard work or believe in their own ability to tackle challenges may be more persistent in their efforts [165].

To analyze such trends, we use the NRC Emotion Lexicon [100, 103], which contains English words and their associations with positive and negative sentiment as well as eight basic and prototypical emotions [166]: *anger, fear, anticipation, trust, surprise, sadness, joy, and disgust*. Complex emotions, such as *regret* or *gratitude*, can typically be viewed as combinations of these basic emotions. The lexicon contains 14,182 general domain words, each of which can be linked to multiple emotions.

Emotions in General Discourse. Building on our previous observation about the prevalence of emotional words, we now compare the rate of use among persistent

and non-persistent people. We compute the total proportion of emotions expressed for each person by averaging the counts of emotion words used across the person's posts. Comparing the persistent and non-persistent people, we found that most of the emotions are equally found in posts by both groups. However, non-persistent users express more anger in general, which may indicate a tendency to be more easily discouraged when faced with difficulty in everyday situations.

Emotions of Self-improvement. We use the same emotion lexicon to extract the expressed emotions in each initial post to r/disciplined. The expressed emotions in first posts that do not differ significantly between persistent and non-persistent users (Table 4.5). However, we see that there is a general trend among everyone of expressing positive sentiment, anticipation, and trust, which signals that they are hopeful with respect to self-improvement and are open to discussing problems and solutions. There is also negative sentiment, which can indicate dissatisfaction towards their current situation and therefore desire to change.

4.5 Predicting Persistence in Change

Our analyses have identified that the people who persist in their self-improvement efforts exhibit consistent linguistic differences in topics, writing style, and emotional expression, versus those who do not persist. As a natural next step, we ask whether we can leverage these characteristics to automatically distinguish between these two groups. We set up a prediction task to determine whether a user is likely to become a persistent or non-persistent user on r/getdisciplined by considering: (1) their language use within six months prior to their initial post on r/getdisciplined; (2) their language use in their first post; and (3) their combined language use within the six months prior and their first post on r/getdisciplined.

To provide more fine-grained semantic representation of the post language, we also construct word embeddings [125] from the text of each post, using word2Vec

embeddings pre-trained on news text.[4] Word embeddings are useful in capturing fine differences between words, such as differences in sentiment valence between similar words (e.g. "good" vs. "great"). For each initial post in r/disciplined, we average the word embeddings of each word in the post to generate a per-post embedding. To represent prior posts, we average the per-post embeddings for all posts of each user from the six months prior to joining r/getdisciplined. For readability, we also include an aggregate readability score based on a number of different readability metrics, in addition to the Flesch score used earlier.[5]

We compare the performance of classifiers that use different combinations of the linguistic features that we have shown to correlate with persistent behavior. Our task is the binary prediction of whether a user will continue to engage (persistent user) or leave after an initial post (non-persistent user). The experiments are performed using SVM classifiers [167] and evaluated using 10-fold cross validation.[6] Since our dataset is balanced, both the random and majority class baselines correspond to an accuracy of 50%.

We present the results in Table 4.6, with classification performance shown for each feature set derived from a user's prior behavior, their first post in r/getdisciplined, and the combination of all features. Using all features, our models are able to achieve an average accuracy of over 60%. This shows that people who persist with change can be distinguished from those who do not, even before they commit to change by posting in r/getdisciplined. That said, the models that use only features from each user's initial post in r/getdisciplined yield the highest performance overall. This is in line with previous work showing that the initial posts that someone makes in a conversation can reliably predict future outcomes, such as whether a debate will derail [168] or a user will remain loyal to a community [169]. Moreover, someone's first post

[4] `https://code.google.com/archive/p/word2vec/`

[5] `https://pypi.org/project/textstat/`

[6] We used the SVM classifier, with default parameters, as applied in Scikit-learn: `https://scikit-learn.org`

Features	Acc	Prec	Rec	F1
1st post				
Readability	0.61	0.59	0.72	0.65
Post Length	0.60	0.57	0.83	0.67
Emotionality	0.54	0.54	0.57	0.55
W2V	0.60	0.64	0.46	0.53
LDA	0.58	0.59	0.53	0.56
Combined	0.62	0.59	0.79	0.67
Prior six months				
Readability	0.53	0.52	0.63	0.57
Post Length	0.56	0.56	0.57	0.57
Emotionality	0.54	0.54	0.49	0.51
W2V	0.56	0.55	0.58	0.57
Subreddits	0.55	0.54	0.62	0.58
LDA	0.62	0.63	0.59	0.61
Combined	0.55	0.55	0.59	0.57
All	0.61	0.58	0.77	0.66

Table 4.6: Prediction results for binary classification of persistence in r/getdisciplined. Metrics: accuracy, precision, recall, and F1 score.

encapsulates how they approach self-improvement such as whether they think it is possible or is an insurmountable goal, which is reflected in their language use.

4.6 Discussion

The readability of a user's initial post appears highly indicative of their future engagement level. As shown previously in Section 4.4.2, persistent users tend to have lower readability in initial posts than non-persistent users. This could be because they come with the intention of engaging with the subreddit, and therefore devote more time to their introductory post hoping for a similar reaction of engagement from the forum. Post length is also a strong signal for our models both when we're considering only each user's first post as well as their prior posts on Reddit. Similar to the readability feature, one possible explanation is the higher engagement with the community through longer posts. Users having longer posts prior to joining r/getdisciplined indicates a more consistently personal style of extensive writing and

engagement, and therefore more willingness for self-disclosure.

The emotionality features provided some signal for the model, but were not as helpful as our other features. However, emotionality features derived from the 1st post resulted in higher recall than those derived from the prior six months, which could indicate that there is more expressed through emotion in the 1st post than in general text.

Prediction performance was consistently high when using word embeddings, which shows that the latent semantic information in embeddings is helpful. However, it is not significantly better than the other top features, indicating that there is room for improvement in representing more subtle linguistic information such as intent or attitude.

Topical content features derived through LDA were among the best performing features for activity from the prior 6 months, while a user's subreddit activity history was less predictive. The subreddits in which someone participates might be too coarse-grained for our task, whereas topic models can better capture the fine-grained behavior that relates to self-improvement and mindset.

Our results demonstrate how people with persistent interest in personal change act differently from those who do not maintain persistent interest. Our analyses showed that those with persistent change intent had higher prior engagement with topics that foster personal change, such as education. This kind of behavior represents a form of *gathering information* related to the intended form of change. Information gathering is an important aspect of a person's reflecting and considering their motivation for potential future change [148]. In addition to topics, we revealed differences in linguistic style between the two groups of people. Persistent users tended to have longer initial posts with lower readability.

Implications for Tailored Interventions We can use our findings and further work to tailor behavior change interventions towards people with different charac-

teristics. Those characterized with lower persistence may be in an earlier behavior change stage, necessitating a different approach than those in later stages [147]. For example, a social intervention could consist of a community moderator, or persistent community member, being paired with a likely non-persistent member (based on language use) to encourage them to stay committed to their goal [170]. Alternatively, a community-based intervention system could automatically recommend posts from persistent people, for the non-persistent people to read as a way to learn how to approach change in a healthier way [171].

4.7 Conclusion

In this chapter, we explored the behavior of users from an online community, r/getdisciplined, as a proxy for measuring persistent intent towards personal change. By analyzing user behavior prior to and immediately after joining the community, we showed quantitative differences between users who sustained intent towards general self-initiated change versus those who did not. Those who have persistent intent tended to engage more with change-oriented topics such as education even prior to expressing explicit intent to change.

We then leveraged these linguistic characteristics to build predictive models that were able to automatically distinguish people who continued engagement in r/getdisciplined and sustained their intent for self-improvement from those who did not continue, even before their first post.

Our results provide actionable insight for research areas that investigate behavior change. Understanding the underlying mechanisms associated with persistence in change can support the development of new approaches to help people change for the better.

CHAPTER V

Forecasting Donation Behavior By Surfacing Attitudes Towards Donation Interests

5.1 Introduction

Understanding and predicting user behavior from their digital traces is important for many applications, such as recommender systems [172], information filtering [173], or dialogue agents [174], as well as numerous behavioral interventions in healthcare, education, economics, and more. Prior research efforts have modeled user interests to understand or predict future behavior such as purchases [175] or click-through likelihood [176], using signals like engagement with social media content or purchase history.

Traditional approaches to user behavior prediction use machine learning models that make use of input features in a linear fashion. These models, including the more advanced neural network architectures, assume that individual data samples are provided one at a time, and are mainly independent of one another. Example user modeling approaches include using recurrent neural networks to encode the behavioral history of each user [177] or linearly aggregating different parts of a user's background and behavior, such as their demographics and online posting patterns [178]. Such approaches do not take full advantage of the relations between entities; for instance,

two products in one's purchase history may be different but still be related to one another; or two users may have interests that are seemingly diverse, but which have some degree of similarity. Richer input representations that incorporate such relations can improve the performance of downstream machine learning models used to predict user behavior.

Graph models are a prominent way of representing relational information between entities. In particular, knowledge graphs have been used widely in the context of recommender systems. For example, one can construct a knowledge graph consisting of clothing brands and items and retrieve the most relevant or similar items to recommend to a user based on their most recent clothing purchase [179, 180]. Further, interactions between users and entities can also be included in the graph, such as clicks or purchases. Such a graph and its resulting node embeddings can better capture the relations between entities that arise from the aggregate behaviors of all the users.

However, these relations still only come from explicitly observed interactions like someone clicking on one entity and then also purchasing another entity, or multiple people co-clicking or co-purchasing the same entity. In many contexts, the resulting knowledge graph is sparse, as there is an absence of many co-occurring user-entity interactions due to factors such as a very large number of entities, or users having on average a very low number of interactions. As such, the learning models applied on these sparse graphs can be lacking.

In this chapter[1], we explore user behavior prediction by using text-aware graph representations. We conduct experiments in the context of university alumni donations. We model alumni donation behavior through text and graph-based representations and evaluate our methods by predicting how likely a potential alum will donate to specific charitable funds. We conduct our experiments using the history of

[1]The work in this chapter was aided by input from Rada Mihalcea and help from Xueming Xu, Yiwei Zhang, and Ian Stewart.

donations and university engagement newsletters of a large Midwest public university.

We start by building a graph representation of alumni and associated entities, such as academic majors, university funds, and articles in engagement newsletters. Alumni actions, such as donating to a fund or clicking on an article in an engagement newsletter, are represented as edges connecting an alumni node with a fund or article node. Node embedding representations derived from this graph are thus capturing how different funds or engagement articles are related with respect to the alumni who donated to or clicked on them. We then use this graph to predict the likelihood of an alum to donate to a given charitable fund.

Specifically, this chapter makes the following two main research contributions. First, we propose a graph framework to represent and predict user behavior, and show that it improves significantly over a linear representation that does not incorporate relational information. Second, we show how this graph representation can be further enriched with implicit links drawn using semantic connections between the textual information associated with the graph entities, leading to additional performance improvements in user behavior prediction. Overall, through experiments on a large alumni donations dataset, we demonstrate the effectiveness of using graph representations enhanced with implicit information for the purpose of user behavior prediction.

5.2 Related Work

5.2.1 Combining Graphs and Text

Graph models and knowledge bases are commonly used in a wide range of tasks. However, given the nature of dealing with discrete entities and relations, they can suffer from incomplete coverage or difficulty reasoning over entity relationships.

Advancements in representation learning on graphs have proven helpful in predic-

tive tasks, such as link prediction [181], node classification [182], and node retrieval or recommendation [183, 184]. Many methods build embedding representations of graph nodes [185] derived from the graph's link structure, using adjacency matrix factorization methods [186] or random walks [187].

Work has also been done towards creating text-aware graph embedding models. Methods include representing an entity through a text embedding of the entity name [188] and jointly learning embeddings for entities and words [189, 190].

In this chapter, we leverage node embedding methods to build continuous vector representations of university alumni and charitable funds, and show that they improve over text-based representations.

5.2.2 Predicting User Behavior

Much research has focused on predicting future user behavior based on user characteristics or prior behavior. Types of predicted behavior spans a wide spectrum, including what online content someone will consume [191], what types of everyday activities someone does [192], and whether someone will persistent in personal improvement efforts [193].

In the space of charitable giving, much prior work has targeted identifying factors behind why people choose to make monetary contributions. These factors include socio-demographic and personality characteristics such as age, level of education, income, agreeableness, and empathy [194–197]. In our context of university donations, prior work has looked predicting how likely it is for an alum to donate a substantial amount of money based on their educational and professional background [198]. While this shed light on signals of individual capacity and general inclination to donate, this did not look at which specific causes donors choose to give to.

There is substantially less insight into which specific charitable causes donors are likely to choose. Studies have primarily focused on giving among one or two types

of charities, such as secular and religious causes [199], or international and national causes [200, 201]. mainly based on surveys [202] asking people to recount their recent donations and describe personal dispositions such as values [203], empathy [204], and beliefs about the cause [205]. Most such studies are limited in the number of donors, donations, and charities observed.

In this chapter, we model donor behavior and donation choices using a large dataset of donations to different causes, connected with known histories of donor interactions with engagement efforts that indicate personal interests.

5.3 University Alumni Dataset

We conduct our experiments on a dataset of alumni information maintained by a large, public university in the Midwestern region of the United States. Each alum is tied to their educational history; we primarily use their major during their highest level of study at the university. The language used in the data is English.

We focus on those who have donated any amount back to their alma mater and who have also engaged with engineering alumni online newsletters, which are typically distributed by email on a regular basis. We have 2 years of newsletter content from January 2018 to March 2020, accompanied by the interaction history of alumni. The interaction history consists of when and how many times a click occurred, as well as what article was specifically clicked in the newsletter.

Likewise, we also have a history of donations that individual alumni have made to various causes at the university. Given our focus on those who have engaged with newsletters, the corresponding history of donations for these alumni span between January 2015 to June 2020. We show statistics about entities in our dataset in Table 5.1.

Entity type	Number
Alumni	5883
Funds	1644
Articles	283
Majors	251

Table 5.1: Statistics of entities in the alumni donation dataset.

Fund Name	Fund Description
Engineering Diversity, Equity, and Inclusion Initiatives	This fund helps provide a vibrant and inclusive climate, which leverages our strengths, broadens our perspectives and paves the way for innovation.
Engineering Student Emergency Fund	This expendable fund supports the emergency needs related to the health, safety and well-being of our Engineering students, especially during the current coronavirus pandemic.
Jane Doe Dance Scholarship Fund	This endowment provides scholarship support for undergraduate dance majors.

Table 5.2: Examples of funds and descriptions.

5.3.1 Donation Funds

At this university, alumni typically donate to funds with designated purposes. For instance, the "Engineering Student Emergency Fund" supports emergency needs related to the well-being of Engineering students. They have a title and an optional textual description of the fund's purpose. Examples of funds and their descriptions are shown in Table 5.2. We see that fund descriptions can range from short and generic to lengthier and more detailed. Similarly, titles can also range in their descriptiveness of the fund's purpose.

The set of all funds span different schools and countless initiatives. In this chapter, we consider only the 1644 funds (Tab. 5.1) that have been donated to by people who have clicked on engineering alumni engagement newsletters.

5.3.2 Engagement Newsletters

The university under consideration sends online newsletters to their alumni on a regular basis. These newsletters contain university news, such as student accomplishments, novel research findings, and alumni events. They consist of links to articles with an accompanying graphic and a short summary of the article.

User actions are recorded, such as clicking on a particular article within the newsletter. Engagement with newsletter is indicative of what alumni are interested in beyond their formal studies. For instance, a computer science graduate may primarily read articles about the solar car racing team or the university's efforts to lower its carbon footprint, showing that this alum has personal interests in sustainability. This would not necessarily be apparent in their educational or employment history. Therefore, we utilize user interaction with engagement newsletters to model personal user interests. There are 283 articles in our dataset (Tab. 5.1), drawn from 49 total newsletters.

5.4 Representing Alumni and Funds

We aim to represent each alum primarily with their clicks. As seen in the previous section, every article linked within a newsletter has an accompanying short preview or summary that is displayed in the newsletter. Since this is what alumni initially see and what prompts their clicks, we use this text in our experiments, rather than the full article text.

5.4.1 Node Representation

Prior work has successfully represented entities in a graph as the average of the word vectors corresponding to its name [188]. We therefore also encode our entities using word vectors. We represent an alum as their history of newsletter article clicks,

which indicates their interests. We construct an alum embedding that is the averaged GloVe embedding of all newsletter article summaries that they have clicked on. We first compute an average GloVe embedding for each article snippet and then average over all of the article snippet embeddings to get the overall alumni embedding. Similarly, we represent a fund using the average GloVe embedding of the words in the fund's name, department, and description.

5.4.2 Graph Representation

We construct a graph to encapsulate the connections between alumni, alumni majors, funds, and newsletter articles. Each unique alumni, major, fund, and newsletter article are nodes in the graph. We include an edge between an alumni and a fund if they have donated to it, weighted by the value of the total amount of donations they've given to this fund. We also connect an alumni to a newsletter article if they have clicked on it, with the edge weighted by the number of clicks the person made. Funds included in the graph are only those associated with donations in the training set of our experiments. All newsletter clicks made by alumni are included, as was done in the text-only setting.

We then use a graph representation learning method to create embedding representations of the nodes. Specifically, we use the node2vec model proposed by [187]. We also conducted experiments using LINE [186], but found that they yielded similar results, and therefore we only show results for node2vec.

5.4.2.1 Similarity Edges

While the explicit connections between entities through actions such as clicking and donating can contain a lot of information, there can still be additional connections made with additional info. Since it's unlikely that many alumni donate and click on exactly the same funds and articles, it may be difficult to capture all relations between

Graph edge type	Number
Alum - Fund Edges	15,604
Alum - Article Edges	20,184
Alum - Major Edges	7,625
Fund - Fund Edges	72,136
Article - Article Edges	3,020

Table 5.3: Statistics of the graph derived from alumni clicks and donations, enhanced with implicit textual similarity edges.

them based on alumni behavior alone. For instance, two articles may contain very similar content but not have many overlapping clicks due to the sparseness of click data. Given the graph we have currently, the graph embedding model likely would not capture that the articles are similar based only on clicks. Similarly, two funds may be similar in their purpose and descriptions but have few overlapping donors, resulting in embeddings that do not capture their relevance to each other.

To better capture these relations among articles and funds, respectively, we propose the addition of similarity edges. The addition of the proposed edges can add these relevance connections that we know inherently exist. This can allow the graph to encode that two funds are related even in the absence of explicit evidence, such as someone donating to both funds or two people clicking on the same article and donating to the same fund.

In preliminary experiments, we found that connecting all pairs of entities weighted by similarity results in lower performance embeddings, as well as much longer training times. We suspect this is due to adding too much noise to the representation through extraneous connections.

To minimize this, we only add edges if the similarity is above a certain threshold. We also give every such edge an equal weight of 1. For every pair of articles, we compute the cosine similarity between their average GloVe embeddings and add an edge between the corresponding nodes if their similarity is above 0.7. We do the same for every pair of funds, adding an edge if the similarity is above 0.8. We choose

these thresholds empirically by looking at the distribution of similarities for all pairs of articles and funds, respectively, approximately keeping the upper quartile of similarity values. We give the numbers of different types of edges in the resulting graph in Table 5.3.

5.4.3 Analysis: Similarity between Alumni and Newsletter Articles

To gain further insights into the donor behavior graph model, we perform an analysis of the relationships between alumni and funds using their graph representations. We would expect the embeddings for alumni to be more similar to the embeddings of funds that they are more likely to donate to. This graph could then be used for querying for relevant entities. For instance, we could find the top funds that may be of interest to an alum.

To examine this, we compute the cosine similarity between pairs of alumni and funds where the alum has donated to the fund, and compare with pairs where the alum did not donate to the fund. We use node2vec embeddings based on the graph that has all similar edges incorporated. Further, we ensure that the model is not simply remembering known donations in this analysis by focusing on the subset of donations that occur in 2020 and *removing links between alumni and funds corresponding to these donations* from the graph, no matter which year the donation was made during. This way, we are looking at similarity of alumni and funds that are known to be related, but *that the model does not explicitly have knowledge about*; their similarity therefore comes solely from other alumni behavior and semantic connections. We show the distribution of similarities in Figure 5.1. Using a two-sided T-test, we calculate the statistical significance between the donation and non-donation samples of similarity values; we designate those with a significance level of $p < 0.1$.

Notably, we see that the GloVe-based similarities do not distinguish well between alum-fund pairs where a donation occurred and where a donation did not occur. In

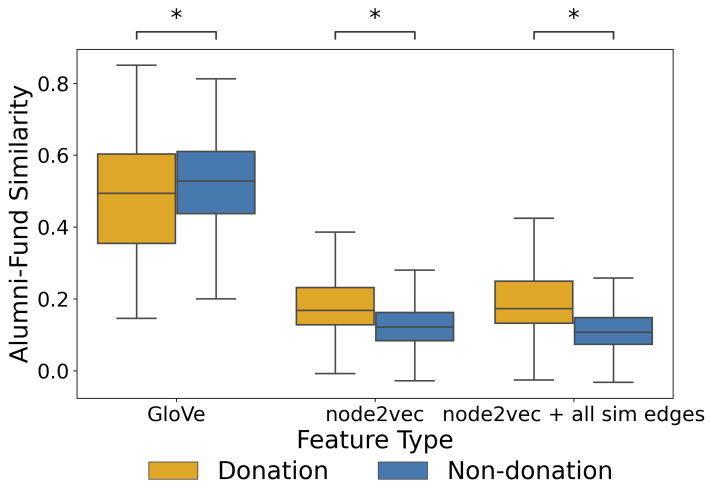

Figure 5.1: Distribution of similarities between pairs of alumni and funds where alumni have either donated to the fund or not. We show distributions of embedding cosine similarity based on text-only GloVe features and node2vec graph features with and without the addition of similarity edges. Statistical significance is determined using a two-sided T-test, and designated with a star (*) if $p < 0.1$.

fact, the non-donation pairs actually have higher similarity than the donation pairs. This implies that it is not sufficient to use only textual semantic similarity between alumni and funds for determining donation interest.

On the other hand, we see significantly higher similarities between alumni and funds that they have donated to than between negative samples of alumni and funds when using graph embeddings. Further, this is more pronounced when similarity edges are included in the graph, yielding greater separation between pairs who have donated and those who have not. This shows that the graph embeddings are indeed encoding alumni behavior and interest.

5.5 Predicting User Behavior

We have seen that the alumni behavior graph model encapsulates relationships between entities in the resulting embedding space. We evaluate the alumni behavior graph model for downstream predictive use in the context of donation prediction. We

construct a task where we predict whether an alumni is likely to donate to a particular fund, showing that we can distinguish *which funds* someone is likely to donate to.

5.5.1 Experimental Setup

We focus on alumni who have both clicked on newsletter articles and have made donations before. This set of alumni, along with the funds that they have donated to, their majors, and newsletter articles they have clicked on, is the dataset we use for our experiments. We look at all pairs of alumni and particular funds they've donated to as data samples. Donations made prior to the beginning of 2020 are considered training data for our predictive models and donations made in 2020 are test data. Splitting our data by time reflects the real task that universities face, where we know an alumni's history and want to predict their future donation behavior.

Funds that do not appear prior to 2020 are not included, as our graph representation models are based solely on the training data and would not be able to produce a representation for a previously unseen entity. Similarly, alumni who only appear in 2020 would be excluded from the experiments as they have no prior history and therefore would have no corresponding representation features.

We then use negative sampling to construct sample pairs of an alum and a fund where the alum has not donated to the fund. For both the training and test sets, we include an equal number of such negative samples to obtain a balanced dataset. To construct a negative sample, we randomly select an alumni and a fund out of all alumni and funds, respectively, that we are considering in our dataset. Then, we check if the alum-fund pair appears as a positive sample in the corresponding data split and keep the pair if it does not appear. We continue constructing negative samples until we have the same number of positive and negative data samples.

Donation prediction with unique alum-fund pairs. We also conduct experiments in a modified setting where we predict the donation interests of alumni without

Alum-Fund Pairs	# Train	# Test
Complete	19,882	3,236
Unique	18,888	3,058

Table 5.4: Number of samples in the training and test sets of our task. The training samples are donations that were made prior to 2020. The test samples are donations made in 2020.

knowledge of their past donations to the same funds they've donated to in 2020. We remove all alum-fund pairs from the training set that occur in the test set, which corresponds to removing past donations that are identical to ones in 2020. Other prior donations that alumni have made are kept. This is a more difficult task, as prior donations to a fund can be highly indicative of future donations to the same fund. Therefore, we must rely more on alumni background and the implicit relationships between different funds as well as between newsletter articles and funds.

5.5.2 Classification

We train a logistic regression classifier to predict whether an alumni has donated to a given fund in 2020, based on the data described previously.

There are funds that receive thousands of donations while others receive far fewer individual donations. This can be due to the fund being very general, such as a general scholarship fund, or a popular interest, such as a sports-related fund. On the other hand, funds with more specific or niche subjects may receive fewer donations. Such large data imbalances can lead predictive models to simply memorize the most frequently occurring funds, rather than using the embedded features to make more complex connections between alumni and funds. We empirically find that less than 1% of the funds we consider have received over 200 donations. Therefore, we downsample the number of unique donations each fund has to 200 samples.

Features	Complete Donation Pairs Alumni	Fund	Both	Unique Donation Pairs Alumni	Fund	Both
Text-only	0.516	0.782	*0.784*	0.500	*0.799*	0.798
Graph representations	0.574	0.781	*0.812*	0.532	*0.791*	0.778
+ article sim edges	0.574	0.774	*0.817*	0.543	*0.789*	0.778
+ fund sim edges	0.575	0.804	*0.846*	0.541	0.816	*0.824*
+ article and fund sim edges	0.564	0.798	*0.841*	0.533	0.816	*0.830*
All (GloVe + node2vec w/ all edges)	0.569	0.824	***0.856***	0.537	0.848	***0.855***

Table 5.5: Results from the donation behavior prediction task. Left: Training set contains the complete prior donation history of alumni in test set. Right: Donations made in 2020, in the test set, are removed from the training set. Italicized values designate the highest performance for a given feature type and experimental setting. Bold values designate the highest performance in the experimental setting overall.

5.6 Results

We compare the use of text-only GloVe features and graph-based node2vec features in our experiments to evaluate the benefit of our alumni behavior graph model. Further, we evaluate our graph representations both when enhanced with text similarity-based edges and without to show the effects of this adding this implicit information to the graph. We show our alumni donation interest prediction results in Table 5.5.

In the results, we see that the graph embedding features generally perform better than the text-only features. This is in line with our hypothesis, since the text only contains information about the semantic content, but nothing about how it is related to any other entities. Further, such relations would be difficult for the machine learning model to pick up through the prediction task, as alumni generally do not individually donate to many funds and there is likely little overlap between different people. This sparsity of connections are typical in many recommendation systems contexts. Our framework of encoding user behavior into a graph could therefore be applied to other types of downstream tasks that aim to predict future behavior.

We see that adding implicit edges derived from the textual content of the funds and

articles generally improves performance over only having explicit action edges that designate donations and clicks. Similarity links between articles are more helpful when we have knowledge of an alumni's entire prior donation history.

Next, we examine how information from alumni and funds respectively contribute to the overall performance by conducting the donation interest prediction task using only alumni features and only fund features. We compare this to the use of both alumni and fund features.

We see that prediction accuracy based only on using the alumni features is lower compared to using other features, as one would expect, as we have no explicit information provided about the fund being donated to. However, alumni-only features perform better when represented as graph embeddings as opposed to text embeddings. The alumni nodes are connected to the funds and therefore fund information could be embedded within alumni nodes as well, so alumni representations could become associated with larger themes or topics that the classifier then picks up on.

Interestingly, accuracy based on using only fund features is much higher than random, showing that the model is learning trends in which types of funds, in terms of content and theme, are generally more well-received. We know the classifier isn't simply picking up on specific popular funds, since we downsampled frequently occurring funds.

When we use both features from alumni and funds, we generally see better performance, especially when using graph features and with fund edges added. This shows that the prediction model is learning how the alumni and funds are related.

When we use only unique donation pairs, we see that the results remain largely comparable with using complete donation pairs. However, the performance is lower than with the use of complete donation pairs when using only features derived from alumni, showing that the complete donation pairs prediction model learned more about donation trends of specific alumni whereas the unique donation pairs model

has to understand more of the implicit relatedness between funds and articles.

Finally, we see that combining GloVe features with node2vec features yields the highest performance. This implies that there is still use in having both the semantic content of the entities and their relational information, and that they are complementary to each other.

Prior Donations	Top 3 Similar Funds (Similarity Score)
Engineering General Scholarship Fund Professorship in Rheumatology	Professorship in Gastroenterology and Hepatology Fund (0.40) Gastroenterology Nurse Education Fund (0.36) Gastroenterology Education and Research Fund (0.32)
Aerospace Engineering Support Aerospace Engineering Centennial Fund	Aerospace Engineering Junior Faculty Support Fund (0.47) Aerospace Graduate Research Excellence Fellowship (0.42) Aerospace Graduate Teaching Award and Scholarship (0.38)
Iconic Mastodons Movement Fund Majungasaurus Exhibit Fund	Mammoth Museum Exhibit Fund (0.44) Museum of Natural History Discretionary Fund (0.42) Museum of Natural History Membership (0.39)

Table 5.6: Prior donations made by a given alum the top 3 most similar funds with respect to the alum, determined by embedding cosine similarity. To preserve anonymity, we remove all names and specific details from fund titles. Text of the fund descriptions are not shown for brevity.

Qualitative Analysis. For a qualitative analysis, we use the node2vec model that includes all similarity edges, built from the training data with unique donations. We analyze how the model is able to retrieve relevant alumni and funds for a given alum.

Retrieving relevant funds. In Table 5.6, we show examples of funds that alumni have previously donated to and the funds that the model determined to have the highest cosine similarity. In the first example, the model retrieves funds that are related to the medical field and supporting research and education in the fields,

Alum's Prior Donations and Clicks	Nearest Alum's Donations and Clicks
F: Engineering General Scholarship Fund F: Mechanical Engineering Special Gifts Fund A: A high altitude long endurance aircraft	F: Engineering General Scholarship Fund F: Mechanical Engineering Special Gifts Fund A: Second place finish for the solar car team A: 3D printing 100 times faster with light
F: Engineering Entrepreneurship Fund F: Engineering Faculty Scholar Award A: Autonomous car preventing traffic jams A: Nobel Prize nomination for powerful laser pulse	F: Engineering Dean's Discretionary Fund A: Driverless future A: Solar car test A: Smart wearables improving elderly mobility

Table 5.7: Examples of the most similar alum for a given alum. To preserve anonymity, we do not show names and remove all identifying information within fund descriptions and article titles. We show the donations and clicks made by the alumni. F - Fund; A - Article

which matches well with the alum's actual prior donations to funds supporting student scholarships and an endowed professorship. The second and third examples similarly show that the given alum's previous donations and most similar funds share common themes of aerospace engineering and natural history, respectively.

Retrieving relevant alumni. In Table 5.7, we show examples of click and donation activities of alumni and their alumni neighbors that the model determined to have highest cosine similarity. In the first example, the chosen alum's donations and clicks are related to mechanical engineering. The most similar alum has also donated to mechanical engineering funds and clicked on mechanical engineering-related articles, which shows that nearest alumni neighbors' interests and behaviors match well with the chosen alumni. Likewise, the alum in the second example and their most similar alum both share interest in autonomous vehicles and research advancements.

5.7 Conclusion

In this chapter, we explored the use of text-aware graph representations for user behavior prediction. Using a large dataset consisting of university alumni donations and their interests as expressed through click-throughs on a university newsletter, we showed that the the use of a graph framework to explicitly encode the relations between user behaviors and user interests leads to significant improvements over simple linear representations. Moreover, we showed how further improvements can be obtained by enhancing the graph with implicit links inferred based on the semantic distance between the textual data associated with the entities in the graph. Our results demonstrate the role played by graph representations using explicit and implicit relations for the prediction of user behavior.

CHAPTER VI

Quantifying Community Subjective Wellbeing and Resilient Attitude

6.1 Introduction

The COVID-19 pandemic has had widespread effects on people's subjective well-being. However, the local surrounding environment can greatly influence and be indicative of how people cope with an adverse event and shifting conditions. For instance, stronger social ties have been associated with higher wellbeing and community resilience[206, 207]. The aspects of life that a community gives attention to, such as leisure, family, and friends, can also be indicative of how that community may fare when impacted by a negative event.

Community resilience is the ability of a community to adapt to changing conditions and to withstand and recover quickly from disruptions [208, 209]. With the COVID-19 global pandemic, boosting and maintaining mental wellbeing has become a prominent issue as everyone continues to grapple with the ongoing daily life restrictions and overall uncertainty of the situation. In this chapter[1], we aim to understand the subjective affective wellbeing recovery patterns of communities in cities across the United States (US) and gain insight on relationships between cities' characteristics

[1]The work in this chapter was aided by input from Rada Mihalcea and help from Sophia Sun and Laura Biester.

and the pandemic's effect on the cities' subjective wellbeing over time.

Many in-person interactions migrated online during the pandemic. Online forums such as Reddit offered a way for locals to stay connected and stay current with ongoing concerns such as whether certain restaurants were open or the status of vaccinations in their area. Reddit's city-focused subreddits, such as r/seattle and r/annarbor, correspond to cities across all states in the US. Discussions on these subreddits reflect people's concerns. While surveys could help measure wellbeing, they are often limited in scale. By looking at everyday conversational behavior in a community, we can get an aggregate picture of wellbeing.

In our work, we characterize trends in how community subjective wellbeing shifted during the beginning of the COVID-19 pandemic (until the end of 2020) across 112 cities spread across the US. Using cities with similar wellbeing recovery patterns, we quantify what community characteristics correlate with lessened impact on wellbeing from the pandemic, as well as recovery speed given a negative impact. The features we examine are derived from online user interaction patterns and linguistic content.

We seek to answer the following research questions:

1. **How has the pandemic impacted the affective wellbeing of US cities?**

   We quantify wellbeing using positive and negative affect expressed in daily discussions on city-focused subreddits. We then measure the pandemic's impact on a city by comparing a city's observed wellbeing with its expected wellbeing, as forecasted by time series models derived from data prior to the pandemic. We define three patterns of wellbeing seen during 2020.

2. **What distinguishes city communities that are more or less impacted by the pandemic?**

   We analyze a set of community traits derived from prior to the pandemic, encompassing linguistic features and user interaction patterns. We predict

whether a city's wellbeing is heavily impacted by the onset of pandemic and analyze differences among cities that were more impacted and those that were not.

3. **How do impacted city communities differ in their speed of recovery?**

   Using the same community features, we predict whether an impacted city makes a recovery within the year, similarly analyzing distinguishing characteristics.

6.2 Related Work

Subjective Wellbeing. Research on subjective wellbeing (SWB) delves into how people feel and think about their lives [210]. Subjective wellbeing is not a single concrete entity and studies look largely at two components, affective and cognitive wellbeing [211]. Affective wellbeing is defined as the positive and negative emotions that people feel, such as happiness and anxiety. More positive affect and less negative affect indicates higher affective wellbeing. On the other hand, cognitive wellbeing is defined as one's evaluation of one's life and resulting level of life satisfaction. This can refer to overall satisfaction, or satisfaction with respect to specific life domains such as jobs or relationships. These are two distinct constructs, and differ in many ways, such as their stability over time and how they are impacted by life events [212]. Though affective wellbeing tends to return to baseline levels through hedonic adaptation [213, 214], the adaptation rate can greatly differ for different people and circumstances [215, 216].

In our work, we present a longitudinal study of affective wellbeing and adaptation rates in different cities across the US in response to the COVID-19 pandemic, derived from large-scale naturally occurring social media data.

Community Resilience and Social Capital. A community can be viewed as as a group of people who are bound by some common tie such as geographic

location; they often have shared social norms [217], values, and other characteristics. Constituent parts of a community can influence one another in complex ways. The study of community resilience investigates the qualities that allow a community to cope with and adapt to a collective disaster experience [218], in terms of physical and mental health outcomes.

A core underlying concept is social capital [219, 220]. In the context of community resilience, social capital consists of the weak and strong social ties and networks in a community. This can come from social support, like family and friends, as well as social participation in the broader community. Improved connections can help provide support to members of the community.

In our work, we study the online counterparts of cities across the United States and how signals of social capital may influence their community wellbeing and resilience within the context of the COVID-19 pandemic.

Studying Online Communities. Online communities have been the focus of research studying user behavior and community features. For instance, prior work has examined behavioral and linguistic factors that drive social support [221–224] and signal community success [2, 225]. A parallel line of work uses online communities as a lens to study the linguistic manifestation of mental health symptoms [226–229]. Work in this area [e.g., 9, 230–232] frequently makes use of Linguistic Inquiry and Word Count (LIWC) [233], a tool that we use to measure city's affective wellbeing. With the onset of the COVID-19 pandemic, many have turned towards identifying the effects of the pandemic on mental health through the analysis of social media [234, 235].

We adapt some of these characterizations of online communities and use them to distinguish between cities with different patterns of subjective wellbeing shifts during the COVID-19 pandemic. We measure community-level subjective wellbeing using social media, and draw insights from longer-term patterns of subjective wellbeing.

	Avg	Min	Max
Authors	21,717	1,799	116,338
Submissions	17,104	1,506	80,231
Comments	349,756	9,021	2,490,517

Table 6.1: Reddit city communities dataset statistics. Total values are computed per subreddit from all of 2017 - 2020, which are then aggregated over all subreddits. The corresponding cities cover 48 US states.

6.3 Reddit City Communities

With the onset of the COVID-19 pandemic and corresponding stay-at-home measures, most of the socializing that normally happened in person shifted online. In our study, we focus on data collected from Reddit, a community-based online forum for information sharing and discussion. Sub-communities on Reddit, known as subreddits, can be centered on any interest. Some subbreddits have the intention of connecting people in a city and allowing them to discuss local news, events, attractions, and more. These allow us to study naturally occurring online counterparts to physical cities; from manual examination, we found that these communities appear to be largely composed of local residents who discuss city-related topics.

During the COVID-19 pandemic, these subreddits became an important way for people to not only get local information, but also to express their worries with to who might best understand, and to find support. Many posts are laden with emotion, such as fear or outrage towards policies being implemented, sadness when talking about a favorite business closing permanently, or loneliness when venting about social isolation and being unable to see friends and family.

Following prior work on subjective wellbeing [236, 237], we examine the levels of positive and negative emotions expressed in language by members of the communities to gauge affective wellbeing in aggregate at the community level. We look at the posts and comments in these city subreddits to observe how affective wellbeing of communities in different US cities was affected over time by the COVID-19 pandemic.

We select the top five most populous cities from every US state, as well as Washington D.C., and manually find the corresponding subreddits. Not all cities have subreddits, yielding 233 cities. For the cities with multiple subreddits, we choose the subreddit with the most members. We also manually label each city with its corresponding county.

We collect all publicly available Reddit posts and comments from the 233 city subreddits starting from January 1st, 2017 to December 31st, 2020.[2] We collect posts made prior to the pandemic as a way to compare with normal community behavior before the pandemic. Additionally, the period we study precedes widespread vaccine availability, allowing us to see how communities dealt with the pandemic when the main solutions involved limiting social interaction.

To ensure that the cities have sufficiently active subreddits, we only use subreddits that have posts on at least 300 unique days in each of the years considered, resulting in a final list of 112 cities that cover 45 US states. We give statistics about the final dataset in Table 6.1, and a map of the cities is presented in Figure 6.1.

6.4 Quantifying Subjective Wellbeing and Resilience

People can perceive the world in drastically different ways, even when they are experiencing similar events. How someone reacts to a traumatic or important situation can say a lot about how they are coping with the situation. In particular, someone's emotional response is a crucial part of their reaction. We therefore focus on emotional expression as a way of quantifying mental wellbeing, which we define as a balance between the presence of positive emotion and a lack of negative emotion.

To compute wellbeing, we examine all posts of the members of a community and count the number of occurrences of words from the PosEmo and NegEmo categories of the Linguistic Inquiry and Word Count (LIWC2015) lexicon [233]. These

[2]We used the Pushshift.io API.

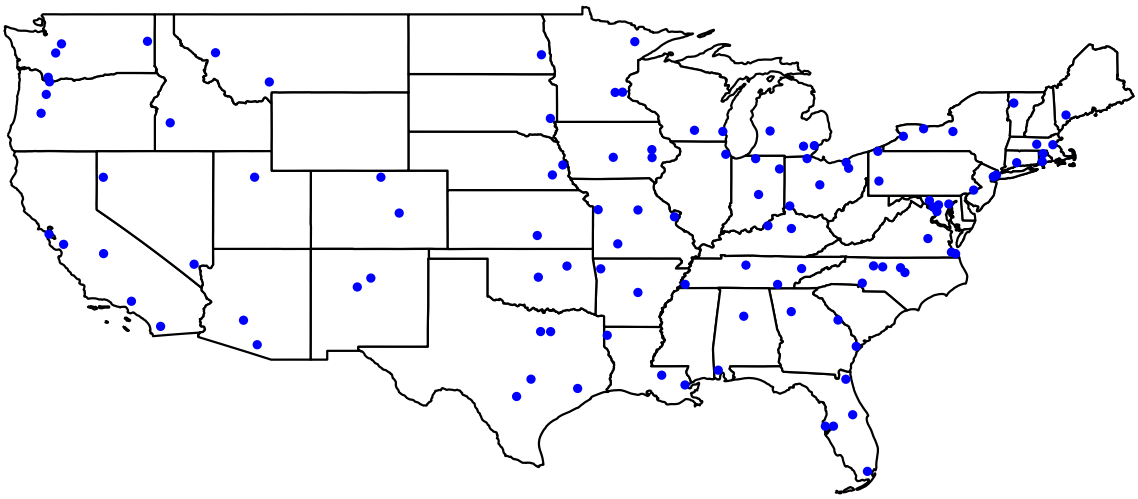

Figure 6.1: A map of the cities included in our city subreddit list, which are spread throughout the United States. Anchorage and Honolulu are also included in our city subreddits.

correspond to positive emotion and negative emotion respectively. More formally, we define our metric for WellBeing as:

$$\textsc{WellBeing} = \textsc{PosEmo}_{norm} - \textsc{NegEmo}_{norm}$$

where $\textsc{PosEmo}_{norm}$ and $\textsc{NegEmo}_{norm}$ represent values after applying z-normalization to the average raw LIWC values for each day.

6.4.1 Measuring Resilience

Community resilience is the ability of a community's ability to adapt to change, and to withstand and recover quickly from disruptions. From prior studies in psychology, we have seen that people tend to return to a baseline level of subjective wellbeing after life disruptions, even when the adverse situation persists [216]; if a community recovers and adapts more *quickly*, then they are more *resilient*.

Based on this definition of resilience, we track the trend of WellBeing scores during the pandemic. This trend can show how well the cities are coping with the

pandemic, and if the cities recover (which we define as reaching expected WellBeing scores predicted before the pandemic), it indicates strong resilience. On the other hand, if a city's WellBeing decreases and does not recover for a long period, the city is likely having a harder time coping with the pandemic.

We begin by building a time series of WellBeing values for each city subreddit. We first compute the average daily WellBeing scores. Next, we fill in missing values using linear interpolation, and take a 7-day rolling mean for each day to smooth out weekly fluctuations and outliers.

We model the expected WellBeing of a city's subreddit using the Prophet model [238] trained on data prior to March 2020 (Jan. 1, 2017 - Feb. 29, 2020); this model has been used in prior work on the impacts of COVID-19 on social media forums [235, 239]. The model fits the equation

$$y(t) = g(t) + s(t) + \epsilon_t \quad [6.1]$$

$g(t)$ represents a piecewise linear model that is used to represent the trend, while a Fourier series $s(t)$ is used to approximate the yearly seasonality. The error is reprsented by the term ϵ_t. We use the Prophet model trained on pre-COVID data to forecast post-COVID values through the end of 2020 (Mar. 1, 2020 - Dec. 31, 2020), along with the 95% confidence interval.

We consider a city's WellBeing to have significantly deviated below our expectations if the values for at least 25% of the days fall below the 95% confidence interval. We consider the early stage of the pandemic to be April 1 - June 30, 2020, and the middle stage of the pandemic to be July 31 - December 31, 2020. Based on these intervals, we define three classes to represent a city's resilience:

Unaffected: In the early stage of the pandemic, the city's WellBeing does not significantly deviate below the expected values.

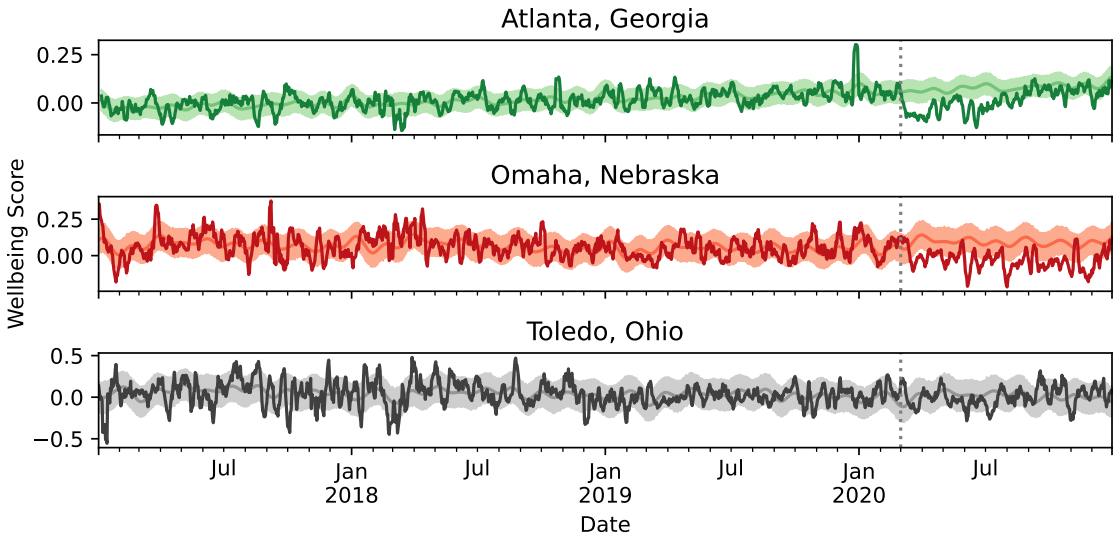

Figure 6.2: Average daily WellBeing scores from a selection cities over time. Atlanta, Georgia *recovered*, Omaha, Nebraska *did not recover*, and Toledo, Ohio was *unaffected*. The lighter line is the Prophet forecast, the shaded area is the 95% prediction interval, and the darker line is the true value. The dotted line marks March 1st, 2020.

Recovered: In the early stage of the pandemic, the city's WellBeing significantly deviates below the expected values, but during at least one of the three-month periods in the middle stage, it no longer deviates.

Non-Recovered: In the early stage of the pandemic, the city's WellBeing significantly deviates below the expected values, and it continues to deviate during each of the three-month periods in the middle stage.

The number of cities matching each recovery pattern are shown in Table 6.2. Figure 6.2 shows plots of the daily WellBeing values for a selection of cities, which are used to determine their recovery patterns. In our experiments, we focus on two distinctions: (1) the distinction between cities that are affected by the pandemic (with respect to WellBeing score), those that are not; and (2) the distinction between cities that recovered and those that did not.

Recovery Pattern	Number of Cities
Unaffected	36
Recovered	32
Non-recovered	44
All	*112*

Table 6.2: Number of cities that fall into each recovery pattern.

6.4.2 Comparing with Traditional Resilience Metrics

Community resilience, or the ability of a community to cope with a crisis, is a prominent topic in disaster and policy research [220]. Many have worked to define and quantify aspects of resilience, such as social capital and economic prosperity. Resilience metrics have been assigned to communities by analyzing empirical data about the community.

One such well-known metric is Baseline Resilience Indicators for Communities (BRIC) [219]. BRIC measures the inherent resilience of counties in the US with respect to six different domains: social, economic, housing and infrastructure, institutional, community, and environmental. The metric covers over 60 different variables, such as mental health support, educational attainment, and employment rate. Resilience scores are assigned to each county in the US based on these measures.

We measure the Spearman correlation between our three classes of social media-derived resilience labels with the different components of the real-valued BRIC resilience scores. Each city is paired with its county's BRIC score; a county and score can be linked to multiple cities. The results are shown in Table 6.3.

We see that there is correlation between our resilience labels and BRIC scores in a number of domains. Notably, we see high correlation with social resilience, as well as correlation with environmental and aggregate resilience. This demonstrates that our labels reflect existing measures of community resilience.

BRIC domain	p-val
Social	**0.004*** **
Economic	0.680
Housing and Infra	0.554
Institutional	0.142
Community	0.664
Environmental	**0.018** **
Aggregate	**0.070***

Table 6.3: Spearman correlation between the social media-derived resilience labels with the components of the BRIC resilience scores. Statistically significant values are bolded.
$* * * : p < 0.01$, $** : p < 0.05$, $* : p < 0.10$

6.5 City Community Features

In order to build a model to predict whether a city's wellbeing will be affected and recover during COVID-19, we use a number of features, include demographic features (Section 6.5.1), linguistic features (Section 6.5.2), and interaction features (Section 6.5.3).

6.5.1 Demographic Features

The demographic attributes of cities can be indicators of community resilience, as indicated by BRIC [219, 240]. We examine a number of community demographic dimensions that may be related to city WellBeing patterns during the pandemic.

Population Density. Denser populations may provide more opportunities for people to aggregate and do things together, including unplanned interactions, therefore potentially facilitating the buildup of more social connections and social capital [241].

Age. Age is correlated with subjective wellbeing. For instance, in the US, subjective wellbeing tends to be lower during middle age and higher in younger and older adults [242].

Rent vs Own. Areas where people own homes may be more invested in and

connected to the surrounding community, resulting in stronger social bonds.

Household Income. Money can't buy happiness, but it can buy life necessities and provide stability. Higher income has been shown to be correlated with subjective wellbeing [243].

Housing Cost. Housing cost can be related to the overall quality of life in an area, although economic stress from housing cost could hamper subjective wellbeing.

Latitude. Climate and weather can influence subjective wellbeing; lower anxiety has been linked with higher temperatures and more hours of sunshine [244]. Higher-latitude regions, or areas further from the equator, tend to have colder temperatures and receive less light than lower-latitude areas.

Our demographic features are summarized in Table 6.4. We use data from the U.S. Census Bureau[3] and the National Weather Service.[4] All the data is collected based on each city's corresponding Federal Information Processing Standard (FIPS) code.

6.5.2 Linguistic Content Features

Personal Concerns. The sharing of personal concerns and details about personal life can garner social support online [245]. People may discuss a wide range of personal attributes, such as their family, career, and hobbies. To quantify people's discussion of personal concerns, we use the LIWC lexicon [233] and consider the following categories: *family*, *friend*, *work*, *leisure*, *home*, and *health*.

Group Belonging. When people talk frequently about their affiliations, such as work associates, romantic partners, or neighbors, it can indicate a sense of community and presence of social ties. Social ties have been shown to be beneficial for wellbeing [246]. Similarly, people can also indicate a sense a group membership through the use of pronouns like "we" or "our." We therefore consider the *affiliation* and *we*

[3]https://www.census.gov/data/developers/data-sets/acs-1year.html
[4]https://www.weather.gov/gis/Counties

Feature Name	Description
Population Density	Number of people per square mile of land area
Median Age	Median age of residents of the city
Rent vs. Own	The ratio of number of people who rent a property to those who own a property
Median Household Income	Median household income in the past 12 months
Median Monthly Housing Cost	Median monthly housing cost
Latitude	The latitude of the county where a city is located

Table 6.4: Summary of city demographic features. All values are derived from 2019, prior to the pandemic.

categories from the LIWC lexicon.

Time Orientation. One's time orientation is the emphases that one places on each of three relative time periods: the past, present, and future [247]. This can be indicative of one's subjective well-being and mental health. For instance, people who focus on the past in a positive way can find joy in their memories [248]. We consider three measures of time perspective and orientation: *past focus*, *future focus*, and *present focus*. All three are computed using the corresponding categories in LIWC: Focus Past, Focus Future, and Focus Present. The Focus Past category primarily consists of past-tense verbs, which also reveal how much people share about their activities.

In our experiments, we derive a set of values for each city corresponding to how prevalent each of these categories were in their subreddits during normal times prior to the pandemic. For each of the lexicon features, we compute the percentage of the category present in each post and average over all posts in 2019.

6.5.3 Interaction Features

We hypothesize that the way in which people interact with others on their subreddit may be indicative of the community strength and resilience in the face of a disaster event such as the COVID pandemic. In order to measure interactions, we compute *user interaction features*, which represent how individual users interact with one-another; and *post interaction features*, which represent the interactions with user's *posts*. These metrics expand on existing metrics [235, 249–252].

User Interaction Features. The user interaction features are computed based on a graph representing daily interactions between users, where an interaction occurs when user B comments on user A's post or comment. Doing so introduces an edge from B to A and vice-versa, as we use an undirected graph. For each day, we consider the posts and comments that occurred on that day and create edges to their parents

(regardless of when they occurred). Next, we compute the graph metrics that are shown in Table 6.5 for each graph, using the NetworkX package [253]. When computing our user interaction features for classification, we compute the average over all of the days where there is activity in the subreddit for each metric.

The metrics represent many facets of the community structure; node count and edge count reflect the daily activity on the subreddit. Density represents how connected each member of the community is to all other members; mean eccentricity (the maximum distance of each vertex from any other other vertex) represents a similar phenomena. Connected component count represents the number of distinct sub-groups in the community who do not interact on a given day, while mean connected component counts represent the size of those subgroups. Mean shortest path (across all components) and diameter represent the distance between nodes that are directly or indirectly connected.

Post Interaction Features To measure how users interact with posts in each subreddit, we first form a tree for each post that represents the chain of comments created in response to it. In each tree, the post is the root node and each comment is a child of the post or comment it replied to. For each post, we compute five measures shown in Table 6.6, some of which are attributes of its corresponding tree and others which the timing of its comments. The measures are inspired by prior work on social media interaction [254, 255]. The measures are averaged across all posts in the subreddit made during 2019 to compute our final features. We began with a longer list of features and removed many features that were highly correlated with our final set.

The first two metrics, Tree Size and Direct Reply Count are representative of the number of comments a post receives. Leaf Node Count represents the number of comments that are left without a reply, while Max Level Width represents the number of comments at the largest level of the tree. The final metric represents

how long on average it takes for each post to get a comment. When a metric would otherwise be undefined, we only consider posts that have comments.

6.6 Predicting Cities' Pandemic Impact

A community's normal behavior is predictive of how they may cope with an adverse event. For instance, communities with stronger social ties may be more supportive of community members during a disaster, leading to higher subjective wellbeing. Similarly, the general disposition of community members, such as being future oriented or valuing leisure activities, may also indicate coping patterns.

We examine community interaction and linguistic characteristics of city subreddits during normal times prior to the pandemic, in 2019, as previously described in Section 6.5. Using these features, we build models to predict whether a city's subjective wellbeing will be significantly negatively impacted by the onset of the COVID-19 pandemic. We distinguish between cities that are unaffected and all those that are affected, as exhibited in the first few months following the start of the pandemic (Section 6.4).

We train a logistic regression classifier with leave-one-out cross validation (LOO-CV) across all the cities. The features are scaled such that they all take on values from 0 to 1. Given the greater prevalence of affected cities and limited number of city data samples, we balance our training data by oversampling the minority class of unaffected cities using SMOTE, which synthesizes new minority instances by interpolating between existing minority data samples [256].

6.6.1 Results and Discussion

We show our results for distinguishing affected versus unaffected cities in Table 6.7. All metrics other than accuracy are computed using macro averaging.

Feature Name	Description
Node Count $\|N\|$	Number of unique users who posted or commented
Edge Count $\|E\|$	Number of unique users who interacted through a reply to a post or comment
Mean Degree	Mean number of edges per node
Density $\frac{2\|E\|}{\|N\|(\|N\|-1)}$	Number of edges in graph over number of possible edges
Connected Component Count	Number of subgraphs in which all pairs of nodes are connected by an edge
Mean Eccentricity	Mean of eccentricity across nodes; eccentricity is the maximum distance between node n and any other node
Mean Connected Component Size	Mean number of nodes in a connected component
Mean Shortest Path	Mean distance between each pair of vertices that are connected by a path
Diameter	Maximum distance between any pair of nodes within a connected component

Table 6.5: Summary of user interaction features. All values are derived from 2019, prior to the pandemic.

Feature Name	Description
Tree Size	Number of nodes in tree
Direct Reply Count	Number of children of the head node
Leaf Node Count	Number of leaf nodes in the tree
Max Level Width	Number of nodes in the largest level of the tree
Min Response Time	Time between creation of original post and the first comment it received

Table 6.6: Summary of post interaction features. All values are derived from 2019, prior to the pandemic.

Features	Acc	P	R	F1	AUC
Random	0.500	—	—	—	—
Demographics	0.607	0.601	0.615	0.591	0.693
User Interaction	0.741	0.718	0.743	0.722	0.801
Post Interaction	0.661	0.638	0.655	0.638	0.698
LIWC	0.688	0.663	0.682	0.665	0.716
All	**0.750**	**0.729**	**0.757**	**0.733**	**0.805**

Table 6.7: Unaffected vs. Affected classification results.

We see that all of our community interaction and linguistic feature sets exceed the random baseline of 50%. Further, user interaction features perform the best.

Additionally, we examine how different features correlate with community resilience by analyzing the coefficients of our regression model that is trained using all of the features together. We use the mean coefficient values across all folds in the LOO-CV. The results are listed in Table 6.8.

Demographic Features. Unaffected cities are associated with higher Latitude. Though higher latitudes have colder weather which has previously been correlated with lower wellbeing in general, people living there may also be more resilient

since they are used to dealing with some amount of discomfort throughout much of the year. These cities also are associated with higher Population Density, and therefore more people and social ties. They also are likely to have a higher Rent vs Own ratio, and higher Median Monthly Housing Cost, which likely is tied to the higher population density, as most people in higher density cities rent and also have a higher cost of living. Finally, these cities are associated with a higher Median Age, which may indicate more stability and long-term occupants who are invested in the community.

Affected cities are associated with higher household income, which is counterintuitive as higher income often correlates with higher subjective wellbeing. However, since we are looking at resilience to wellbeing impact and not absolute wellbeing, it may be that areas with higher income are less accustomed to large negative events and their affective wellbeing was more impacted.

User Interaction Features. Subreddits where WellBeing was not highly impacted by the pandemic are associated with higher density and edge count, meaning that their members tend to interact more with others.

Affected subreddits are associated with more separate groups of connected users (Connected Component Count), meaning that individuals are not interacting as much with the larger community, and that interactions are more localized. These subreddits are also associated with being less connected, with higher Eccentricity, longer Mean Shortest Path, and larger Diameter.

Post Interaction Features. Cities whose WellBeing is not impacted by COVID-19 tend to have posts with more interaction, as shown by correlation with greater Tree Size, Direct Reply Count, Max Level Width, and Leaf Node Count. On the other hand, affected cities are correlated with higher Min Response Time.

This indicates that subreddits with *slower* response times and *less* interaction

end up with greater WellBeing impact during the first three months of the COVID pandemic.

Linguistic Features. Unaffected cities are correlated with the Focus Past, Focus Present, and Home LIWC categories. The Focus Past and Focus Present categories are largely verbs in the present and past tense, and therefore this may indicate that people are more willing to share about the activities they do, and details about their lives overall. People talking about their home more, prior to the pandemic, may indicate that their home environment is an important part of their normal life. Therefore, when the pandemic hit and people were largely confined to their homes, this may have not been as significant of a negative shock.

We see that affected cities are more likely to talk about their friends, family, and other affiliations. Though many of these would generally be positive social factors in a disaster, the nature of the social policies put in place may have turned these into negative factors, since social interaction was specifically restricted.

We also see that the affected cities are more likely to focus on the future. Similarly, while this might normally mean hopefulness towards the future, it also may mean result in greater wellbeing impact from the pandemic, since the state of the world was under such great uncertainty.

People generally talking about their health more may mean they had more health concerns, which could have been exacerbated by avoiding the hospital and also being more concerned in general about health issues.

6.7 Predicting Cities' Ability to Recover

We now look at distinguishing between those cities that recovered by the end of 2020 versus those that had not, among those that were affected. We maintain the same experimental setup as in the previous section.

6.7.1 Results and Discussion

We show the classification results in Table 6.9. We see that the task of distinguishing between recovered and non-recovered cities is a more difficult task; none of the features, except LIWC features, are able to predict recovery better than random chance. Therefore, our analysis here focuses only on the LIWC features. Logistic regression coefficients from a LIWC-only model are given in Table 6.10.

We see that recovered cities are more associated with past-tense language. As noted before, this likely indicates people sharing more about what they have done, as the Focus Past LIWC category consists primarily of past-tense verbs. They also talk more about their friends and homes.

On the other hand, non-recovered cities talk more about family, affiliations, and work. These may have been large aspects of life for people in these cities, which were then greatly impacted by the pandemic due to social distancing policies keeping people from seeing their extended families and co-workers. They also refer to themselves as part of a group more, using We words, indicating that they considered themselves affiliated with groups. This feeling of connection was likely impacted by the isolation brought on by the pandemic.

6.8 Broader Implications and Ethical Considerations

Our work, in conjunction with existing work [235, 257], shows that the pandemic had different effects on different communities. Further, we show that signals from social media can be predictive of how a community copes with adversity. We found that cities more affected by the pandemic tended to have less connected members and had previously placed more importance on life aspects that were most impacted by social distancing during the pandemic, such as seeing friends and participating in group activities. Our features were predictive of whether a city's wellbeing was

affected by the pandemic. However, predicting the subsequent recovery trajectory of affected cities proved to be more difficult, implying that there are other factors involved and further work is needed to understand community resilience over time.

Our findings indicate that differential policies should be put in place for communities, based on the pandemic's local impact. Cities more impacted by social distancing measures may need to place higher priority on re-establishing social activities, such as local events and cultural festivals. Furthermore, policymakers could use automatically-derived signals from social media as a real-time source of feedback for their policies, especially during times that require quick decision-making like during the pandemic. However, it's important that such factors are considered holistically.

A limitation of our findings is that they do not necessarily reflect the general public. Our work focuses on a single social media site (Reddit) where the users tend to be young[5] and male.[6] However, further studies using other social media, as well as surveys, can help support our insights.

Because our study is based solely on observational data, we cannot establish causal links between the community characteristics we have identified and the wellbeing recovery outcomes. To address this, future work could involve collecting ground truth data about city recovery and resilience, such as through large surveys of individuals in each city regarding their wellbeing during the pandemic.

Finally, our study should not be construed to be a comprehensive study of wellbeing. Subjective wellbeing does not consist solely of the presence of positive affect. It is more complex and multi-faceted, involving other aspects such as life satisfaction which are impacted in different ways [258]. Future work could study how community factors relate to these additional aspects of wellbeing. Furthermore, the relation of our wellbeing metric to metrics such as self-reported life satisfaction has not been

[5]https://www.statista.com/statistics/261766/share-of-us-internet-users-who-use-reddit-by-age-group/

[6]https://www.statista.com/statistics/1255182/distribution-of-users-on-reddit-worldwide-gender/

studied; a misalignment between stance expressed in social media and public opinion surveys has been noted in prior work [259], and we leave it to future work to study how wellbeing as expressed in social media posts relates to self-reported wellbeing.

6.9 Conclusion

We characterized the affective wellbeing patterns of cities across the US during the COVID-19 pandemic prior to vaccine availability, as exhibited in subreddits corresponding to the cities. We then derived linguistic and interaction features from the subreddit communities based on data prior to the pandemic and used them to predict how the affective wellbeing of each community would be impacted by the pandemic. We showed that communities with interaction characteristics corresponding to more closely connected users and higher engagement were less likely to be significantly impacted. Notably, we found that communities that talked more about social ties, such as friends, family, and affiliations, were actually more likely to be impacted. This may result from the social isolation policies affecting precisely these social ties. Additionally, we used the same features to predict how quickly each community would recover after the initial onset of the pandemic. We similarly found that communities that talked more about family, affiliations, and identifying as part of a group were more likely to recover more slowly. We showed that general community traits can be predictive of community resilience.

Features	Feature Name	Coef
Demographic	Latitude	1.215
Demographic	Population Density	1.000
User Interaction	Density	0.808
Demographic	Rent vs Own Rate	0.682
Demographic	Median Age	0.466
Demographic	Median Monthly Housing Cost	0.399
LIWC	Focus Past	0.317
LIWC	Focus Present	0.309
Post Interaction	Tree Size	0.230
Post Interaction	Direct Reply Count	0.167
LIWC	Home	0.164
Post Interaction	Max Level Width	0.137
Post Interaction	Leaf Node Count	0.136
User Interaction	Edge Count	0.102
User Interaction	Mean Degree	-0.034
LIWC	We	-0.067
User Interaction	Node Count	-0.101
User Interaction	Mean Connected Component Size	-0.121
LIWC	Work	-0.277
LIWC	Focus Future	-0.347
LIWC	Affiliation	-0.371
LIWC	Friend	-0.388
LIWC	Family	-0.405
LIWC	Leisure	-0.439
Post Interaction	Min Response Time	-0.516
User Interaction	Connected Component Count	-0.517
Demographic	Household Income	-0.553
LIWC	Health	-1.211
User Interaction	Mean Eccentricity	-1.238
User Interaction	Diameter	-1.511
User Interaction	Mean Shortest Path	-1.519

Table 6.8: Unaffected vs. Affected coefficients. Positive coefficients indicate that the feature is more associated with the subreddits of cities that are unaffected; negative coefficients indicate association with the affected subreddits of cities.

Features	Acc	P	R	F1	AUC
Random	0.500	—	—	—	—
Demographics	0.382	0.392	0.393	0.381	0.392
User Interaction	0.461	0.482	0.483	0.458	0.479
Post Interaction	0.447	0.433	0.433	0.433	0.317
LIWC	**0.539**	**0.542**	**0.543**	**0.537**	**0.562**
All	0.513	0.515	0.516	0.511	0.480

Table 6.9: Non-Recovered vs. Recovered classification results.

Features	Feature Name	Coef
LIWC	Focus Past	0.794
LIWC	Friend	0.598
LIWC	Home	0.361
LIWC	Focus Future	0.152
LIWC	Focus Present	0.024
LIWC	Leisure	-0.051
LIWC	Family	-0.422
LIWC	Affiliation	-0.484
LIWC	Health	-0.577
LIWC	Work	-0.624
LIWC	We	-0.680

Table 6.10: Non-Recovered vs. Recovered coefficients. Positive coefficients indicate that the feature is more associated with the subreddits of cities that recovered; negative coefficients indicate association with the subreddits of cities that did not recover. Only LIWC features are included, as other features did not yield performance exceeding random chance.

CHAPTER VII

Conclusion

In this thesis, we explored how we can use computational language models model and gain insight into people's attitudes and behaviors from large-scale linguistic data. We showed that uncovering implicit attitudes requires more than straightforward applications of existing NLP models, and develop computational methods to capture this more nuanced information. Further, we examined how attitudes manifest in language and behavior. We now revisit the questions posed in Chapter 1.

RQ 1: How can we computationally model the attitudes that people hold towards entities in their world?

We showed that it is possible to extract and predict people's implicit attitudes towards social roles through corpus statistics and dependency-based embedding models. We introduced a dataset of social roles and their associated descriptors in two cultures, India and US, and used this to conduct evaluations focused on predicting social roles. Our models showed stronger predictive ability when the train and test set cultures match, indicating that cultural differences can be automatically accounted for in our models, and that attitudes informed by a culture can be captured as well.

RQ 2: How can we predict the behaviors that people are likely to exhibit in a given context based on their personal characteristics?

In Chapter 3, we predicted alumni donation behavior through enhancing sparse

textual content. We introduced a dataset of alumni donations and explored four different methods of expanding sparse text, including lexicon generation methods and text embedding methods. We showed that we can classify large donors from non-donors with an accuracy up to 80%, and that enriching sparse text with textual features does benefit model performance.

In Chapter 4, we also explored behavior in the context of an online community, the subreddit r/getdisciplined, where users express intent towards self-improvement and personal change. Leveraging affect, linguistic style, and topic features, we built computational models that are able to distinguish between people with continued, persistent engagement in r/getdisciplined from those who did not continue.

Through these two chapters, we saw that latent information in language can be predictive of behavior and that natural language processing methods can be adapted and developed to capture this information.

RQ 3: How do attitude and behavior give insight into each other?

To understand the ties between attitudes and behavior, we conduct work in two directions. First, in Chapter 5, we predict attitude towards philanthropic causes based on engagement behavior with emails and personal background information.

In Chapter 6, we also analyzed shifts in subjective wellbeing across the US in response to COVID-19, and characterized how attitudinal traits expressed prior to the pandemic are predictive of the recovery behavior patterns and resilience.

From this, we showed that attitudes and behavior are intertwined, and can be indicative of one another.

7.1 Limitations, Broader Considerations, and Future Directions

While working with big data, it can be appealing to simply dig into the data, find salient patterns that emerge, and report those as findings. However, care should be taken to ensure that the underlying social science questions and constructs are properly operationalized and scoped within the computational work. Otherwise, conclusions can be very limited in how well they can generalize to a broader population, and therefore limited in usefulness.

If one is crafting a new measure for a social science construct derived from data, one can validate the measure with pre-existing measures for the same construct. For instance, in our work quantifying community subjective wellbeing in the wake of the pandemic (Chapter VI), we compare our data-derived social resilience measures with existing resilience measures. We found that they were correlated, as expected.

For NLP models to be useful in gaining social science insights, we need to validate that they accurately reflect the desired construct for the study. To do this, one can consult people who can manually label data with the construct. For example, we explicitly survey people about their attitudes towards social roles (Chapter II). We then use the survey results to validate whether our models can automatically extract social role aspects.

However, other work that we presented did not have the benefit of manual labeling, such as when we looked at personal persistence (Chapter IV) or community wellbeing (Chapter VI), and may therefore limited in how well they can truly represent the nature of these constructs. While these exploratory studies provide interesting insights, they would benefit from further studies where the construct in question is more narrowly defined and, to the extent possible, manually validated.

In most cases, a given study cannot fully capture all relevant aspects of a partic-

ular construct. Measures may only reflect a limited portion of the whole construct. One could either expand the measures or focus the scope of the study. For instance, subjective wellbeing encompasses many different facets, but we focused on affective wellbeing in Chapter VI. Our study does not present a comprehensive view of overall subjective wellbeing and therefore our insights do not apply broadly to all of subjective wellbeing. However, prior work has shown that different aspects of wellbeing (e.g. affective vs cognitive) are affected differently. Therefore, future work into how cognitive wellbeing was affected by the pandemic, in combination with our current work, could yield broader insights on wellbeing.

Similarly, while we aimed to measure persistence in Chapter IV, persistence in an online forum likely does not fully generalize to persisting in real life settings. However, future work can link social media use with real world persistence behavior, such as through self reports of exercise or surveys of self-efficacy over time.

Finally, while we are able to gain useful insight into the problems presented here, the relationships between variables are based on correlation. Exploring causal relations between attitudes and behavior presents a promising way to yield stronger conclusions.

BIBLIOGRAPHY

[1] Sharma E, De Choudhury M. Mental health support and its relationship to linguistic accommodation in online communities. In Conference on Human Factors in Computing Systems - Proceedings, 2018. doi:10.1145/3173574.3174215.

[2] Cunha T, Tan C, Jurgens D, Romero DM. Are all successful communities alike? Characterizing and predicting the success of online communities. In The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019, 2019. doi:10.1145/3308558.3313689.

[3] Newell E, Jurgens D, Saleem HM, Vala H, Sassine J, Armstrong C, Ruths D. User migration in online social networks: A case study on Reddit during a period of community unrest. In Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016, 2016.

[4] Zagheni E, Garimella VRK, Weber I, State B. Inferring international and internal migration patterns from twitter data. In WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web, 2014. doi:10.1145/2567948.2576930.

[5] Chancellor S, Counts S. Measuring employment demand using internet search data. In Conference on Human Factors in Computing Systems - Proceedings, 2018. doi:10.1145/3173574.3173696.

[6] Olteanu A, Vieweg S, Castillo C. What to expect when the unexpected happens: Social media communications across crises. In CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing, 2015. doi:10.1145/2675133.2675242.

[7] Althoff T, Jindal P, Leskovec J. Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior . doi:10.1145/3018661.3018672.

[8] Singla P, Richardson M. Yes, There is a Correlation-From Social Networks to Personal Behavior on the Web . .

[9] De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In Seventh international AAAI conference on weblogs and social media, 2013. pp. 128–139.

[10] Vijayaraghavan P, Vosoughi S, Roy D. Twitter demographic classification using deep multi-modal multi-task learning. In ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 2017. doi:10.18653/v1/P17-2076.

[11] Burger JD, Henderson J, Kim G, Zarrella G. Discriminating Gender on Twitter 2011. pp. 1301–1309.

[12] Mukherjee A, Liu B. Improving Gender Classification of Blog Authors. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing 2010. p. 158–166.

[13] Garimella A, Banea C, Mihalcea R. Demographic-aware word associations. Technical report 2018. doi:10.18653/v1/d17-1242.

[14] Welch C, Kummerfeld JK, Pérez-Rosas V, Mihalcea R. Compositional demographic word embeddings 2020. doi:10.18653/v1/2020.emnlp-main.334.

[15] Hovy D. Demographic factors improve classification performance. In ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, 2015. doi:10.3115/v1/p15-1073.

[16] Plank B, Hovy D. Personality Traits on Twitter—or—How to Get 1,500 Personality Tests in a Week. 2015. doi:10.18653/v1/w15-2913.

[17] Wilson S, Mihalcea R, Boyd R, Pennebaker J. Disentangling Topic Models: A Cross-cultural Analysis of Personal Values through Words. 2016. doi:10.18653/v1/w16-5619.

[18] Mitra T, Counts S, Pennebaker JW. Understanding Anti-Vaccination Attitudes in Social Media. Technical report 2016.

[19] Preotiuc-Pietro D, Hopkins DJ, Liu Y, Ungar L. Beyond binary labels: Political ideology prediction of twitter users. In ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 2017. doi:10.18653/v1/P17-1068.

[20] De Choudhury M, Counts S, Horvitz E. Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth. ; 2013.

[21] Hill RJ, Fishbein M, Ajzen I. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Contemporary Sociology 1977. doi:10.2307/2065853.

[22] Sauper C, Haghighi A, Barzilay R. Content models with attitude. In ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011. doi:10.1145/2010000/2002517/p350-sauper.pdf.

[23] Jiang Y, Song X, Harrison J, Quegan S, Maynard D. Comparing Attitudes to Climate Change in the Media using sentiment analysis based on Latent Dirichlet Allocation. 2018. doi:10.18653/v1/w17-4205.

[24] Abu-Jbara A, Diab M, Dasigi P, Radev D. Subgroup detection in ideological discussions. In 50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 - Proceedings of the Conference, 2012. doi:10.7916/D8H421R3.

[25] Hassan A, Qazvinian V, Radev D. What's with the Attitude? Identifying sentences with attitude in online discussions. In EMNLP 2010 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2010.

[26] Rodríguez-Penagos C, Grivolla J, Codina-Fibá J. A hybrid framework for scalable Opinion Mining in Social Media: detecting polarities and attitude targets. Proceedings of the Workshop on Semantic Analysis in Social Media 2012. .

[27] Somasundaran S, Wilson T, Wiebe J, Stoyanov V. QA with attitude: Exploiting opinion type analysis for improving question answering in online discussions and the news. In ICWSM 2007 - International Conference on Weblogs and Social Media, 2007.

[28] Stoyanov V, Cardie C, Wiebe J. Multi-perspective question answering using the OpQA corpus. In HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2005. doi:10.3115/1220575.1220691.

[29] Volkova S, Wilson T, Yarowsky D. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2013.

[30] Garimella A, Banea C, Hovy D, Mihalcea R. Women's syntactic resilience and men's grammatical luck: Gender-bias in part-of-speech tagging and dependency parsing. In ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 2020. doi:10.18653/v1/p19-1339.

[31] Fazio RH, Zanna MP. Direct experience and attitude-behavior consistency. Advances in Experimental Social Psychology 1981. doi:10.1016/S0065-2601(08)60372-X.

[32] Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods 2010. doi:10.1177/0261927X09351676.

[33] Mohammad S, Turney P. Emotions Evoked by Common Words and Phrases. Proceedings of the NAACL-HLT 2010. .

[34] Hirsh JB, Peterson JB. Personality and language use in self-narratives. Journal of Research in Personality 2009. doi:10.1016/j.jrp.2009.01.006.

[35] Lee CH, Kim K, Young SS, Chung CK. The relations between personality and language use. Journal of General Psychology 2007. doi:10.3200/GENP.134.4.405-414.

[36] Beukeboom CJ, Tanis M, Vermeulen IE. The Language of Extraversion: Extraverted People Talk More Abstractly, Introverts Are More Concrete. Journal of Language and Social Psychology 2013. doi:10.1177/0261927X12460844.

[37] Mohammad SM, Kiritchenko S, Zhu X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013) 2013; 2:321–327.

[38] Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. In Computational Intelligence, 2013. doi:10.1111/j.1467-8640.2012.00460.x.

[39] Kilgarriff A, Fellbaum C. WordNet: An Electronic Lexical Database. Language 2000. doi:10.2307/417141.

[40] Hamilton WL, Clark K, Leskovec J, Jurafsky D. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP-16) 2016. pp. 595–605.

[41] Wilson SR, Shen Y, Mihalcea R. Building and validating hierarchical lexicons with a case study on personal values. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018. doi:10.1007/978-3-030-01129-1{_}28.

[42] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 2013.

[43] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations ofwords and phrases and their compositionality. In Advances in Neural Information Processing Systems, 2013.

[44] Pennington J, Socher R, Manning CD. GloVe: Global vectors for word representation. In EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014. doi:10.3115/v1/d14-1162.

[45] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL HLT 2019 -

2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2019.

[46] Tan C, Lee L, Pang B. The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter. In 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference, 2014. doi:10.3115/v1/p14-1017.

[47] Guerini M, Pepe A, Lepri B. Do linguistic style and readability of scientific abstracts affect their virality? In ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 2012.

[48] Wu Y, Zhang Q, Huang X, Wu L. Phrase dependency parsing for opinion mining. In EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009, 2009. doi:10.3115/1699648.1699700.

[49] Chinsha TC, Joseph S. A syntactic approach for aspect based opinion mining. In Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC 2015, 2015. doi:10.1109/ICOSC.2015.7050774.

[50] Shweder RA. Thinking through cultures: Expeditions in cultural psychology. : Harvard University Press; 1991.

[51] Garimella A, Mihalcea R, Pennebaker J. Identifying cross-cultural differences in word usage. In Proceedings of the International Conference on Computational Linguistics (COLING 2016), 2016.

[52] Hovy D, Purschke C. Capturing regional variation with distributed place representations and geographic retrofitting. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018. pp. 4383–4394.

[53] Paul M, Girju R. Cross-Cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 2009. pp. 1408–1417.

[54] Garimella A, Banea C, Mihalcea R. Demographic-aware word associations. In Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP 2017), 2017.

[55] Jurgens D, Tsvetkov Y, Jurafsky D. Writer profiling without the writer's text. In International Conference on Social Informatics, 2017. pp. 537–558.

[56] Ashforth BE, Kreiner GE, Fugate M. All in a day's work: Boundaries and micro role transitions. Academy of Management review 2000;25:472–491.

[57] Biddle BJ. Recent developments in role theory. Annual review of sociology 1986;12:67–92.

[58] Eagly AH, Karau SJ. Role congruity theory of prejudice toward female leaders. Psychological review 2002;109:573.

[59] Katz D, Kahn RL. The social psychology of organizations, volume 2. : Wiley New York; 1978.

[60] Ritzer G, et al. The McDonaldization of society. : Pine Forge Press; 1992.

[61] Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. 1991.

[62] Sunstein CR. Social norms and social roles. Columbia law review 1996;96:903–968.

[63] Triandis HC, Marin G, Hui CH, Lisansky J, Ottati V. Role perceptions of hispanic young adults. Journal of Cross-Cultural Psychology 1984;15:297–320.

[64] Bamman D, O 'connor B, Smith NA. Learning Latent Personas of Film Characters . pp. 352–361.

[65] Salton G, Lesk M. Computer evaluation of indexing and text processing. Journal of the ACM 1968;15:8–36.

[66] Wang S, Manning CD. Baselines and bigrams: Simple, good sentiment and topic classification. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers 2012;2:90–94.

[67] Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. Journal of machine learning research 2003;3:1137–1155.

[68] Mikolov T, Yih Wt, Zweig G. Linguistic regularities in continuous space word representations. In NAACL HLT, 2013. pp. 746–751.

[69] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014. pp. 1532–1543.

[70] Rogers A, Hosur Ananthakrishna S, Rumshisky A. What's in your embedding, and how it predicts task performance. In Proceedings of the 27th International Conference on Computational Linguistics, 2018. pp. 2690–2703.

[71] Andrews M, Vigliocco G, Vinson D. Integrating experiential and distributional data to learn semantic representations. Psychological review 2009;116:463.

[72] Bamman D, Dyer C, Smith NA. Distributed representations of geographically situated language. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2014. pp. 828–834.

[73] Bruni E, Boleda G, Baroni M, Tran NK. Distributional semantics in technicolor. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1, 2012. pp. 136–145.

[74] Feng Y, Lapata M. Visual information in semantic representation. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010. pp. 91–99.

[75] Levy O, Goldberg Y. Dependency-Based Word Embeddings . pp. 302–308.

[76] Kent GH, Rosanoff AJ. A study of association in insanity. American Journal of Psychiatry 1910;67:37–96.

[77] Nelson DL, McEvoy CL, Schreiber TA. The university of south florida free association, rhyme, and word fragment norms. Behavior Research Methods, Instruments, & Computers 2004;36:402–407.

[78] Miller GA. WordNet: a Lexical database for English. Communications of the Association for Computing Machinery 1995;38:39–41.

[79] Berinsky AJ, Huber GA, Lenz GS. Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. Political analysis 2012; 20:351–368.

[80] Clifford S, Jewell RM, Waggoner PD. Are samples drawn from mechanical turk valid for research on political ideology? Research & Politics 2015; 2:2053168015622072.

[81] Esuli A, Sebastiani F. SentiWordNet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), 2006.

[82] Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Computational Intelligence 2013;29:436–465.

[83] Plutchik R. The Emotions. : New York: Random House; 1962.

[84] Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S. Opinionfinder: A system for subjectivity analysis. In Proceedings of HLT/EMNLP 2005 Interactive Demonstrations, 2005.

[85] Zafar L, Afzal MT, Ahmed U. Exploiting polarity features for developing sentiment analysis tool. In EMSASW@ ESWC, 2017.

[86] Daga SS, Raval VV, Raj SP. Maternal meta-emotion and child socioemotional functioning in immigrant indian and white american families. Asian American Journal of Psychology 2015;6:233.

[87] Raval VV, Raval PH, Salvina JM, Wilson SL, Writer S. Mothers' socialization of children's emotion in india and the usa: A cross-and within-culture comparison. Social Development 2013;22:467–484.

[88] Roseman IJ, Dhawan N, Rettek SI, Naidu R, Thapa K. Cultural differences and cross-cultural similarities in appraisals and emotional responses. Journal of cross-cultural psychology 1995;26:23–38.

[89] Chaudhari DL, Damani OP, Laxman S. Lexical co-occurrence, statistical significance, and word association. In Proceedings of the conference on empirical methods in natural language processing, 2011. pp. 1058–1068.

[90] De Deyne S, Navarro DJ, Storms G. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. Behavior research methods 2013;45:480–498.

[91] Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, 2014. pp. 55–60.

[92] Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993–1022.

[93] Brown PF, Desouza PV, Mercer RL, Pietra VJD, Lai JC. Class-based n-gram models of natural language. Computational linguistics 1992;18:467–479.

[94] Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. Journal of the American society for information science 1990;41:391–407.

[95] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, 2008. pp. 160–167.

[96] Turian J, Ratinov L, Bengio Y. Word representations: a simple and general method for semi-supervised learning. In Proceedings of the 48th annual meeting of the association for computational linguistics, 2010. pp. 384–394.

[97] Zou WY, Socher R, Cer D, Manning CD. Bilingual word embeddings for phrase-based machine translation. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013. pp. 1393–1398.

[98] Gupta A, Boleda G, Baroni M, Padó S. Distributional vectors encode referential attributes. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015. pp. 12–21. doi:10.18653/v1/D15-1002.

[99] Banerjee S, Pedersen T. An adapted lesk algorithm for word sense disambiguation using wordnet. In International conference on intelligent text processing and computational linguistics, 2002. pp. 136–145.

[100] Mohammad S, Kiritchenko S, Zhu X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In Proceedings of Semeval, 2013.

[101] Jiang L, Yu M, Zhou M, Liu X, Zhao T. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011. pp. 151–160.

[102] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Computational linguistics 2011;37:267–307.

[103] Mohammad SM, Turney PD. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, 2010. pp. 26–34.

[104] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005. pp. 347–354.

[105] Hu M, Liu B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004. pp. 168–177.

[106] Esuli A, Sebastiani F. Sentiwordnet: A high-coverage lexical resource for opinion mining. Evaluation 2007. pp. 1–26.

[107] Miller GA. Wordnet: a lexical database for English. Communications of the ACM 1995;38:39–41.

[108] Rao D, Ravichandran D. Semi-supervised polarity lexicon induction. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009. pp. 675–682.

[109] Esuli A, Sebastiani F. Pageranking wordnet synsets: An application to opinion mining. In ACL, 2007, volume 7. pp. 442–431.

[110] Hamilton WL, Clark K, Leskovec J, Jurafsky D. Inducing domain-specific sentiment lexicons from unlabeled corpora. arXiv preprint arXiv:160602820 2016. .

[111] Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman ME, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one 2013;8:e73791.

[112] Pennebaker JW, Graybeal A. Patterns of natural language use: Disclosure, personality, and social integration. Current Directions in Psychological Science 2001;10:90–93.

[113] Ott M, Choi Y, Cardie C, Hancock JT. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011. pp. 309–319.

[114] Le QV, Mikolov T. Distributed representations of sentences and documents. In ICML, 2014, volume 14. pp. 1188–1196.

[115] Levy O, Goldberg Y. Neural word embedding as implicit matrix factorization. In Advances in neural information processing systems, 2014. pp. 2177–2185.

[116] Levy O, Goldberg Y, Dagan I. Improving distributional similarity with lessons learned from word embeddings. Transactions of the Association for Computational Linguistics 2015;3:211–225.

[117] Kiros R, Zhu Y, Salakhutdinov RR, Zemel R, Urtasun R, Torralba A, Fidler S. Skip-thought vectors. In Advances in neural information processing systems, 2015. pp. 3294–3302.

[118] Faruqui M, Dodge J, Jauhar SK, Dyer C, Hovy E, Smith NA. Retrofitting word vectors to semantic lexicons. In Proc. of NAACL, 2015.

[119] Kenter T, De Rijke M. Short text similarity with word embeddings. In Proceedings of the 24th ACM international on conference on information and knowledge management, 2015. pp. 1411–1420.

[120] Yu L, Hermann KM, Blunsom P, Pulman SG. Deep learning for answer sentence selection. CoRR 2014;abs/1412.1632.

[121] Kenter T, Borisov A, de Rijke M. Siamese cbow: Optimizing word embeddings for sentence representations. CoRR 2016;abs/1606.04640.

[122] Hoyt JE. Understanding alumni giving: Theory and predictors of donor status. Online Submission 2004. .

[123] Meer J, Rosen HS. Does generosity beget generosity? alumni giving and undergraduate financial aid. Economics of Education Review 2012;31:890–907.

[124] McDearmon JT. Hail to thee, our alma mater: Alumni role identity and the relationship to institutional support behaviors. Research in Higher Education 2013;54:283–302.

[125] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, 2013. pp. 3111–3119.

[126] Argamon S, Whitelaw C, Chase P, Hota SR, Garg N, Levitan S. Stylistic text classification using functional lexical features. Journal of the American Society for Information Science and Technology 2007;58:802–822.

[127] Maas A, Daly R, Pham P, Huang D, Ng A, Potts C. Learning word vectors for sentiment analysis. In Proceedings of the Association for Computational Linguistics (ACL 2011), 2011.

[128] Farnadi G, Tang J, De Cock M, Moens MF. User profiling through deep multimodal fusion. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018. pp. 171–179.

[129] Mukherjee A, Liu B. Improving gender classification of blog authors. In Proceedings of the Conference on Empirical Methods in natural Language Processing, 2010. pp. 207–217.

[130] Schler J, Koppel M, Argamon S, Pennebaker J. Effects of age and gender on blogging. In Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs, 2006. pp. 199–204.

[131] Sarawgi R, Gajulapalli K, Choi Y. Gender attribution: tracing stylometric evidence beyond topic and genre. In Proceedings of the Fifteenth Conference on Computational Natural Language Learning, 2011. pp. 78–86.

[132] Baranski EN, Morse PJ, Dunlop WL. Lay conceptions of volitional personality change: From strategies pursued to stories told. Journal of personality 2017; 85:285–299.

[133] Covey SR, Covey S. The 7 habits of highly effective people. : Simon & Schuster; 2020.

[134] Sandberg S. Lean In: Women, Work and the Will to Lead. : Random House, Inc.; 2013.

[135] Marcus BH, Forsyth LH, Stone EJ, Dubbert PM, McKenzie TL, Dunn AL, Blair SN. Physical activity behavior change: issues in adoption and maintenance. Health psychology 2000;19:32.

[136] Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Silva APC, Meira Jr W, Beleigoli AMR. Factors associated with weight change in online weight management communities: a case study in the loseit reddit community. Journal of medical Internet research 2017;19:e17.

[137] Kanner RE, Connett JE, Williams DE, Buist AS, Group LHSR, et al. Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the lung health study. The American journal of medicine 1999;106:410–416.

[138] Achananuparp P, Lim EP, Abhishek V. Does journaling encourage healthier choices? analyzing healthy eating behaviors of food journalers. In Proceedings of the 2018 International Conference on Digital Health, 2018. pp. 35–44.

[139] Chung CF, Agapie E, Schroeder J, Mishra S, Fogarty J, Munson SA. When personal tracking becomes social: Examining the use of instagram for healthy eating. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017. pp. 1674–1687.

[140] Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. American journal of health promotion 1997;12:38–48.

[141] Schwarzer R, Renner B. Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. Health psychology 2000;19:487.

[142] Dong M, Jurgens D, Banea C, Mihalcea R. Perceptions of social roles across cultures. In International Conference on Social Informatics, 2019. pp. 157–172.

[143] Prochaska JO, Marcus BH. The transtheoretical model: Applications to exercise. 1994. .

[144] Laub JH, Sampson RJ. Understanding desistance from crime. Crime and justice 2001;28:1–69.

[145] Semenza JC, Hall DE, Wilson DJ, Bontempo BD, Sailor DJ, George LA. Public perception of climate change: voluntary mitigation and barriers to behavior change. American journal of preventive medicine 2008;35:479–487.

[146] Bentley F, Tollmar K, Stephenson P, Levy L, Jones B, Robertson S, Price E, Catrambone R, Wilson J. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. ACM Transactions on Computer-Human Interaction (TOCHI) 2013; 20:1–27.

[147] DiClemente CC, Prochaska JO. Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. 1998. .

[148] Schwarzer R. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. Applied psychology 2008; 57:1–29.

[149] Claro S, Paunesku D, Dweck CS. Growth mindset tempers the effects of poverty on academic achievement. Proceedings of the National Academy of Sciences 2016;113:8664–8668.

[150] Dweck C. Mindset: The New Psychology of Success. : Random House Publishing Group; 2006.

[151] Velicer WF, Diclemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: An integrative model. Addictive behaviors 1990;15:271–283.

[152] Ammari T, Schoenebeck S, Romero D. Self-declared throwaway accounts on reddit: How platform affordances and shared norms enable parenting disclosure and support. Proceedings of the ACM on Human-Computer Interaction 2019; 3:1–30.

[153] Jurgens D, McCorriston J, Ruths D. An analysis of exercising behavior in online populations. In Ninth international aaai conference on web and social media, 2015.

[154] Chung JE. Social networking in online support groups for health: how online social networking benefits patients. Journal of health communication 2014; 19:639–659.

[155] White M, Dorman SM. Receiving social support online: implications for health education. Health education research 2001;16:693–707.

[156] Kummervold PE, Gammon D, Bergvik S, Johnsen JAK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in norway. Nordic journal of psychiatry 2002;56:59–65.

[157] Turner RJ, Frankel BG, Levin DM. Social support: Conceptualization, measurement, and implications for mental health. Research in community & mental health 1983. .

[158] An J, Weber I. # greysanatomy vs.# yankees: Demographics and hashtag use on twitter. In Tenth International AAAI Conference on Web and Social Media, 2016.

[159] Shen Y, Wilson SR, Mihalcea R. Measuring personal values in cross-cultural user-generated content. In Social Informatics, I Weber, KM Darwish, C Wagner, E Zagheni, L Nelson, S Aref, F Flöck, eds., 2019. pp. 143–156.

[160] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research 2003;3:993–1022.

[161] Scherer KR. Personality markers in speech. : Cambridge University Press; 1979.

[162] Pennebaker JW. Using computer analyses to identify language style and aggressive intent: The secret life of function words. Dynamics of Asymmetric Conflict 2011;4:92–102.

[163] Kincaid JP, Fishburne Jr RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch 1975.

[164] Hutchinson JC, Sherman T, Martinovic N, Tenenbaum G. The effect of manipulated self-efficacy on perceived and sustained effort. Journal of Applied Sport Psychology 2008;20:457–472.

[165] Strecher VJ, McEvoy DeVellis B, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health education quarterly 1986; 13:73–92.

[166] Plutchik R. A general psychoevolutionary theory of emotion. In Theories of emotion, Elsevier, 1980, pp. 3–33.

[167] Cortes C, Vapnik V. Support-vector networks. Machine learning 1995;20:273–297.

[168] Zhang J, Chang JP, Danescu-Niculescu-Mizil C, Dixon L, Hua Y, Thain N, Taraborelli D. Conversations gone awry: Detecting early signs of conversational failure. arXiv preprint arXiv:180505345 2018. .

[169] Hamilton W, Zhang J, Danescu-Niculescu-Mizil C, Jurafsky D, Leskovec J. Loyalty in online communities 2017. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15710/14848.

[170] Vlahovic TA, Wang YC, Kraut RE, Levine JM. Support matching and satisfaction in an online breast cancer support community. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014. pp. 1625–1634.

[171] Cosley D, Frankowski D, Terveen L, Riedl J. Suggestbot: using intelligent task routing to help people find work in wikipedia. In Proceedings of the 12th international conference on Intelligent user interfaces, 2007. pp. 32–41.

[172] Resnick P, Varian HR. Recommender systems. Communications of the ACM 1997;40:56–58.

[173] Belkin NJ, Croft WB. Information filtering and information retrieval: Two sides of the same coin? Communications of the ACM 1992;35:29–38.

[174] Mazare PE, Humeau S, Raison M, , Bordes A. Training millions of personalized dialogue agents. In Proceedings of the Conference in Empirical Methods in Natural Language Processing, 2018.

[175] Pradel B, Sean S, Delporte J, Guérif S, Rouveirol C, Usunier N, Fogelman-Soulié F, Dufau-Joel F. A case study in a recommender system based on purchase data. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011. pp. 377–385.

[176] Qin J, Zhang W, Wu X, Jin J, Fang Y, Yu Y. User behavior retrieval for click-through rate prediction. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020. pp. 2347–2356.

[177] Zhang Y, Dai H, Xu C, Feng J, Wang T, Bian J, Wang B, Liu TY. Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks. Proceedings of the AAAI Conference on Artificial Intelligence 2014;28.

[178] Xu Z, Pérez-Rosas V, Mihalcea R. Inferring Social Media Users' Mental Health Status from Multimodal Information 2020. https://aclanthology.org/2020.lrec-1.772.

[179] Wang X, He X, Cao Y, Liu M, Chua TS. Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019. pp. 950–958.

[180] Palumbo E, Rizzo G, Troncy R, Baralis E, Osella M, Ferro E. Knowledge graph embeddings with node2vec for item recommendation. In European Semantic Web Conference, 2018. pp. 117–120.

[181] Wang Z, Zhang J, Feng J, Chen Z. Knowledge Graph Embedding by Translating on Hyperplanes. Proceedings of the AAAI Conference on Artificial Intelligence 2014;28.

[182] Cai H, Zheng VW, Chang KCC. A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications. IEEE Transactions on Knowledge and Data Engineering 2018;30:1616–1637.

[183] Zhao Y, Liu Z, Sun M. Representation learning for measuring entity relatedness with rich information. In IJCAI International Joint Conference on Artificial Intelligence, 2015, volume 2015-Janua. pp. 1412–1418.

[184] Li J, Zhu J, Zhang B. Discriminative Deep Random Walk for network classification. In 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers, 2016, volume 2. pp. 1004–1013. doi: 10.18653/v1/p16-1095.

[185] Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems 2018;151:78–94.

[186] Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. LINE: Large-scale information network embedding. In WWW 2015 - Proceedings of the 24th International Conference on World Wide Web, 2015. doi:10.1145/2736277.2741093.

[187] Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks. KDD : proceedings International Conference on Knowledge Discovery & Data Mining 2016;2016:855–864.

[188] Socher R, Chen D, Manning CD, Ng AY. Reasoning with neural tensor networks for knowledge base completion. In Advances in Neural Information Processing Systems, 2013.

[189] Toutanova K, Chen D, Pantel P, Poon H, Choudhury P, Gamon M. Representing text for joint embedding of text and knowledge bases. In Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, 2015. pp. 1499–1509. doi:10.18653/v1/d15-1174.

[190] Xiao H, Huang M, Meng L, Zhu X. SSP: Semantic space projection for knowledge graph embedding with text descriptions. In 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017. pp. 3104–3110.

[191] Yin H, Cui B, Chen L, Hu Z, Huang Z. A temporal context-aware model for user behavior modeling in social media systems. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 2014. pp. 1543–1554. doi:10.1145/2588555.2593685.

[192] Wilson SR, Mihalcea R. Predicting human activities from user-generated content. In ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 2020. pp. 2572–2582. doi:10.18653/v1/p19-1245.

[193] Dong M, Xu X, Zhang Y, Stewart I, Mihalcea R. Room to Grow: Understanding Personal Characteristics Behind Self Improvement Using Social Media. In Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media, 2021. pp. 153–162. doi:10.18653/v1/2021.socialnlp-1.13.

[194] Bekkers R. Who gives what and when? A scenario study of intentions to give time and money. Social Science Research 2010;39:369–381.

[195] Snipes RL, Oswald SL, Snipes RL, Oswald SL. Charitable giving to not-for-profit organizations: factors affecting donations to non-profit organizations 2010. .

[196] Shier ML, Handy F. Understanding online donor behavior: the role of donor characteristics, perceptions of the internet, website and program, and influence from social networks. International Journal of Nonprofit and Voluntary Sector Marketing 2012;17:219–230.

[197] Kitchen H. Determinants of charitable donations in Canada: A comparison over time. Applied Economics 1992;24:709–713.

[198] Dong M, Mihalcea R, Radev D. Extending sparse text with induced domain-specific lexicons and embeddings: A case study on predicting donations. Computer Speech and Language 2020;59.

[199] Helms SE, Thornton JP. The influence of religiosity on charitable behavior: A COPPS investigation. The Journal of Socio-Economics 2012;41:373–383.

[200] Rajan SS, Pink GH, Dow WH. Sociodemographic and personality characteristics of Canadian donors contributing to international charity. Nonprofit and Voluntary Sector Quarterly 2009;38:413–440.

[201] Micklewright J, Schnepf SV. Who gives charitable donations for overseas development? Journal of Social Policy 2009;38:317–341.

[202] Breeze B. How donors choose charities: the role of personal taste and experiences in giving decisions. Voluntary Sector Review 2013;4:165–183.

[203] Sneddon JN, Evers U, Lee JA. Personal Values and Choice of Charitable Cause: An Exploration of Donors' Giving Behavior. Nonprofit and Voluntary Sector Quarterly 2020;49:803–826.

[204] Neumayr M, Handy F. Charitable Giving: What Influences Donors' Choice Among Different Causes? Voluntas 2019;30:783–799.

[205] Bachke ME, Alfnes F, Wik M. Eliciting Donor Preferences. Voluntas 2014; 25:465–486.

[206] Kawachi I, Berkman LF. Social ties and mental health. Journal of Urban health 2001;78:458–467.

[207] Aldrich DP, Meyer MA. Social capital and community resilience. American behavioral scientist 2015;59:254–269.

[208] Berkes F, Ross H. Community resilience: toward an integrated approach. Society & natural resources 2013;26:5–20.

[209] Walsh F. Traumatic loss and major disasters: Strengthening family and community resilience. Family process 2007;46:207–227.

[210] Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: Three decades of progress 1999. doi:10.1037/0033-2909.125.2.276.

[211] Lucas RE, Diener E, Suh E. Discriminant validity of well-being measures. Journal of personality and social psychology 1996;71:616.

[212] Luhmann M, Hofmann W, Eid M, Lucas RE. Subjective well-being and adaptation to life events: a meta-analysis. Journal of personality and social psychology 2012;102:592.

[213] Frederick S, Loewenstein G. Hedonic adaptation. Well-Being The foundations of Hedonic Psychology/Eds D Kahneman, E Diener, N Schwarz NY: Russell Sage 1999. pp. 302–329.

[214] Lyubomirsky S. Hedonic adaptation to positive and negative experiences. : Oxford University Press; 2011.

[215] Lucas RE. Adaptation and the set-point model of subjective well-being. Current Directions in Psychological Science 2007;16:75 – 79.

[216] Luhmann M, Intelisano S. Hedonic adaptation and the set point for subjective well-being. Handbook of well-being 2018. .

[217] Dong M, Jurgens D, Banea C, Mihalcea R. Perceptions of Social Roles Across Cultures, volume 11864 LNCS. ; 2019. doi:10.1007/978-3-030-34971-4{_}11.

[218] Ae FHN, Stevens SP, Betty AE, Ae P, Ae KFW, Pfefferbaum RL. Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness 2007. doi:10.1007/s10464-007-9156-6.

[219] Cutter SL, Ash KD, Emrich CT. The geographies of community disaster resilience. Global Environmental Change 2014;29:65–77.

[220] Sherrieb K, Norris FH, Galea S. Measuring Capacities for Community Resilience. Social Indicators Research 2010;99:227–247.

[221] De Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In Eighth international AAAI conference on weblogs and social media, 2014.

[222] Ammari T, Schoenebeck S, Romero DM. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support 2019. doi:10.1145/3359237.

[223] Cunha TO, Weber I, Haddadi H, Pappa GL. The Effect of Social Feedback in a Reddit Weight Loss Community . doi:10.1145/2896338.2896353.

[224] Andy A, Chu B, Fathy R, Bennett B, Stokes D, Guntuku SC. Understanding social support expressed in a COVID-19 online forum. In Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis, 2021.

[225] Hamilton W, Zhang J, Danescu-Niculescu-Mizil C, Jurafsky D, Leskovec J. Loyalty in Online Communities. Proceedings of the International AAAI Conference on Web and Social Media 2017;11:540–543.

[226] Fine A, Crutchley P, Blase J, Carroll J, Coppersmith G. Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data. Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science at EMNLP 2020 2020. pp. 50–54. doi:10.18653/v1/2020.nlpcss-1.6.

[227] Benton A, Mitchell M, Hovy D. Multitask learning for mental health conditions with limited social media data. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 2017.

[228] Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2014. pp. 51–60. doi:10.3115/v1/W14-3207.

[229] Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017.

[230] Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting changes in suicide content manifested in social media following celebrity suicides. In Proceedings of the 26th ACM conference on Hypertext & Social Media, 2015. pp. 85–94.

[231] Pavalanathan U, De Choudhury M. Identity management and mental health discourse in social media. In Proceedings of the 24th International Conference on World Wide Web, 2015. pp. 315–321. doi:10.1145/2740908.2743049.

[232] Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015. pp. 11–20. doi:10.3115/v1/W15-1202.

[233] Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of liwc2015. Technical report 2015.

[234] Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into us mental health during the covid-19 pandemic: longitudinal analysis of twitter data. Journal of medical Internet research 2020;22:e21418.

[235] Biester L, Matton K, Rajendran J, Provost EM, Mihalcea R. Understanding the impact of covid-19 on online mental health forums 2021;12.

[236] Kramer ADI. An Unobtrusive Behavioral Model of "Gross National Happiness". Proceedings of the 28th international conference on Human factors in computing systems - CHI '10 2010. doi:10.1145/1753326.

[237] Mihalcea R, Liu H. A corpus-based approach to finding happiness. AAAI Spring Symposium - Technical Report 2006;SS-06-03:139–144.

[238] Taylor SJ, Letham B. Forecasting at scale. The American Statistician 2018; 72:37–45.

[239] Cao I, Liu Z, Karamanolakis G, Hsu D, Gravano L. Quantifying the effects of COVID-19 on restaurant reviews. In Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media, 2021.

[240] Walton A, McCrea R, Leonard R. Csiro survey of community wellbeing and responding to change: Western downs region in queensland. Australia: CSIRO Land and Water Retrieved from http://gisera org au/publications/tech_reports_papers/socioeco-proj-3-community-wellbeing-report pdf 2014. .

[241] Malecki EJ. Regional social capital: Why it matters. Regional Studies 2012; 46:1023–1039.

[242] Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. The Lancet 2015;385:640–648.

[243] Gardner J, Oswald AJ. Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. Journal of health economics 2007;26:49–60.

[244] Howarth E, Hoffman MS. A multidimensional approach to the relationship between mood and weather. British Journal of Psychology 1984;75:15–23.

[245] Choudhury MD, Kiciman E. The Language of Social Support in Social Media and Its Effect on Suicidal Ideation Risk. Proceedings of the International AAAI Conference on Web and Social Media 2017;11:32–41.

[246] Kawachi I, Berkman LF. Social Ties and Mental Health. Journal of Urban Health: Bulletin of the New York Academy of Medicine 2001;78.

[247] Wallace M, Rabin AI. Temporal experience. Psychological Bulletin 1960;57:213.

[248] Zimbardo PG, Boyd JN. Putting time in perspective: A valid, reliable individual-differences metric. In Time perspective theory; review, research and application, Springer, 2015, pp. 17–55.

[249] Tang L, Liu H. Graph mining applications to social network analysis. In Managing and Mining Graph Data, Springer, 2010, pp. 487–513. doi:10.1007/978-1-4419-6045-0_16.

[250] Wilson C, Boe B, Sala A, Puttaswamy KP, Zhao BY. User interactions in social networks and their implications. In Proceedings of the 4th ACM European conference on Computer systems, 2009. pp. 205–218. doi:10.1145/1519065.1519089.

[251] Williams HT, McMurray JR, Kurz T, Lambert FH. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global environmental change 2015;32:126–138.

[252] Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, Swarup S. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017;35:3621–3638.

[253] Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using networkx. In Proceedings of the 7th Python in Science Conference, G Varoquaux, T Vaught, J Millman, eds., 2008. pp. 11 – 15.

[254] Wei Z, Liu Y, Li Y. Is this post persuasive? ranking argumentative comments in online forum. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2016.

[255] De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, CHI '16. p. 2098–2110. doi:10.1145/2858036.2858207.

[256] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 2002;16:321–357.

[257] Ashokkumar A, Pennebaker JW. Social media conversations reveal large psychological shifts caused by covid-19’s onset across u.s. cities. Science Advances 2021;7:eabg7843.

[258] Kettlewell N, Morris RW, Ho N, Cobb-Clark DA, Cripps S, Glozier N. The differential impact of major life events on cognitive and affective wellbeing. SSM-population health 2020;10:100533.

[259] Joseph K, Shugars S, Gallagher R, Green J, Quintana Mathé A, An Z, Lazer D. (mis)alignment between stance expressed in social media data and public opinion surveys. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021.