MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220188

Identification of Transportation Barriers in Patient Portal Messages via Deep Semantic Embeddings and Clustering

Ming Huang [a], Aditya Khurana [b], George M. Mastorakos [b], Jungwei Fan [a*]

[a] *Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA*
[b] *Mayo Clinic Alix School of Medicine, Mayo Clinic, Scottsdale, AZ, USA*

Abstract

Patient portals have been widely used by patients to enable timely communications with their providers via secure messaging for various issues including transportation barriers. The large volume of portal messages offers an invaluable opportunity for studying transportation barriers reported by patients. In this work, we explored the feasibility of cutting-edge deep learning techniques for identifying transportation issues mentioned in patient portal messages with deep semantic embeddings. The successful creation of annotated corpus and identification of 7 transportation issues showed the feasibility of this strategy. The developed annotated corpus could aid in developing an artificial intelligence tool to automatically identify transportation issues from millions of patient portal messages. The identified specific transportation issues and the analysis of patient demographics could shed light on how to reduce transportation gaps for patients.

Keywords:

Patient Portals, Transportation of Patients, Health Services Accessibility

Introduction

Transportation is a necessary and important step for ongoing healthcare access. Transportation barriers can affect a person's access to healthcare services. Transportation barriers often lead to missed, canceled, or rescheduled appointments and limit individuals to access healthcare. Transportation barriers are often cited as a major type of barriers to healthcare access [10; 14]. In 2005, Wallace et al. performed a retrospective analysis of National Health Interview Surveys and National Transportation Availability and Use Surveys in 2002 to estimate magnitude of transportation barriers to health care and found that in the United States, about 3.6 million people missed at least one medical trip and fail to access healthcare because of transportation barriers [16]. Recently in 2020, Mary et al. used the data from the National Health Interview Survey in 2017 to examine transportation barriers to healthcare in the United States and discovered that 5.8 million (1.8%) persons in the United States delayed healthcare because of lack of available transportation. For seniors, they may face multiple access barriers due to disability, illness and need for frequent visits to their providers. Across the United States, transportation barriers pose the third leading cause of missing a medical appointment for seniors. Flores et al. conducted face to face survey on barriers to health care access and 21% participants reported the transportation barrier as a reason they failed to bring child in for a medical encounter [2]. The impact of transportation issues on patients and providers is significant. These issues may result in missed or delayed health care appointments, increased health expenditures and overall poorer health outcomes [1].

Patient portals have been widely used by patients to timely communicate with their providers via secure messaging for the various issues including transportation barriers [5]. Patient portals give patients unlimited access to their health information (e.g., clinical notes, test results, medications, and discharge summaries) from anywhere with Internet connection [6] and are becoming increasingly common. Over 90% of healthcare organizations (e.g., Veterans Administration, Kaiser Permanente, and Mayo Clinic), had provided patient portal services to their patients [4]. The convenient access and management of personal health information have been shown to improve patient self-management of diseases by promoting the awareness of disease knowledge, status and progress [15]. Additionally, patient portals provide a significant function of portal messaging for asynchronous communication between patients and their providers or care teams on a wide spectrum of tasks such as discussing transportation issues. The large volume of portal messages offers an invaluable opportunity for studying transportation barriers reported by patients.

In this study, we explore a pipeline enhanced by artificial intelligent to identify transportation issues mentioned in millions of patient portal messages at Mayo Clinic, a large multi-specialty academic health system. Blessed by modern advanced computational technology, artificial intelligent and natural processing now enables us to automatically screen the deep semantics of vast text documents with minimum feature engineering. Recently deep learning techniques have obtained very high performance across many different natural language processing (NLP) tasks [9]. More recently, the pre-trained deep learning models learn universal language representations from very large corpora. These pre-trained models achieved state-of-the-art performance on a large number of NLP tasks when they were published. The identified patient-reported transportation issues such as impaired driving ability and financial hardship will provide us a good understanding of transportation issues for patients' access to healthcare services. These findings could greatly inform the pertinent stakeholders to develop corresponding solutions in order to reduce transportation gaps for access to healthcare by patients.

Methods

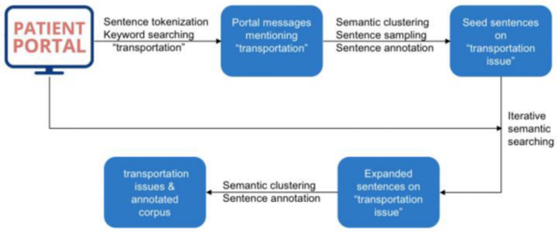

Figure 1 Workflow to identify transportation issues in patient portal messages

Figure 1 shows the overall workflow to identify transportation issues mentioned in patient portal messages leveraging deep semantic embedding. We collected all the patient portal messages created by patients in 2019 and tokenized them into sentences. We then use simple keyword searching to identify patient portal messages with the word "transportation" and manually annotated these sentences whether they discussed transportation issues for healthcare access as a starting seed dataset. After transforming these sentences into deep semantic embeddings, we iteratively expanded the transportation issue dataset by identifying portal message sentences mentioning transportation issues via searching semantically similar sentences and manual confirmation. Once obtaining enough relevant sentences, we performed clustering analysis to identify the specific themes of transportation issues mentioned in these portal message sentences.

Retrieval and Preprocessing of Patient Portal Messages

We collected all the patient portal messages generated by patients in 2019 at Mayo Clinic. Mayo Clinic is a large multi-specialty academic medical center focused on integrated patient care, education, and research and has operated the patient online services (patient portal) since 2010 [7]. Patients at Mayo Clinic use the secure online platforms to conveniently access information from their electronic health records and asynchronously interact with their providers [5].

We removed the patient-generated portal messages by request of their providers such as questionnaire messages before and after appointments and E-visit messages. We deleted quoted information, special characters and hyperlinks in these patient portal messages. Finally, we retrieved a total of 2,763,955 patient portal messages in 2019 for analysis. Sequentially, we tokenized each message into sentences using the NLTK sentence chunker [8] and collected 9,378,876 portal message sentences for identifying transportation issues reported by patients in patient portal messages.

Iterative Identification of Transportation Issue Sentences

After data collection and preprocessing, we constructed a seed dataset of portal message sentences mentioning transportation issues to iteratively identify transportation issue sentences. Specifically, we applied simple keyword match with the key word "transportation" over these patient portal messages. Over 2 million patient portal messages (9 million portal message sentences) were screened, the transportation keyword hit 1,160 patient portal messages (1,246 portal message sentences). We applied the Siamese BERT network [12] to generate deep semantic embeddings of the over 1,000 portal message sentences and performed agglomerative hierarchical clustering method to detect 20 clusters of these sentences. Stratified sampling was then used to select about 10 sentences from each of the 20 clusters (hence 200 in totals), which were independently annotated by two annotators into the following three classes:

1. The sentence is irrelevant to healthcare access or not about transportation.
2. The sentence mentioned transportation for healthcare access that did not cause a barrier.
3. The sentence mentioned a transportation issue that did impede access to healthcare.

The inter-annotation agreement (IAA) measured by Cohen's kappa between two annotators for the 200 portal message sentences was 0.78. If we combined the class 1 and class 2 as one class "1+2", the IAA for the binary annotation was 0.90. We identified a total of 44 portal message sentences that mentioned a transportation issue that obstruct the healthcare access as a seed dataset for sequential expansion.

We converted over 9 million portal message sentences into their deep semantic embeddings using the Siamese BERT network. We measured the Cosine similarity of each sentence pair between the seed dataset and the entire dataset (over 9 million portal message sentences) based on their deep semantic embeddings. We then selected the top 10 sentences similar to each seed sentence to annotate whether it mentioned a transportation issue. The confirmed sentences with transportation issues were included into the seed dataset for the next iteration of relevant sentence identification till the sentences in the seed dataset were enough. Finally, we collected an annotated corpus with 407 portal message sentences that mention the transportation issues during healthcare access for topic analysis.

Topic Analyses of Transportation Issues

On top of the identified portal message sentences with transportation issues, we used agglomerative hierarchical clustering method to cluster these sentences. We performed the silhouette analysis [13] to determine the optimal number of clusters for topic analysis. We manually annotated the specific topics of transportation issues mentioned in these portal message sentences.

This is the section where the authors describe the methods used at the level of detail necessary to convey the sample size, setting, procedure, datasets, analytic plan, and other relevant particulars to the reader.

Results

Summary of the Annotated Corpus and Cohort

Based on the initial seed dataset with 44 portal message sentences, we constructed an annotated corpus after four iterative expansion through searching semantically similar sentences in the entire sentence dataset in 2019. The annotated corpus contained 1,165 portal message sentences generated by 1,537 patients. Among them, 407 portal message sentences discussed transportation issues that impede access to healthcare as listed in Table 1.

Table 1 Statistics of the annotated corpus in terms of sentences, messages, and patients as well as classes

Number	Class T	Class N	Total
Annotated sentences	407	1,178	1,615
Portal messages	403	1,173	1,599
Unique patients	390	1,147	1,537

*Class T denotes that a sentence mentions a transportation issue that impede healthcare access; Class N refers to a sentence not mentioning a transportation issue or healthcare access situation.

We analyzed the distribution of different patient populations by stratifying the patient users with respect to their personal and social conditions including age, gender, marriage, ethnicity, race, language, and residence as shown in Table 2.

Table 2 Characteristics of portal users during the study period

Demographics		Class T (N=390)	Class N (N=1,172)
Age	<18	2.47	4.18
	18-29	7.14	7.73
	30-39	15.38	13.60
	40-49	20.60	14.76
	50-64	30.22	29.42
	65+	24.18	30.31
Gender	Female	67.31	64.89
	Male	32.69	35.11
Marriage	Married or Life Partner	51.37	61.61
	Not married or Legally Separated	48.63	38.39
Ethnicity	Not Hispanic or Latino	96.64	95.73
	Hispanic or Latino	3.36	4.27
Race	White	93.04	92.29
	Asian	1.95	1.90
	Black or African American	1.95	2.09
	American Indian or Alaskan Native	1.67	0.73
	Native Hawaii or Pacific Islander	0.00	0.18
	Other	1.39	2.81
Language	English	98.35	99.02
	Arabic	0.55	0.09
	Spanish	0.27	0.18
	Other	0.82	0.71
Residence	Urban	57.44	69.37
	Rural	35.38	26.54

We found that the patient users who mentioned transportation issues (class T) had a significantly different distribution from these who did not report transportation issues (class N). For the patients who are 30-64 years old, Female, not married or legally separated, White, or lives in rural area, the percentage of patients who reported transportation issues for access to healthcare was higher than those who did not mention transportation issues in patient portal messages.

Topic Analysis of the Transportation Issues

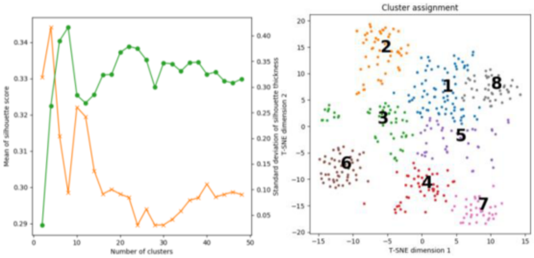

Figure 2 Determining the optimal number of clusters (left) and t-SNE scatter plot of the sentences and their cluster assignments (right)

With the collected 407 portal message sentences mentioning transportation issues for healthcare access, we performed agglomerative hierarchical clustering with the optimal number of clusters (n=8) determined by the silhouette analysis as shown in the left panel of Figure 2. The t-SNE scatter plot in the right panel of Figure 2 illustrates the spread and adjacency of the clusters.

Table 3 Topic clusters and examples of sentences mentioning transportation issues

ID	#	Description	Example
1	80	No ride	I have no ride to my appointment.
2	52	Transportation issue unspecified	I have transportation issues to get there.
3	50	Weather issue	I had to cancel my appointment due of the snow storm.
4	49	No transportation	I don't have transportation.
5	42	No ride	Similar to the cluster #1
6	51	Road issue	I cancelled my appointment due of the roads
7	37	Financial issue	I can't afford gas to even go to appointments.
8	46	Car problem	I had to cancel today appt because of car problems.

We identified 7 different transportation issues in the annotated corpus: no ride, transportation issue unspecified, weather issue, no transportation, road issue, financial hardship, and car problem. The clusters 1 and 5 have redundant themes since we scanned the number of clusters from 2 to 50 with a step of 2. The two clusters were spatially close to each other and we could further improve the clustering by given the cluster number of 7 as an input.

Discussion

Patient portals as secure online platforms allow patients to conveniently access information from their electronic health records and asynchronously interact with their providers for any medical issues including transportation barriers for access to healthcare [5]. Millions of patient portal messages offers us a great opportunity for studying transportation barriers reported by patients. In this work, we explored the feasibility of cutting-edge deep learning techniques for identifying transportation issues mentioned in patient portal messages with deep semantic embeddings as language representation. The successful creation of annotated corpus and identification of 7 transportation issues proved the feasibility of this strategy. The developed annotated corpus pave the way for developing an intelligent tool to automatically identify transportation issues from millions of patient portal messages in the future.

We characterized the patient users who reported the transportation issues with respect to their personal and social factors such as age, gender, race, and geographic location. The patients who are 30-64 years old, not married or legally separated, or rural residence have higher possibility to report transportation issues for healthcare access, compared to the general users. The single patient may experience no ride to make an appointment without spouse or life partner's support. The patients in the rural regions could have transportation issues due to long distances and lengthy times to reach needed healthcare services [3]. Studies show that transportation barriers to health care have a disproportionate impact on individuals who lives in rural communities [11].

For the annotated sentences with transportation issues, we performed topic analysis using clustering methods and identify 7 different transportation issues: no transportation, no ride, car problem, weather issue, road issue, and transportation cost. It is interesting that our method detects financial challenge, one of social determinants that implicitly correlates to transportation issues. The patient-reported transportation issues such as lack of vehicle access, inadequate infrastructure and transportation costs could offer a good picture of transportation issues about access to healthcare services. The pertinent stakeholders could make multiple strategies based on these findings to reduce transportation barriers for access to healthcare by patients.

Conclusions

The increasing volume of patient portal messages provides us a valuable opportunity for studying transportation barriers reported by patients. The successful creation of annotated corpus and identification of 7 different transportation issues showed the feasibility of our method to identify transportation issues from vast text documents with deep semantic embedding methods. The identified specific transportation issues could shed light on how to reduce transportation gaps for patients during access to healthcare. The disparity among different patient populations suggest an opportunity for reducing transportation barriers for certain patients for improving patient-centered care.

Acknowledgements

N/A

Ethics approval

No patients were exposed to any intervention. We used the data from the Mayo Clinic Unified Data Platform for analysis. The study was approved by the Mayo Clinic Institutional Review Board (19-002211).

Address for correspondence

Dr. Jungwei Fan
Department of Artificial Intelligence and Informatics
Mayo Clinic
200 1ST SW
Rochester, MN, United States, 55905
E-mail: Fan.Jung-wei@mayo.edu
Telephone: 507-778-1191

References

[1] A.L. Fitzpatrick, N.R. Powe, L.S. Cooper, D.G. Ives, and J.A. Robbins, Barriers to health care access among the elderly and who perceives them, *American journal of public health* **94** (2004), 1788-1794.

[2] G. Flores, M. Abreu, M.A. Olivar, and B. Kastner, Access barriers to health care for Latino children, *Archives of pediatrics & adolescent medicine* **152** (1998), 1119-1125.

[3] T.G. Heckman, A. Somlai, J. Peters, J. Walker, L. Otto-Salaj, C. Galdabini, and J. Kelly, Barriers to care among persons living with HIV/AIDS in urban and rural areas, *AIDS care* **10** (1998), 365-375.

[4] J. Henry, W. Barker, and L. Kachay, Electronic Capabilities for Patient Engagement among US Non-Federal Acute Care Hospitals: 2013-2017, *ONC Data Brief* **45** (2019).

[5] T. Irizarry, A.D. Dabbs, and C.R. Curran, Patient portals and patient engagement: a state of the science review, *Journal of medical Internet research* **17** (2015), e148.

[6] C.S. Kruse, K. Bolton, and G. Freriks, The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review, *Journal of medical Internet research* **17** (2015), e44.

[7] Mayo Clinic, Patient Online Services, in, 2021.

[8] NLTK Project, Natural Language Toolkit, in, 2019.

[9] D.W. Otter, J.R. Medina, and J.K. Kalita, A survey of the usages of deep learning for natural language processing, *IEEE Transactions on Neural Networks and Learning Systems* (2020), 1-21.

[10] V. Pesata, G. Pallija, and A.A. Webb, A descriptive study of missed appointments: families' perceptions of barriers to care, *Journal of Pediatric Health Care* **13** (1999), 178-182.

[11] J.C. Probst, S.B. Laditka, J.-Y. Wang, and A.O. Johnson, Effects of residence and race on burden of travel for care: cross sectional analysis of the 2001 US National Household Travel Survey, *BMC health services research* **7** (2007), 1-13.

[12] N. Reimers and I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, *arXiv preprint arXiv:1908.10084* (2019).

[13] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of computational and applied mathematics* **20** (1987), 53-65.

[14] S.T. Syed, B.S. Gerber, and L.K. Sharp, Traveling towards disease: transportation barriers to health care

access, *Journal of Community Health* **38** (2013), 976-993.

[15] L. Tieu, U. Sarkar, D. Schillinger, J.D. Ralston, N. Ratanawongsa, R. Pasick, and C.R. Lyles, Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study, *Journal of medical Internet research* **17** (2015), e275.

[16] R. Wallace, P. Hughes-Cromwick, H. Mull, and S. Khasnabis, Access to health care and nonemergency medical transportation: two missing links, *Transportation research record* **1924** (2005), 76-84.