Deception Detection in Political Statements

by

Manvi Kamboj

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science (Information Systems and Technology) in the University of Michigan-Dearborn 2019

Master's Thesis Committee:

Assistant Professor Mohamed Abouelenien, Chair
Professor Brahim Medjahed
Assistant Professor Anys Bacha

Table of Contents

List of Tables

List of Figures

Abstract

Politicians use deceit as a successful strategy for political gains and positive image creation. Political fact checking organizations assess the political statements for being truthful or deceitful. This is a challenging problem and has enormous social and political impacts. Automatic deception detection has gained interest in applications, such as security purposes, criminal investigations, and social interactions over the past years, however, there has not been significant statistical analysis for deceit detection in political statements. This gap in analysis can be attributed to the lack of any labeled multimodal dataset for political deception detection. In this thesis, we collected a novel dataset for analyzing political deceit from videos which are labeled by fact checking organizations. The dataset consists of 180 videos from 88 politicians from the two main political parties, including 87 Democratic and 93 Republican videos. An empirical analysis is conducted using the new dataset to investigate political deception detection using linguistic, acoustic and visual modalities. The experimental results indicate the feasibility of detecting political deception using multimodal automated systems as well as specify behavioral patterns that are associated with truthful and deceptive statements from different politicians and different parties.

Chapter 1 Introduction

Democracy requires political trust. How citizens show faith in political institutions, such as political parties and government when faced with uncertainty explains their political trust. Politicians are known to lead people down the garden path. Statements by politicians can be beguiled by false promises and an attempt to delude the reality. The use of selective statements by the opposition and presenting it in an opposite sense is common in politics. Politicians have an artful way of twisting reality in their favor. In everyday life it is easy to mislead by telling the truth. Politicians palter for their party and personal benefit. They exaggerate the benefits of their proposed policies and their achievements. When contradicted for their wrongdoing they indulge in carefully crafted denials. “Post-truth politics" is a term that was first used in 1992, but became the Oxford Dictionary 2016 Word of the Year. It is defined in the Oxford Dictionary as- “Relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.” The increased use of the term points to how the grey area between truths and lies has increased over the years. Therefore, the question that arises is, how can citizens assess the reliability of what politicians say and how to differentiate between truths, perceived truths and lies?

The need to assess deceit in political statements has influenced the emergence of fact checking as an independent stream in news reporting and thus has led to the rise of political fact checking establishments, such as PolitiFact, FactCheck.org and the Washington Post’s fact checker. These fact checking establishments assess the statements made by politicians and try to decipher if the statement was truthful or deceptive. Some recent studies have revealed that the faith of citizens in social media has reduced over time. The Edelman trust [1] results of 2018 shows that 71 to 75% of people who were surveyed were concerned about the dangers of fake news. Another study by Stanford [8] shows that people who were using social media for news could not differentiate between real and fake news stories thus plummeting the trust in social media as a news source.

It is not possible to use the traditional covert methodologies, such as polygraphy to detect deceit in everyday political statements. Inspired by these difficulties we want to extend the use of learning-based deception detection on political statements. The learning-based methodologies do not require special equipment, but rely on data collected from truthful and deceptive speeches. Early work in automatic deception detection relied on data collected in a special lab setting [31] and required subjects to speak about what they believed about certain topics and then their opposite opinion to provide the deceptive statements In [32] the subjects were interviewed on a mock crime scenario, where they lie about committing a crime in order to provide a dataset to understand the behavior of liars. The drawback in these datasets is that the subjects were in an artificial setting when they, enacted the statements, which can hinder the real emotional expressions observed in real-life deceitful and truthful scenarios. In [34] trial data was used to perform only textual analysis.

A new real-world multimodal high-stake dataset was introduced in [33] and was used for deception detection using textual and gesture modalities. Authors of [23,4,35,36] used the dataset proposed in [33] and applied multimodal techniques involving acoustic, visual and lexical analysis for deception detection. The results of these papers showed that multimodal techniques can be successfully used for deception detection for videos recorded in an unimpeded environment.

There is no existing dataset to detect political deceit in the context of a learning-based multimodal system, as per our best knowledge. To bridge the gap in this work we propose a new dataset derived from videos collected based on political statements classified by Politifact as truthful or deceptive. Politifact does extensive work to classify statements made by public figures. The statements are freely available on their site and list what steps were involved in marking the high-stake statement as one of their lie categories. As explained in the dataset section, we collected the videos of these statements and extracted the part of the video where the statement is made. These video clips are used to perform multimodal analysis using the vision, linguistic, and acoustic channels to detect deceit in these publicly made political statements. In particular, we build systems that integrate and learn from these multimodal features to detect political lies. Moreover, we analyze the linguistic, visual and acoustic patterns that exhibit higher associations with truth and deceit.

Chapter 2 Related Work

Human emotion detection has always gained researchers' interest. Traditionally, psychological measures have been used for analyzing human behavior and emotions. Modalities that have been analyzed to detect deceptive behavior include psychological, physiological, thermal, linguistic, visual and acoustic [51, 52, 53, 54, 55, 56].

In the psychology domain for deception detection researchers studied behavioral changes associated with deceit. They posed questions related to how deceiver presents themselves and observed for suppressed or exaggerated facial expressions, verbal changes and fear level [57]. The approaches focused on aspects of felt emotion (anxiety, scratching, evasive response), arousal (eye blinking, pupil dilation, speech disturbance), control (planned or rehearsed behavior lacking spontaneity) and cognitive processing (hesitation, longer response time). [57] studied 120 independent sample and compiled 158 cues to detect deception.

The finding that deceivers and truth tellers show different cues has motivated researchers to use thermal, language and visual features to identify lies. Polygraph, a physiological method, which was developed between 1895 and 1945 [27], is still the most widely used tool for deception detection. It has limitations as it involves attaching physical instruments to monitor blood pressure, electrodermal skin conductance and respiratory rate. It is also very time consuming involving multiple rounds of interview and background checks to help polygraph examiners get additional information for inferring the interview. The emotional status of the subject, such as fear and anxiety can affect the result of a polygraph test [48]. Additionally, they are subject to human error and bias [47]. Another drawback of polygraph is that subjects can fake innocence by using specific training and countermeasures such as tongue biting, muscle tensing or stating lie in the pre-test questionnaire [49] making it unreliable.

The application of deception detection is not restricted to criminal justification but has found requirement in other areas, such as daily life, social media, online transactions and interviews, generating a necessity for finding alternative approaches. Work of [28] showed the usefulness of functional magnetic resonance imaging (fMRI) scanning to identify deception. The electrodermal activity signals captured from subjects in their study showed that there was significant activation in the right middle frontal cortex and the anterior cingulate gyrus parts of the brain, related mostly with emotions, communication, working memory semantic processing and decision making. However, application of such methodology is not feasible for large-scale application.

In [29, 32] thermal imaging was suggested as an alternative to polygraphs. The authors used mid-infrared cameras to acquire thermal data concentrating on the periorbital area of the subject's face. Blood flow distribution toward the musculoskeletal tissue was spotted as a nervous system response towards deceitful action. The system was found comparable to polygraph test and was found to attain equivalent results. Further experiments were conducted to detect areas that would provide more accuracy for deceit detection using thermal imaging. [3] found that features from the forehead area achieved improved performance when compared with different thermal facial regions.

Numerous studies presented the relationship between deceptive behavior and linguistic selection. [40] examined linguistic indicators of deceit in written stories. They tested and measured different linguistic dimensions of linguistic sets that were earlier found to be correlated with deception, negative emotion, self-references and indicators of cognitive complexity [50]. The study observed responses of subjects divided into deceptive and truthful groups acting in a "Mock Crime" scenario. The analysis provided evidence of linguistic differences between deceptive and truthful responses. Computational approaches were attempted to replicate the psychological experimentation findings. As mentioned in previously, [31] used classifiers to distinguish between deceptive and truthful writings covering three topics: opinions about death penalty, opinions on abortion and feelings about a best friend. In this data-driven approach, authors used unigrams features with classifiers such as Naive Bayes and Support Vector Machines. Results presented a clear separation between deceptive and truthful writings.

Computational linguistics has also found its applications in identification of deception involving applications with computer-mediated communication including forums, chats, social networks, online dating websites and product review websites. [58] analyzed linguistic patterns in deceptive reviews to identify spam reviews by using n-grams and LIWC dictionary. The authors found that the machine learning algorithms can be used to identify fake reviews with accuracies higher than human baselines. [59] used Support Vector Machines (SVM) and Spanish version of LIWC dictionary to detect deception in Spanish essays and [60] explored various strategies, such as part-of-speech, utterance length, LIWC features and lemmas to detect deceit in Italian court cases. Therefore, computational linguistic research is not limited to English as a language. [61] studied the variances among deceptive and truthful essays written by Spanish, Romanian and English speakers and found interesting lying and truthful patterns among cultures.

Computer vision provides an important tool for covert visual deception detection method. [30] used blob analysis to extract information about hands and head movement and augmented it with features extracted from the face.

For visual deceit detection, psychologists have studied microexpressions. Microexpression is a voluntary and involuntary response that occurs when the emotion center of the brain responds to a stimulus. According to [2] microexpressions last for only 0.5 or less of a second which is much smaller compared to an emotion which can last from 0.5 to 4 seconds. Because they appear for such a small fraction of time, it is extremely difficult to hide these types of expressions.

Moreover, they can be used to recognize emotions, such as surprise, fear, disgust, sadness, anger, happiness, contempt, etc. Based on these microexpressions, Facial Action Unit encoding System (FACS) [24] is being used extensively for automatic facial emotion recognition. FACS involves encoding of the slight facial changes that are related to some specific facial muscles. These FACS are widely used in emotion recognition in both phycology works and automatic emotion recognition in computer science. Along with verbal communication facial expressions can be used for detecting depression, deceit and other emotions.

Researchers discovered multimodal approaches for creating more informed deceit detection systems where features from different modalities are integrated. This aims to avoid the uncertainty related with the use of single modalities and presents the benefit of enriching the dataset with information from different sources. [62] utilized verbal and nonverbal features to find cues for deception detection.

In 2016, [4] analyzed the same dataset introduced in [33]. However, their approach added the use of acoustic features along with the linguistic and vision analysis that was previously used. [4] used five categories of features, the first being prosody features, such as intensity, pitch and loudness. The second is energy features that describe the amount of energy in the sound and relate it to how humans perceive loudness. The third is voicing probabilities features that present an estimate of unvoiced and voiced energy percentage. The fourth is spectral features that are frequency-based features, and cepstral features. The authors also used bag of words model using words with a frequency above five. These were then associated with the frequency of each unigram for every script. They used SenticNet to add emotional information. To further enhance their feature vector, they used Part of Speech (POS) weighted vectors. We build on the previous findings to develop our political deception detection system and we start by describing the data collection process.

Chapter 3 Dataset

Most of the earlier deception detection work was based on data collected under a controlled environment. There have been very few works which used data from real-life-scenarios. One dataset used in many recent deception detection works was introduced in [33]. Their dataset consists of 121 video clips collected from real life court trials. This work presented a multimodal approach for high-stake deception detection problems, but manual labeling of micro expressions was used. Some other studies [23,4,12,35,36] were also based on the same dataset. Another work which created a real-life dataset was [5]. They collected videos from the internet involving forensic cases of emotional pleaders asking for help to find their missing relatives or pleading to find a murderer or one who dismembered them. They collected videos of 69 suspects with an average length of 20 seconds. However, they did not have any ground truth of indicating which segment of the complete video recording the subject is lying or telling the truth. Their label is based on the final verdict for the suspect being guilty or truthful, but not the exact moment when the truth or lie is being said. They created temporal clips and assumed the label for all these clips as the decision of the final verdict. This could create errors in classification.

Additional significant lexical dataset related to fake news detection was introduced in [7]. Their work was based on statements classified by PolitiFact.com. It includes a list of 12,836 statements collected from natural context, such as TV ads, debates, Twitter, Facebook posts, interviews etc. The dataset is only useful for lexical analysis; however, it lacks the multimodal input.

To address the absence of any existing publicly available multimodal dataset for political deception detection we propose a novel dataset for training deception detection models. We are focused on political statements where the stakes are high and are fact-checked by the labeler. The goal is to build a multimodal collection of high-stake occurrences for real-life political deception as well as build classification models and conduct 'political lies' analysis using this dataset.

3.1 Data Collection

We target those statements which could be analyzed by the current multimodal technologies that are freely available and would not require any special or advanced instruments. We aim to perform a multimodal analysis, which requires synchronized data from all the three modalities: audio, lexical and visual. Thus, if a statement was made and reported in newspaper columns, tweets, radio and/or phone interviews, they were unsuitable for our analysis as they lack data from at least one of the three modalities.

During the data collection process, we faced multiple challenges. For visual analysis the subject's face needed to be clearly visible in the video, especially while the statement is being made. If the subject was far away from the camera the face detection tool was unable to detect it. Also, the subject's face needed to be visible for most of the duration of the video with a visual quality adequate to extract the visual features. The camera frame needed to have only the subject of interest in view without any other people in focus. If faces other than the subject's were visible, the tool might focus on someone other than the subject of interest and capture incorrect features. Sample pictures with these challenges are shown in Figure 1. To overcome the challenge of multiple people in a video, we manually cropped the videos with multiple faces to have only subject in focus. Thus, the resulting video contained only the main subject, allowing it to be useful for the analysis.

Figure 1 Video selection challenges

For the audio feature extraction process, it is required that the subject's voice be clearly audible without any music or noise in the background. There are many classified statements in PolitiFact that are based on political advertisements. We had to remove those videos as they included loud music playing in the background while the subject was making the statement.

For the transcript we had to search the actual source of the statement and then manually reconstruct the complete statement. This was required because PolitiFact often paraphrased instead of providing word to word transcriptions of the statements made.

The final process involved searching for a statement in which all the modalities were present, downloading its source video from YouTube, Facebook or Twitter and locating and trimming to the part of the clip where the statement is made and cropping the video if required. Given the challenges and our set goal, the data collection process was very demanding and arduous involving several iterations of web data mining, cleaning and analysis.

The final dataset consists of 180 videos with an average length of 18.65 seconds of 88 subjects. The distribution between parties is 87 Democratic videos and 93 Republican videos. The party wise distribution of data is shown in Figure 2. The party label of the subjects is according to the PolitiFact site or Wikipedia.org.

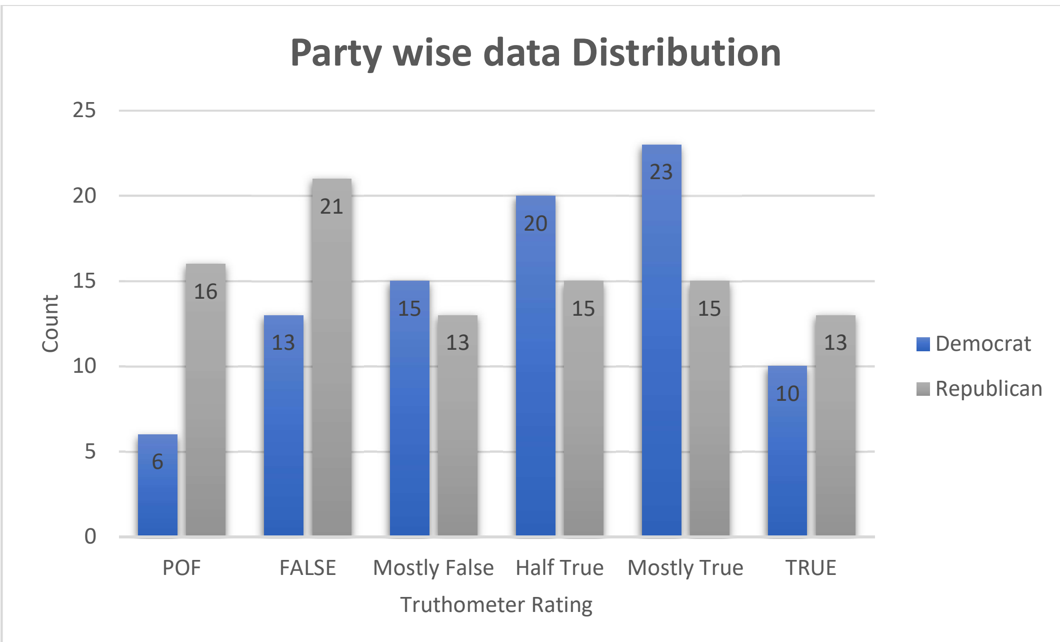

Figure 2 Party wise Truthometer Distribution

We collected data for the following six-point labels of truthfulness as rated by PolitiFact: 0 Pants on Fire, 1 False, 2 Mostly False, 3 Partly True, 4 Mostly True, and 5 True. The distribution per category is shown in Figure 3. The data sheet contains details for the video id, source (if available), date when the statement was made, flag to indicate if the statement was made in a prepared (in a speech etc.) or spontaneous (in a debate, interview etc.) setting, the political party, transcript, gender and duration of the video. There are 146 male subject videos and 34 female subjects. A sample from the dataset is shown in Table 1.

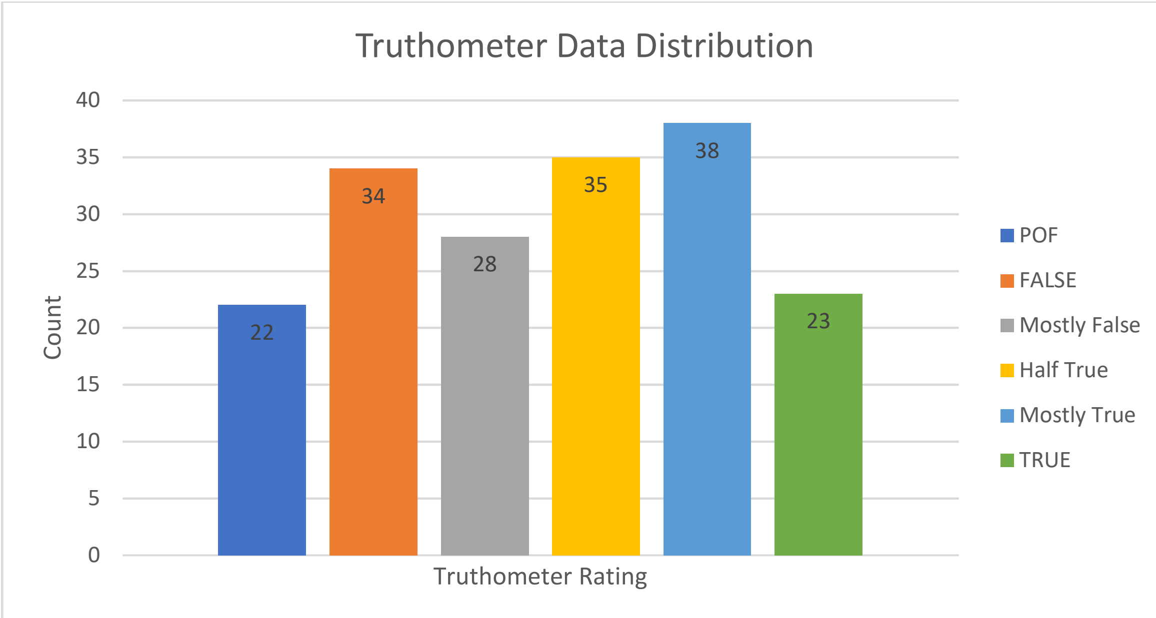

Figure 3 Truthometer Data Distribution

Video File	1103
Source	abc this week
Subject	Tim Scott
Date	Nov 26 2017
Prepared (Y/N)	N
Political Party	Republican
PolitiFact Rating (0-5 with 0 being pants on fire)	1
Transcript	Well, if you don't pay income taxes and we increase your refund by 40 percent, that is a direct dollar impact. In other words, you'll have more money to use to keep those ends together, those single mothers like mine, who are working paycheck to

	paycheck, they will now not get a $9,300 deduction.
Gender	M
Duration	20
Video File	1302
Source	WSB TV2
Subject	Jon Ossoff
Date	Jun 6 2017
Prepared (Y/N)	N
Political Party	Democrat
PolitiFact Rating (0-5 with 0 being pants on fire)	3
Transcript	And he is able to get coverage right now because there are protections for children like that with preexisting conditions. But secretary Handel supports a bill that will gut the protections for Americans with preexisting condition hundreds and thousands of them. Well if I might and with all due respect Secretary Handell first of all when it comes to preexisting conditions, I'm afraid that you are mistaken. The bill that passed the house guts protection for preexisting conditions for Georgians.

Gender	M
Duration	26

Table 1 Sample of data sheet details.

Chapter 4 Methodology

This section describes the different approaches we used to process multiple modalities.

4.1 Linguistic Modality

Linguistic features are commonly used in research papers involved in both emotion detection and deceit detection. We extract various linguistic features which have shown correlation with deceptive speech. The feature set used for feature extraction is extracted from the manual transcripts created for each video. It was ensured that the written transcript matches with the video recording. In particular, the following techniques are used:

LIWC: To extract features related to psychological state of a subject when stating the truthful and deceitful statements, we selected Linguistic Inquiry and Word Count (LIWC) [39]. LIWC is a transparent textual analysis program for creating word counts for psychologically meaningful groupings and has been used in textual deceit detection [40,41]. It extracts the cognitive, structural and emotional components present in text. We used the LIWC 2015 which has a dictionary containing almost 6,400 words, select emoticons and word stems [42] and produces 90 features for each statement received as input. The 90 features are grouped into 7 broad sub-categories, such as summary language variables (which relate to emotional tone, analytical thinking, clout and authenticity), standard linguistic dimensions (verbs, nouns etc.), general descriptors, psychological concepts related words, personal concerns (home, work), informal language markers and punctuation markers. To generate our feature set using LIWC we extracted the frequency of words of every category for each video transcript.

For additional understanding, we inspected our dataset to find some interesting insights related to LIWC categories. When comparing the words per sentence (WPS), shown in Figure 4, as seen in previous studies [64], the POF statements have higher count when compared to other truthometer rated text. This indicates that politicians tend to use more words when stating a lie.

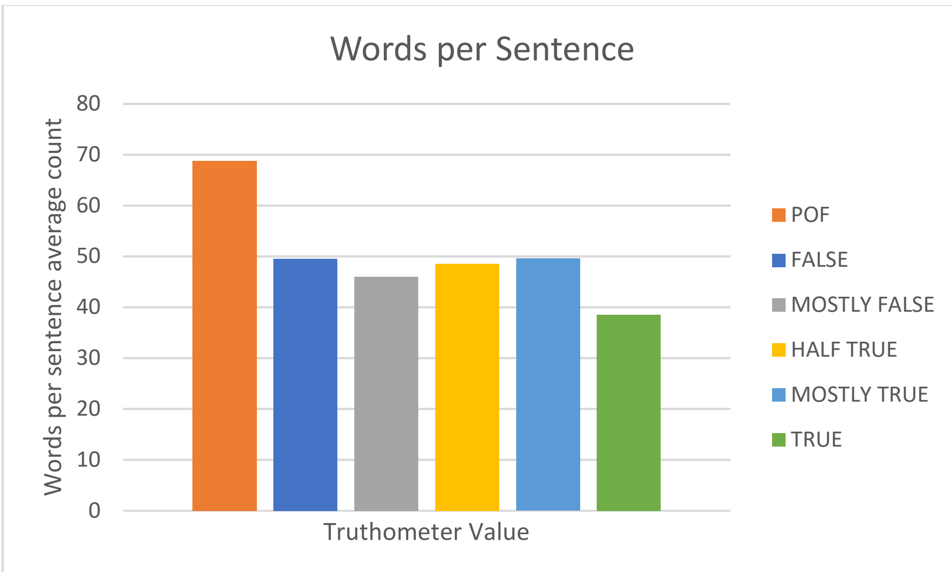

Figure 4 Distribution for Average words per Sentence for each Truthometer rating.

Some other interesting indications are shown in Figure 5, which depicts the comparison of interesting linguistic and other grammar categories for the truthometer ratings. It clearly depicts the use of the self-reference singular pronoun ‘I’ when stating a lie (POF rating) compared to the higher use of plural pronoun ‘we’ when stating a truthful statement (TRUE rating). Another noticeable insight is the higher use of interrogatives (how, when, what) when the statement is marked as POF. The figure also shows a higher use of numbers in truthful statements compared to deceitful statements. As numbers are inclined to be associated with facts, we can see their association with more truthful scenarios.

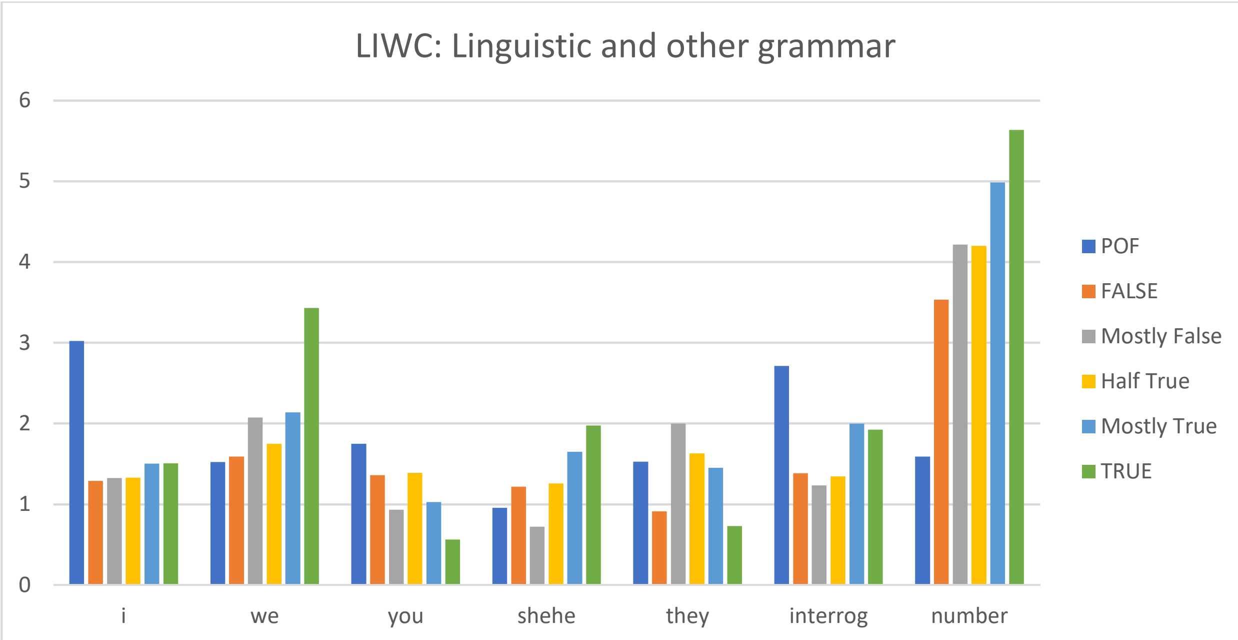

Figure 5 Use of LIWC classes for different Truthometer ratings

Table 2 shows the top 10 LIWC classes distribution for Republicans and Democrats. Observation from Table 2 is the higher value of emotional tone by Democrats when stating a false statement compared to true statements. A higher value of tone is associated with more positive style, implying Democrats speak more positively when stating a lie compared to when they are making a true statement. For Republicans the tone is not varying with true and false categories.

Republican	
FALSE	TRUE
Dic	Dic
Clout	Clout
Analytic	Analytic
WC	function
WPS	WC
function	WPS
Tone	Tone

Democrat	
FALSE	TRUE
Dic	Dic
Analytic	Clout
Clout	Analytic
Tone	WC
function	WPS
WC	function
WPS	Authentic

Authentic	Authentic	Authentic	Tone
Sixltr	Sixltr	Sixltr	Sixltr
verb	verb	verb	relativ

Table 2 Top 10 LIWC classes for Republican and Democrats

POS tagging: We used the NLTK's POS tagger to generate the Part-Of-Speech (POS) tags for the linguistic features. The tagger uses the PerceptronTagger which is a greedy averaged perceptron tagger, trained on Wall Street Journal corpus [37]. For our linguistic features, the extracted POS features are encoded using frequency distribution for each POS tag in the dataset.

Semantic features: We used GloVe (Global Vectors for Word Representation) [38] for creating a global vector for each video transcript. The GloVe is an unsupervised learning algorithm developed by Stanford for generating word embeddings. The embeddings are created using statistics derived from global word-word co-occurrence in a corpus. It is a global log-bilinear regression model using both global matrix factorization and local context window models. We use the Wikipedia 2014+ Gigaword5 pretrained corpus with word embedding vector of size 100. The transcripts from the videos are first lemmatized and then corresponding word embedding vectors are created using the GloVe corpus.

Unigrams: In [4] researchers used bag-of-words to represent the transcripts of video recordings. Bag of words was used to extract the unigram counts which were used as the linguistic feature set. The paper shows that the best results were obtained by a combination of LIWC and unigrams. Inspired by their finding, we decided to use BoW unigrams as an additional lexical feature set. The feature set consists of 2029 unique words.

We also extract the term frequency-inverse document frequency (TF-IDF) unigrams to provide normalized frequency distribution over the vocabulary consisting of all the words in our transcripts. Term frequency (TF) is the measure of occurrences of a word in a document. This term frequency is then weighted by an inverse document frequency count (IDF) which is a measure of occurrences of a word in the entire corpus. Mathematically, we can define TF-IDF as follows:

$$\underline{TFIDF = tf(w) * idf(w)}$$

Where tf(w) represents the number of times the word "w" appears in a document divided by total number of words in the document and idf(w) represents the log (total number of documents / number of documents containing word w).

Each unigram is encoded as TF-IDF values, that help us understand the importance of each word with respect to the whole corpus. We then set a threshold of 3.5 for binary class and 4.2 for multi-class classification. We only keep terms having high TFIDF score as informative candidate features for our model. Thus, all stop words, such as 'the', 'and', 'of', 'to', 'in', 'that', 'is', 'have', 'it', 'for', 'you', which are used frequently in the entire corpus, are eliminated from the feature set.

4.2 Acoustic Modality

We use OpenSmile (Open-Source Media Interpretation by Large Feature-space Extraction) to extract acoustic features [43]. OpenSMILE is a toolkit written in C++ in the scope of the European EU-FP7 research project SEMAINE and is used there as the acoustic emotion recognition engine in a real-time affective dialogue system. It can run on different main-stream platforms such as Linux, Windows, and MacOS. OpenSMILE is used for extracting features for signal processing and machine learning applications in order to classify speeches and music signals, all via a simple configuration file. The interesting part about OpenSMILE is that it can be used for audio-signal features and to analyze other modalities, such as physiological signals, visual signals, and other physical sensors. We can communicate with OpenSMILE using famous data formats used in the field of data mining and machine learning such as PCM WAVE for audio files, CSV (Comma Separated Value, spreadsheet format) and ARFF (Weka Data Mining) for text-based data files.

We extracted WAV format audio from the mp4 videos because OpenSmile requires the input to be in audio format for analysis. We used FFmpeg, which is a multimedia framework for this conversion. In order to distinguish between deceptive and truthful speech, we examined the Interspeech 2009 Emotion Challenge feature set (IS09) which is used for emotion recognition and since emotion and deception are correlated [45], we try to predict deception using acoustic emotion features.

The IS09 feature set [46] contains 32 descriptors that are divided into 16 delta regression coefficients and 16 low-level descriptors (LLD). These include 12 mel-frequency cepstral Coefficients (MFCCs 1-12), zero-crossing-rate (ZCR), root mean square (RMS) frame energy, pitch frequency (F0), and Noise-to-Harmonics ration (NHR), which indicates that the total number of features from this set is 384 features as 12 functionals are applied: Arithmetic mean, Standard deviation, Skewness, Kurtosis, maximum and minimum value, range, Relative position of max. and min. value, Linear regression slope, offset, and quadratic error. Accordingly, using this OpenSMILE's IS09 feature set, we were able to extract 384 features from the audio files.

In previous research in deception in spoken dialogue [44], the author compared OpenSMILE and Praat for classification. In [44], researchers identified the top 20 features from OpenSmile using the SelectKBest function in scikit-learn and f-classif score function (that uses the ANOVA F-value) in order to rank the features. We create our acoustic feature set using these top 20 features.

4.3 Visual Modality

4.3.1 IDT Features

The real-life videos in our dataset were collected from various online sources and are therefore of varying quality in terms of resolution, frame rates, and camera angles. For this reason, we found the improved dense trajectories (IDT) method to be an ideal candidate for extracting motion-based descriptors. This is accomplished by employing a series of computer vision techniques such as feature detection/tracking, geometric image transformations, and optical flow estimation to each set of consecutive frames throughout the video. Finally, each frame is processed at multiple scales to calculate action trajectories, which are robust against camera motion.

We applied the same series of operations to calculate the IDT features as originally proposed by [14]. However, we used a slightly different set of features for tracking and estimating the homography due to several reasons. First, we wanted to take advantage of the hardware acceleration features offered through the OpenCV CUDA API's due to their expected improved performance. The second reason was to limit the number of tracking points down to those most likely to capture human motion. Therefore, we use the features from Accelerated Segment Test (FAST) feature detector in favor of dense sampling. These points are then projected onto the optical flow field to calculate the motion trajectories. Similarly, we only used SURF descriptors to compute the holography estimation instead of using a combination of SURF and harris corners as done in previous work.

Finally, we used a fisher vector encoding to translate the low-level motion-based histograms into higher level descriptors to be used as training data for supervised learning [16]. We used the same encoding scheme used in several previous works [12,14,16]. Each MBH descriptor consists of 192 dimensions, which is reduced to 32 dimensions (D=32) using Principal Component Analysis (PCA). A codebook of visual words is created by estimating a Gaussian Mixture Model (GMM). We set the number of Gaussians to 16 (K=16) and randomly sample a subset of 256,000 samples. Stratified sampling was used to create an evenly distributed random sample. Therefore, 42,667 samples were randomly selected from each of the 6 classes defined by PolitiFact's Truth-O-Meter rating system. After encoding, each video is represented as a 2DK (1,024) dimensional fisher vector [17].

4.3.2 Facial Feature extraction

OpenFace is used to extract facial behavior features from the videos. It uses Constrained Local Neural Field (CLNF) [10] for facial landmark detection. CLNF represents an improvement over Constrained Local Model (CLM) [9] which struggled to perform in poor lighting condition, the presence of blockage and detecting landmarks in datasets which are not previously seen. The CLNF includes a Local Neural Field (LNF) patch expert which learns about both the adjacent and long-distance pixels by gaining information about the similarity and the long-distance sparsity constraints. This provides local variation of each landmark's appearance. The second main component for facial landmark detection is Point Distribution Model which captures variation in the shape of facial landmarks. When processing videos, OpenFace initializes the CLNF model based on facial landmarks detected in previous fames. This provides the detection of 68 facial landmarks [10].

To get the head pose OpenFace projects the 3D detected facial landmarks on the image, using an orthographic camera projection. For detecting eye gaze, they detect eye-region landmarks first and then calculate the pupil location based on the intersection of a ray passed through the pupil and the eye ball sphere. This provides the pupil location in 3D camera which is used for creating a feature vector for each eye [10].

For extracting facial appearance features, OpenFace uses a similarity transformed from the presently noticed facial landmarks to a neutral expression frontal landmark representation. HOG (Histogram of Oriented Gradients) is extracted from the aligned face generating a high dimensional vector (4464-dimensional vector). Then PCA is applied in order to reduce the dimensionality. This dimensionality reduced HOG features and facial features from CLNF are used for AU prediction and intensity measure [10]. It is capable of recognizing following subset of Action Units (AU): 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45 and reports the intensity and presence of these actions. The presence is recorded as 0 (absent) and 1 (present) respectively and the intensity ranges from 0 to 5, where 0 means the AU is not present, 1 is the lowest intensity level and 5 represents the maximum intensity.

We use AU, gaze and head pose features in addition to the IDT features to represent our visual modality feature set. OpenFace also provides the landmarks as their output, however including landmarks as features resulted in adding noise to the feature set as explained in experimental results section.

Figure 6 provides more insight on how deceptive and truthful statements associate with the AU features extracted from the videos. To find an indication of deceptive vs truthful expressions, the graph bars are calculated by subtracting the deceptive AU average feature values from the truthful ones. Therefore, a positive result specifies an association between an AU and truthfulness, and a negative result indicates an association between an AU and deception. The resulting figure provides interesting observations. The Cheek Raiser and Lip Corner Puller which are related to joy/smile, are mostly associated with truthful statements. Similarly, Upper lid raiser (which can be related to surprise, fear and anger, depending on which other AU is shown along with it) and Lip suck is strongly associated with truthful statements for all the subjects irrespective of their party. It is interesting how Republicans' Lip tightener is associated with truthful statements, while for Democrats it is associated with lying. Similar behavior is seen for Lip corner depressor (associated with sadness and disgust) between the parties, with Republicans showing its association with truth while Democrats showing it for deception. This could potentially indicate that Republicans make more truthful statements related to sadness.

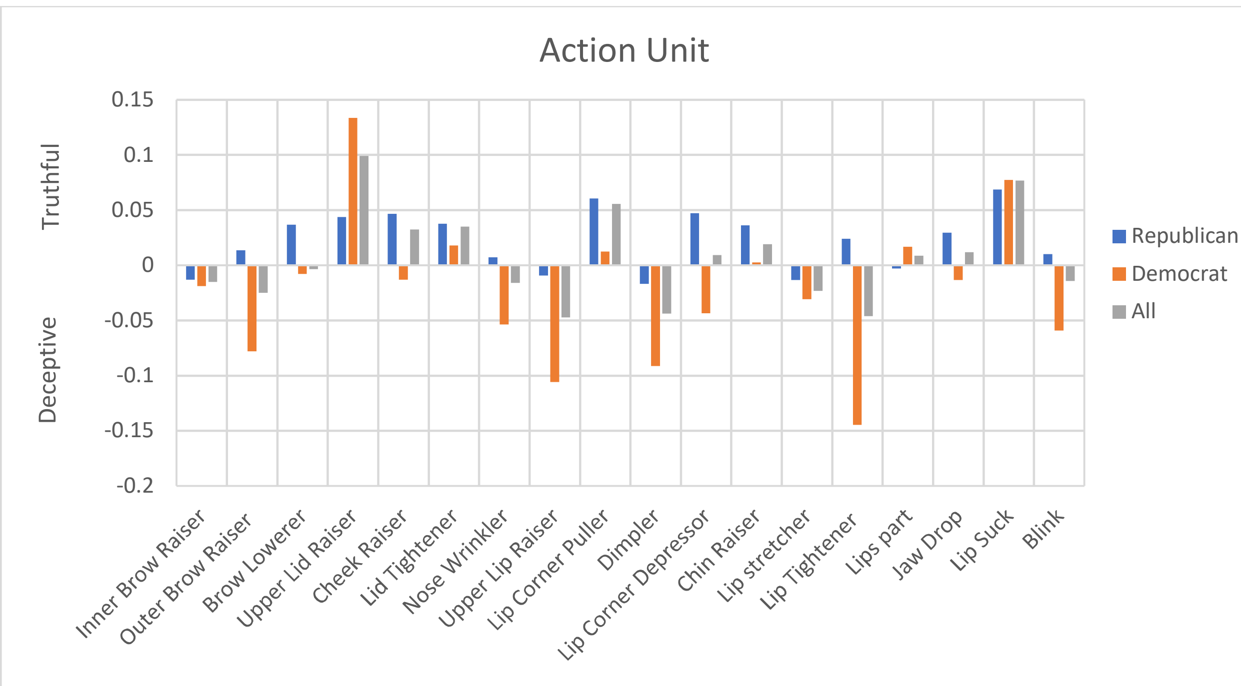

Figure 6 Party differences in deception, as reflected in AUs.

4.4 Multimodal Features

After extracting features from individual modalities, we test their performance independently and combined. The multimodal fusion is achieved by integrating the features collected from the three multimodal streams to create a single feature vector, which is then utilized to make a decision about the classification of the video. A decision tree classifier is used for classification with leave-one-subject-out cross validation scheme, using the scikit LeaveOneGroupOut functionality. This means that all the instances that belong to the same subject are held in reserve for testing, while all the instances belonging to the remaining subjects are used for training in each fold. We report the average overall accuracy and recall of each of the binary and multiclass classes.

Chapter 5 Experimental Discussion

Given our novel dataset of 180 instances for linguistic, acoustic and visual modalities, we started by evaluating classification results for deceit detection from individual modalities and then assess their combinations. In this section, we present the performance comparison on individual and combined modalities using Decision Tree classifier. We tried other well-known classifiers, such as Random Forest and Naïve Bayes however, Decision Tree was observed to perform better for most of the combinations. The dataset is represented in six different categories as provided by PolitiFact. We tried following combinations of these categories for binary and multiclass classifications

Binary Combinations

- Binary 1: POF + FALSE (as Deceptive) and MOSTLY TRUE + TRUE (as Truthful).
- Binary 2: POF + FALSE + MOSTLY FALSE (as Deceptive) and MOSTLY TRUE + TRUE (as Truthful).
- Binary 3: POF + FALSE + MOSTLY FALSE (as Deceptive) and HALF TRUE + MOSTLY TRUE + TRUE (as Truthful).

Multiclass Combinations

- 3-Class: POF + FALSE (as Deceptive), MOSTLY FALSE + HALF TRUE (as Partially Deceptive) and MOSTLY TRUE + TRUE (as Truthful).

- 4-Class: POF (as Deceptive), FALSE + MOSTLY FALSE (as Partially Deceptive), HALF TRUE + MOSTLY TRUE (as Partially Truthful) and TRUE (as Truthful).
- 6-Class: The individual six classes from PolitiFact.

5.1 Dataset

The novel dataset consists of 180 videos with 88 different subjects.

5.2 Individual Modalities

5.2.1 Linguistic

Table 2 lists the classification results of the binary combinations; binary 1, binary 2 and binary 3 for different individual linguistic techniques. The table shows the average recall for the truthful and deceptive classes as well as the overall accuracy for each binary classification type. The baseline of random guessing for the deceptive and truthful classes is shown as well. Table 3 lists the classification results for the multiclass combinations and, similar to the binary classification table, it shows the average recall for each class as well as the overall accuracy.

Binary1																Baseln
LIWC				POS			Bag of Words -Unigrams			Tf-IDF - Unigrams			GlOve			
Classes	ACC	Class	REC	ACC	Class	REC	AC	Class	REC	ACC	Class	REC	ACC	Class	REC	
Decision Tree	42.7	Deceptive	37.5	52.1	Deceptive	55.4	59.8	Deceptive	58.9	27.4	Deceptive	25	38.5	Deceptive	37.5	47.9
		Truthful	47.5		Truthful	49.2		Truthful	60.7		Truthful	29.5		Truthful	39.3	52.1
Binary 2																
Decision Tree	53.1	Deceptive	58.3	54.5	Deceptive	56	43.4	Deceptive	50	58.6	Deceptive	79.8	53.8	Deceptive	57.1	57.9
		Truthful	45.9		Truthful	52.5		Truthful	34.4		Truthful	29.5		Truthful	49.2	42.1
Binary 3																
Decision Tree	51.7	Deceptive	51.2	52.1	Deceptive	45.2	48.3	Deceptive	41.7	48.3	Deceptive	22.6	49.4	Deceptive	46.4	46.7
		Truthful	52.1		Truthful	51		Truthful	54.2		Truthful	70.8		Truthful	52.1	53.3

Table 3 Results for binary classification for LIWC, POS, BoW and TF-IDF linguistic methods.

	LIWC			POS			BoW unigram			TF-IDF unigram			GlOve			Bsln
3 Class																
Classes	ACC	Class	REC	ACC	Class	REC	ACC	Class	REC	ACC	Class	REC	ACC	Class	REC	
Decision Tree	38.9	Deceptive	33.9	33.3	Deceptive	21.4	26.7	Deceptive	17.9	26.1	Deceptive	8.9	32.8	Deceptive	28.6	31.1
		Mostly Deceptive	46		Mostly Deceptive	34.9		Mostly Deceptive	31.8		Mostly Deceptive	42.9		Mostly Deceptive	38.1	35
		Truthful	36.1		Truthful	42.6		Truthful	29.5		Truthful	24.6		Truthful	31.2	33.9
4 Class																
Decision Tree	30	Deceptive	18.2	31.7	Deceptive	9.1	27.8	Deceptive	4.6	44.4	Deceptive	22.7	29.4	Deceptive	27.3	12.2
		Mostly Deceptive	35.5		Mostly Deceptive	40.3		Mostly Deceptive	24.2		Mostly Deceptive	25.8		Mostly Deceptive	38.7	34.4
		Mostly Truthful	32.9		Mostly Truthful	37		Mostly Truthful	39.7		Mostly Truthful	75.3		Mostly Truthful	31.5	40.6
		Truthful	17.4		Truthful	13		Truthful	21.7		Truthful	17.4		Truthful	0	12.8
6 Class																
Decision Tree	18.3	POF	13.6	16.6	POF	0	13.3	POF	9.1	16.1	POF	22.7	12.8	POF	13.6	12.2
		FALSE	20.6		FALSE	5.9		FALSE	8.8		FALSE	14.7		FALSE	8.8	18.9
		Mostly False	14.3		Mostly False	21.4		Mostly False	7.1		Mostly False	0		Mostly False	17.9	15.6
		Half True	31.4		Half True	11.4		Half True	22.9		Half True	31.4		Half True	8.6	19.4
		Mostly True	7.9		Mostly True	34.2		Mostly True	13.2		Mostly True	13.2		Mostly True	18.4	21.1
		TRUE	21.7		TRUE	17.4		TRUE	17.4		TRUE	13		TRUE	8.7	12.8

Table 4 Multiclass classification results for LIWC, POS, BoW and TF-IDF linguistic methods.

For the binary classification results, BoW unigrams features provide the best results with the binary 1 grouping followed by TF-IDF unigrams for binary 2, using decision tree. A general trend that can be observed is that the performance in the majority of cases is close to the baseline with few exceptions, especially with BoW as mentioned earlier. This is expected given that we are using individual modalities and the exploratory nature of the feature we selected.

For multiclass classification, as shown in Table 3, the highest accuracy is recorded with TF-IDF for the 4-class grouping. With the 6-class classification scheme, classifiers are not able to learn from the data effectively when POS and TF-IDF features are used. Decision Tree is able to perform close to the baseline when using BoW and LIWC features. It can be noticed here that decision tree classifier is not able to learn effectively for multiclass classification, which could be a result of the size of the dataset and the distribution of the instances among different classes in comparison to the binary classification schemes.

All Lexical				LIWC + Bag of Words-Unigram			Baseline
Classes	Accuracy	Class	Recall	Accuracy	Class	Recall	
Binary 1							
Decision Tree	43.6	Deceptive	53.6	43.6	Deceptive	44.6	47.9
		Truthful	34.4		Truthful	42.6	52.1
Binary 2							
Decision Tree	47.6	Deceptive	55.1	49	Deceptive	52.4	57.9
		Truthful	38.8		Truthful	44.3	42.1
Binary 3							
Decision Tree	52.8	Deceptive	45.2	57.8	Deceptive	53.6	46.7
		Truthful	59.4		Truthful	61.5	53.3

Table 5 Binary classification results with all the lexical features and combination of LIWC+ BoW.

We tried different combinations of lexical features to determine whether any combination provides better performance for deceit detection, as shown in Table 5, and found that the integration of LIWC and BoW is able perform better, however, no significant improvement is observed over the usage of single linguistic feature sets in Table 3.

Table 6 shows similar results to Table 5 for the multiclass classification schemes for the integration of all linguistic features as well as for the combination of LIWC and BoW. A small improvement in performance over the baseline and the usage of individual linguistic feature sets can be observed for the 3-class classification scheme.

All Lexical				LIWC + Count-Unigram			Baseline
Classes	Accuracy	Class	Recall	Accuracy	Class	Recall	
3 Class							
Decision Tree	34.4	Deceptive	25	38.9	Deceptive	30.4	31.1
		Mostly Deceptive	44.4		Mostly Deceptive	54	35
		Truthful	32.8		Truthful	31.2	33.9
4 Class							
Decision Tree	27.8	Deceptive	14.8	25.6	Deceptive	9.1	12.2
		Mostly Deceptive	32.3		Mostly Deceptive	22.6	34.4
		Mostly Truthful	35.3		Mostly Truthful	34.3	40.6
		Truthful	5		Truthful	21.7	12.8
6 Class							
Decision Tree	17.2	POF	13.6	20.6	POF	13.6	12.2
		FALSE	26.5		FALSE	20.6	18.9
		Mostly False	21.4		Mostly False	21.4	15.6
		Half True	11.4		Half True	40	19.4
		Mostly True	15.6		Mostly True	7.9	21.1
		TRUE	9.1		TRUE	17.4	12.8

Table 6 Multiclass classification results with all the lexical features and combination of LIWC+ BoW

5.2.2 Acoustic

As earlier, the acoustic features are classified using the three binary and three multiclass classification schemes. The binary results are shown in Table 7 and the multiclass classification results are shown in Table 8.

Classes	Accuracy	Class	Recall	Baseline
Binary 1				
Decision Tree	49.6	Deceptive	53.6	47.9
		Truthful	45.9	52.1
Binary 2				
Decision Tree	42.8	Deceptive	47.6	57.9
		Truthful	36.1	42.1
Binary 3				
Decision Tree	50.6	Deceptive	48.8	46.7
		Truthful	52.1	53.3

Table 7 Acoustic feature classification result

The acoustic features show deteriorated performance as compared to the lexical features. The accuracy figures in the results are observed to be either very close to or below the baseline. The results show that the acoustic features are not able to distinguish between truthfulness and deception for political statements.

Classes	Accuracy	Class	Recall	Baseline
3 Class				
Decision Tree	31.7	Deceptive	33.9	31.1
		Partially Deceptive	31.8	35
		Truthful	29.5	33.9

4 Class				
Decision Tree	33.9	Deceptive	9.1	12.2
		Partially Deceptive	38.7	34.4
		Partially Truthful	46.6	40.6
		Truthful	4.4	12.8
6 Class				
Decision Tree	20	POF	9.1	12.2
		FALSE	11.8	18.9
		Mostly False	14.3	15.6
		Half True	28.6	19.4
		Mostly True	36.8	21.1
		TRUE	8.7	12.8

Table 8 Acoustic Multiclass classification results

5.2.3 Visual

OpenFace provides facial landmarks, AU, gaze and pose as output. We first tried training the classifiers with all the features provided by OpenFace and discovered that the landmarks add noise to the data, resulting in overfitting for decision tree. Hence, we decided to use AU, pose and gaze features, which have shown correlation to earlier deception detection work. The visual features are extracted at frame-level by OpenFace, creating 100-900 frames of data per video. For our analysis, we had to present each video as a single feature vector for the learning process. Therefore, we tried different statistical functions (mean, max and standard deviation) and their combinations to find best statistical measure. The combination of max+mean+standard deviation provides richer information for discriminating between truth and deception. The final feature vectors are generated using this statistical combination. The results of binary and multiclass classifications for visual features that are derived from OpenFace and IDT are shown in Table 9 and Table 10. The IDT flow failed to extract features for 4 videos, resulting in a total of 176 video in place of the 180 videos that are used for other methods. Accordingly, the tables have different baseline columns for OpenFace and IDT. As observed with the acoustic features, IDT and OpenFace features are unable to detect deceit from truthfulness, performing below or close to the baseline for most of the classification schemes.

OpenFace				Baseline OpenFace	IDT			Baseline IDT
Classes	Accuracy	Class	Recall		Accuracy	Class	Recall	
Binary1					Binary 1			
Decision Tree	41.9	Deceptive	35.7	31.1	45.1	Deceptive	48.2	47.8
		Truthful	47.5	33.9		Truthful	42.4	52.2
Binary 2								
Decision Tree	54.5	Deceptive	61.9	46.7	52.5	Deceptive	62.2	58.2
		Truthful	44.3	33.9		Truthful	39	51.8
Binary 3								
Decision Tree	54.4	Deceptive	52.4	46.7	44.9	Deceptive	34.2	46.6
		Truthful	56.3	53.3		Truthful	54.3	53.4

Table 9 OpenFace and IDT features binary classification results

We also tried to combine the OpenFace and IDT features, however, we did not observe improvement in performance using this combination.

OpenFace				Baseline OpenFace	IDT			Baseline IDT
Classes	Accuracy	Class	Recall		Accuracy	Class	Recall	
3 Class					3 Class			
Decision Tree	33.3	Deceptive	25	31.1	29.5	Deceptive	32	30.7
		Mostly Deceptive	36.5	35		Mostly Deceptive	30.5	35.8
		Truthful	37.7	33.9		Truthful	26.9	33.5

4 Class								
Decision Tree	29.4	Deceptive	18.2	12.2	40.3	Deceptive	18.2	11.9
		Mostly Deceptive	32.3	34.4		Mostly Deceptive	53.5	34.7
		Mostly Truthful	37	40.6		Mostly Truthful	44.1	40.3
		Truthful	8.7	12.8		Truthful	21.4	13.1
6 Class								
Decision Tree	18.3	POF	18.2	12.2	11.9	POF	8.7	11.9
		FALSE	23.5	18.9		FALSE	14.7	18.8
		Mostly False	14.3	15.6		Mostly False	18.9	15.9
		Half True	14.3	19.4		Half True	9.8	19.9
		Mostly True	18.4	21.1		Mostly True	11.5	20.5
		TRUE	21.7	12.8		TRUE	0	13.1

Table 10 OpenFace and IDT features multiclass classification results

5.3 Multimodal

Different combinations of lexical (LIWC, POS, Unigram, GlOve), acoustic and visual (IDT and OpenFace) features are experimented and compared for performance. The integration of these multimodal features provides richer dataset for political deceit detection. We apply early fusion by integrating the output of selected unimodal feature sets to create a single feature vector for each video. The results are listed in Table 11 and Table 12 for the multimodal binary classification schemes and for the multimodal multiclass combinations, respectively.

The best performing combination is binary 3 with an overall average accuracy of around 69% using the Decision Tree classifier. This represents a significant improvement over the performance of the individual modalities and individual feature sets. The results also show that “Binary 3” provides the best class combination of the six original classes provided by PolitiFact. This combination included all the three truthful and partially truthful categories in one class and all the three deceptive and partially deceptive categories as the other class. The results also indicate that developing a multimodal political deception detection system can aid in determining the truthfulness of the statements and promises made by politicians. On the other hand, no further improvement is noticed using the multimodal features for the multiclass classification schemes in Table 12.

Classes	Accuracy	Class	Recall	Baseline
Binary 1				
Decision Tree	50.4	Deceptive	44.6	31.1
		Truthful	55.7	33.9
Binary 2				
Decision Tree	49	Deceptive	50	46.7
		Truthful	47.5	33.9
Binary 3				
Decision Tree	68.9	Deceptive	63.1	46.7
		Truthful	74	53.3

Table 11 Bow unigrams+LIWC+OpenFace+acoustic binary classification results.

Classes	Accuracy	Class	Recall	Baseline
3 Class				
Decision Tree	32.8	Deceptive	33.9	31.1
		Mostly Deceptive	23.8	35
		Truthful	41	33.9
4 Class				
Decision Tree	34.4	Deceptive	18.2	12.2
		Mostly Deceptive	30.7	34.4
		Mostly Truthful	45.2	40.6
		Truthful	26.1	12.8

6 Class				
Decision Tree	17.2	POF	4.6	12.2
		FALSE	14.7	18.9
		Mostly False	28.6	15.6
		Half True	20	19.4
		Mostly True	15.8	21.1
		TRUE	17.4	12.8

Table 12 Bow unigrams+LIWC+OpenFace+acoustic multiclass classification results

Chapter 6 Conclusion

Deception detection is gaining a lot of interest recently in various applications ranging from security to social interaction. We introduced a novel multimodal dataset for political deception detection. To the best of our knowledge, there is no other dataset available for this task. The statements made in the videos were not restricted to a specific topic and, therefore, can be used in research work for covert automatic deception detection without using any special equipment.

We provided an analysis of the facial action units and linguistic (LIWC) features to discover interesting behavioral differences between deceptive and truthful politicians. It was interesting to find that lip corner depressor has higher association with truthful statements in Republicans. The same observation is also seen in LIWC classes where emotional tone is associated with truth for Republicans and with lie for Democrats.

Furthermore, we extracted multiple types of feature sets to train classifiers in order to automatically detect political lies. Different truthometer combinations were experimented to create multiple binary and multiclass classification schemes. While the individual and combined linguistic features showed promising potential for detecting political deceit, the visual and acoustic features did not exhibit similar performance, especially for the multiclass classification scenarios. However, the integration of the multimodal features achieved a significant improvement for the binary classification scheme. The results indicate that developing a multimodal political deception detection system can aid in determining the truthfulness of the statements and promises made by politicians. For future work, we plan to conduct further feature engineering in order to select the best discriminative set of features from each modality as well as experiment multiple machine learning algorithms.

Bibliography

[1] 2018 Edelman TRUST BAROMETER. (n.d.). Retrieved from https://www.edelman.com/research/2018-edelman-trust-barometer

[2] Inc., G. (2016, September 14). Americans' Trust in Mass Media Sinks to New Low. Retrieved from https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx

[3] (Abouelenien et al., 2014) Abouelenien, M., Pérez-Rosas, V., Mihalcea, R. and Burzo, M. (2014). Deception detection using a multimodal approach. [online] Available at: https://dl.acm.org/citation.cfm?id=2663229

[4] (Jaiswal, Tabibu and Bajpai, 2016) Jaiswal, M., Tabibu, S. and Bajpai, R. (2016). The Truth and Nothing But the Truth: Multimodal Analysis for Deception Detection - IEEE Conference Publication. [online] Ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/document/7836768/

[5] L. Su, M. LevineDoes ?lie to me? lie to you? an evaluation of facial clues to high-stakes deceptionComput. Vis. Image Understand., 147 (2016), pp. 52-68

[6] Michael T. Braun, Lyn M. Van Swol & Lisa Vang (2015) His Lips Are Moving: Pinocchio Effect and Other Lexical Indicators of Political Deceptions, Discourse Processes, 52:1, 1-20, DOI: 10.1080/0163853X.2014.942833

[7] Wang, William Yang. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. 1 May 2017, arxiv.org/abs/1705.00648.

[8] Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." The Journal of Economic Perspectives 31.2 (2017): 211-36. ProQuest. Web. 9 Mar. 2019.

[11] Wang, Heng, and Cordelia Schmid. "Action recognition with improved trajectories." Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013.

[12] Wu, Z., Singh, B., Davis, L., & Subrahmanian, V. S. (2017, December 12). Deception Detection in Videos. Retrieved from https://arxiv.org/abs/1712.04415

[13] Li, Xiaobai, et al. "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods." IEEE Transactions on Affective Computing (2017).

[14] Wang, Heng, and Cordelia Schmid. "Action recognition with improved trajectories." Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013.

[15] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." European conference on computer vision. Springer, Berlin, Heidelberg, 2006.

[16] Perronnin, Florent, Jorge Sánchez, and Thomas Mensink. "Improving the fisher kernel for large-scale image classification." European conference on computer vision. Springer, Berlin, Heidelberg, 2010.

[17] Sánchez, Jorge, et al. "Image classification with the fisher vector: Theory and practice." International journal of computer vision 105.3 (2013): 222-245.

[24] P. Ekman, W.V. Friesen, J.C. Hager Facial action coding system, the manual Research Nexus division of Network Information Research Corporation, USA(2002)

[25] Hurley, Carolyn M., et al. "Background Factors Predicting Accuracy and Improvement in Micro Expression Recognition." Motivation and Emotion 38.5 (2014): 700-14. ProQuest. Web. 12 Mar. 2019.

[27] Elkins, Aaron & Burgoon, Judee & Pantic, Maja & Zafeiriou, Stefanos. (2013). Unobtrusive Deception Detection.

[28] F. A. Kozel et al., “A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men,” J. Neuropsychiatry Clin. Neurosci., vol. 16, no. 3, pp. 295–305, Aug. 2004.

[29] Ioannis T. Pavlidis "Lie detection using thermal imaging", Proc. SPIE 5405, Thermosense XXVI, (12 April 2004); doi: 10.1117/12.547806; https://doi.org/10.1117/12.547806

[30] Michael N., Dilsizian M., Metaxas D., Burgoon J.K. (2010) Motion Profiles for Deception Detection Using Visual Cues. In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol 6316. Springer, Berlin, Heidelberg

[31] R. Mihalcea and C. Strapparava. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the Association for Computational Linguistics (ACL 2009), Singapore, 2009.

[32] I. Pavlidis, N. Eberhardt, and J. Levine. Human behaviour: Seeing through the face of deception. Nature, 415(6867), 2002.

[33] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo. Deception detection using real-life trial data. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI ’15, pages 59–66, New York, NY, USA, 2015. ACM.

[34] T. Fornaciari and M. Poesio. Automatic deception detection in Italian court cases. Artificial Intelligence and Law, 21(3):303–340, 2013.

[35] N. Carissimi, C. Beyan and V. Murino, "A Multi-View Learning Approach to Deception Detection," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, 2018, pp. 599-606. doi: 10.1109/FG.2018.00095
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8373888&isnumber=8373793

[36] M. Gogate, A. Adeel and A. Hussain, "Deep learning driven multimodal fusion for automated deception detection," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1-6. doi: 10.1109/SSCI.2017.8285382
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8285382&isnumber=8280782

[37] https://explosion.ai/blog/part-of-speech-pos-tagger-in-python

[38] Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In EMNLP, volume 14, 1532–1543.

[39] J. Pennebaker and M. Francis. Linguistic inquiry and word count: LIWC, 1999. Erlbaum Publishers.

[40] M. Newman, J. Pennebaker, D. Berry, and J. Richards. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin, 29, 2003.

[41] V. Rubin. On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the American Society for Information Science and Technology, 47(1):1-10, 2010.

[42] Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin

[43] Florian Eyben, Martin W ollmer, and Bj orn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia, pages 1459-1462. ACM, 2010

[44] Levitan, Sarah Ita. Deception in Spoken Dialogue: Classification and Individual Differences. Diss. Columbia University, 2019.

[45] Amiriparian, J. Pohjalainen, E. Marchi, S. Pugachevskiy, and B. Schuller. Is deception emotional? An emotion-driven predictive approach. In INTERSPEECH, pages 2011– 2015. ISCA, 2016.

[46] Schuller, Björn, Stefan Steidl, and Anton Batliner. "The interspeech 2009 emotion challenge." Tenth Annual Conference of the International Speech Communication Association. 2009

[47] T. A. Gannon, A. R. Beech, and T. Ward. 2009. Risk Assessment and the Polygraph, pp. 129–154. John Wiley and Sons Ltd. DOI: 10.1002/9780470743232.ch8 . 419, 430

[48] N. R. Council. 2003. The Polygraph and Lie Detection. The National Academies Press, Washington, DC. http://www.nap.edu/catalog/10420/the-polygraph-and-lie-detection. DOI: 10.17226/10420. 420

[49] G. Ganis, S. M. Kosslyn, S. Stose, W. L. Thompson, and D. A. Yurgelun-Todd. August 2003. Neural correlates of different types of deception: An fmri investigation. Cerebral Cortex, 13(8): 830–836. DOI: 10.1093/cercor/13.8.830. 431

[50] M. Zuckerman, B. M. DePaulo, and R. Rosenthal. 1981. Verbal and nonverbal communication of deception. Advances in Experimental Social Psychology, 14(1): 59. 423, 437

[51] T. Pfister and M. Pietik¨ainen. 2012. Electronic imaging & signal processing automatic identification of facial clues to lies. SPIE Newsroom. http://tomas.pfister.fi/. 420, 428

[52] J. Hillman, A. Vrij, and S. Mann. 2012. Um... they were wearing . . .: The effect of deception on specific hand gestures. Legal and Criminological Psychology, 17(2): 336–345. DOI: 10.1111/j.2044-8333.2011.02014.x. 420, 430

[53] M. Owayjan, A. Kashour, N. Al Haddad, M. Fadel, and G. Al Souki. Dec 2012. The design and development of a lie detection system using facial micro-expressions. In 2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), pp. 33–38. DOI: 10.1109/ICTEA.2012.6462897. 420

[54] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis. May 2013. Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging. IEEE Transactions on Biomedical Engineering, 60(5): 1280–1289. DOI: 10.1109/TBME.2012.2232927. 420

[55] B. Rajoub and R. Zwiggelaar. 2014. Thermal facial analysis for deception detection. IEEE Transactions on Information Forensics and Security, PP(99): 1–1. DOI: 10.1109/TIFS .2014.2317309. 420

[56] S. Feng, R. Banerjee, and Y. Choi. 2012. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, ACL ’12, pp. 171–175. Association for Computational Linguistics, Stroudsburg, PA. 420, 425

[57] B. M. DePaulo. 1992. Nonverbal behavior and self-presentation. Psychological Bulletin, 111(2): 203. DOI: 10.1037//0033-2909.111.2.203. 423

[58] M. Ott, Y. Choi, C. Cardie, and J. Hancock. 2011. 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, HLT ’11, pp. 309–319. Association for Computational Linguistics, Stroudsburg, PA. 425, 427, 443

[59] A. Almela, R. Valencia-Garc´ia, and P. Cantos. 2012. Seeing through deception: A computational approach to deceit detection in written communication. In Proceedings of the Workshop on Computational Approaches to Deception Detection, pp. 15–22. Association for Computational Linguistics, Avignon, France. DOI: 10.5195/lesli.2013.5 . 426, 427

[60] T. Fornaciari and M. Poesio. 2013b. Automatic deception detection in italian court cases. Artificial Intelligence and Law, 21(3): 303–340. DOI: 10.1007/s10506-013-9140-4. 426

[61] V. P´erez-Rosas and R. Mihalcea. 2014. Cross-cultural deception detection. In Proceedings of the Association for Computational Linguistics. DOI: 10.1002/9781118510001.ch8. 426, 427

[62] D. D. Henningsen, K. Valde, and E. Davies. 2005. Exploring the effect of verbal and nonverbal cues on perceptions of deception. Communication Quarterly, 53(3): 359–375. DOI: 10.1080/01463370500101329. 422, 433

[63] M. L. Jensen, T. O. Meservy, J. K. Burgoon, and J. Nunamaker. 2010. Automatic, multimodal evaluation of human interaction. Group Decision and Negotiation, 19(4): 367–389. DOI: 10.1007/s10726-009-9171-0. 433

[64] Michael T. Braun, Lyn M. Van Swol & Lisa Vang (2015) His Lips Are Moving: Pinocchio Effect and Other Lexical Indicators of Political Deceptions, Discourse Processes, 52:1, 1-20, DOI: 10.1080/0163853X.2014.942833