

Leveraging Natural Language Processing to Identify Caregiver Availability for Patients with ADRD from Electronic Medical Records

Elham Mahmoudi¹ | Wenbo Wu¹ | Cyrus Najarian¹ | James Aikens¹ |
Julie Bynum^{1,2} | V.G. Vinod Vydiswaran¹

¹University of Michigan, Ann Arbor, MI, USA

²University of Michigan Medical School, Ann Arbor, MI, USA

Correspondence

Elham Mahmoudi, University of Michigan, Ann Arbor, MI, USA.

Email: mahmoudi@med.umich.edu

Abstract

Background: Prior literature has posited that caregiver availability can potentially prevent adverse health events such as hospital readmissions in the ADRD population. Identifying caregiver availability for patients with Alzheimer's disease and related dementia (ADRD) can inform healthcare systems and hospital administrations on appropriate discharge planning. Caregiver availability is an important and influential domain in assessing risk of hospital readmissions. However, caregiver availability data is not systematically collected in our healthcare system. Our main objective was to use medical notes (from physicians, nurses, or social workers) to assess caregiver availability and types of caregivers for hospitalized patients with ADRD.

Method: We used 2016-2019 telephone-encounter medical notes from a single institution. Rule-based natural language processing (NLP) was used to develop an algorithm to identify whether the patient (1) resides at home; (2) resides at an institution; (3) has a formal caregiver; and (4) has an informal caregiver. We evaluated the data at the patient level. We used chart abstraction as the "gold standard," and examined the validity of the algorithm for both training (n = 749) and test sets (n = 227).

Result: The result from our test set indicated a high level of accuracy and reliability for identifying an informal caregiver (F1 = 0.942; accuracy = 0.947, sensitivity = 0.970, and specificity = 0.928). Identifying whether the patient lived at an institution was the least reliable measure (F1 = 0.638, accuracy = 0.899, sensitivity = 0.512, specificity = 0.978). The most common causes of misclassifications were: (1) incomplete or misspelled names of the facilities; (2) past/uncertain/undecided situations; (3) lack of specificity; (4) use of uncommon abbreviations; and (5) irregular use of templates.

Conclusion: Our NLP algorithm was able to identify whether the patient lived at home vs. an institution and whether the patient had a formal or informal caregiver. There is merit in continuing the NLP approach to identify more granular, caregiver-related information (e.g., type, hours available, etc.). More effort is needed towards generalizing the algorithm. This information can be used by other healthcare systems to detect ADRD patients who do not have a caregiver at home and may be at risk of adverse events including readmission.