Towards Informal Caregiver Identification in Social Media

Reda Al-Bahrani[1], Margaret K Danilovich[2] Ankit Agrawal[1] Alok Choudhary[1]
[1]Department of Electrical Engineering and Computer Science. [2]Department of Physical Therapy and Human Movement Sciences.

1 Background

With the senior population rapidly growing, caregiving is a leading public health concern. Over 30 million adults currently provide an average of 24.4 hours per week of unpaid (informal) caregiving services to an older adult in the United States [2]. By 2030, the older adult population is projected to rise by 101%, however, the number of family members available to provide informal care is only expected to rise 25% [3]. As such, there will be increasing demands and stressors on informal caregivers. Unfortunately, health care providers do not often address caregivers health. Only 16% of caregivers report that a health care provider has addressed caregiver's own personal health [1].

Twitter is an emerging tool for surveying and intervening on large groups of people in real time due to the publicly available tweets as evidenced by the use of Twitter for public health surveillance in conditions such as flu, suicide, and depression. We have previously successfully developed a real-time flu surveillance system using social media [4], where it was found that the mention of a sore throat on Twitter peaks 3-4 days before mentions of cough and fever, indicating that sore throat tweets are an early sign for a flu outbreak.

We propose an approach to identify caregivers through Twitter data mining with analysis of caregiver related content that will drive the development of a social media health promotion intervention to prevent caregiver burnout.

2 Method

Below are the steps used to fine tune our informal caregiver identification system to track caregiver tweets
1. We handcrafted a set of initial key words to track on twitter based on keywords identified through clinical expertise.
2. Through the Twitter Application Program Interface (API), we collected tweets that contain the keywords and then stored in a MySQL database. The Twitter API provides us with a sample, a random selection of tweets, about 1% of all tweets.
3. Each downloaded tweet contains information, such as user location, verified status,time stamp, in addition to the tweet contents. The metadata and contents of tweets are extracted from the raw format and store in the database.
4. We monitored the number of tweets collected per set of tracked words and with the help of clinical expertise we refined the set of tracked words to download more tweets.
5. We periodically refine the keywords and append them to the list of tracked keywords.
6. We performed basic statistics and natural language processing techniques.

3 Results

Our initial analysis revealed that identifying targeted informal caregivers is not a trivial task. Unlike the work on flu using social media[4], identifying caregivers requires a larger set of keywords resulting in a challenging task.

We have collected 1,637,257 tweets of which 886,105 are unique tweets for the period between November 22, 2015 and May 15, 2016. The majority of collected tweets fall under job postings or reference articles related to caregiving. We calculated the number of occurrence of hashtags and words in the tweets. Table 1 shows the top 10 hashtags and words. The majority of the top hashtags are related to caregiver jobs. We identified a set of words related to informal caregiver stress, and we found the top 10 co-occurring words of that set. We noticed the scarcity of tweets that fall under this category. We found mostly negative expressions of caregiving in analyzing the commonly used words in tweets. An interesting finding was that prayer was among the most frequently cited coping strategy by caregivers.

Table 1: Top 10 hashtags in the collected tweets, selected words related to informal caregiver stress, and the top 10 co-occurring words in that category.

Hashtag	Count	Word	Count	Word Pair	Count
#job	85,493	sick	14,774	sick, take	10,504
#caregiv	63,833	burden	7,530	mom, sick	9,842
#hire	23,247	stress	7,395	burden, care	7,322
#carer	10,086	hard	6,574	i'm, sick	4,919
#dementia	9,808	sad	4,096	hard, take	2,097
#healthcar	8,726	pray	2,108	hard, work	1,790
#alzheim	7,811	depress	1,953	hard, mom	1,591
#nurs	4,655	tire	1,780	burden, take	1,376
#endalz	4,399	burn	712	sick, want	1,304
#health	3,761	exhaust	475	hard, it'	1,210

We have identified several themes regarding the experience of caregiving. Caregiving content on Twitter themes are 1) caregiving advertisement or agency posts, and 2) emotional expressions.

> RT @*twitter_user*: Angry while caregiving? You are not alone. #caregiving #stress *url url* - **advertisement or agency tweet**

> Hubs: you seem stressed. Me: stressed? Why'd I be stressed? Dad has cancer, mom has dementia, I care for them 5 days a wk, cook, clean... - **emotions tweet**

> She's dying/frail + senile dementia *and* I'm trying to keep a fulltime office job. I'm burned out. Been looking into respite care for her. - **emotions tweet**

Also, using the user activity over the period of tweet collection we identified accounts with high volume of tweets which do not depict the behavior of an individual caregiver. The *source* attribute describes the application programs from which the tweet was tweeted. This information is helpful to identify individual caregivers versus caregiver agencies since caregivers are more likely to use personal device (e.g. Twitter for iPhone or Android) than business organizations.

4 Conclusions and Future Work

The development and deployment of this platform will assist in improving the health of caregivers and benefit the caregiver care recipient relationship. Also, such intervention can be easily scaled to address a comprehensive offering of health, wellness, and condition self-management interventions to all caregivers across a continuum of health. Our future work includes extending the use of natural language processing techniques and feature engineering to identify informal caregivers.

5 References

[1] Caregiving in the u.s. http://www.caregiving.org/wp-content/uploads/2015/05/2015_CaregivingintheUS_Final-Report-June-4_WEB.pdf. Accessed: 2016-05-23.

[2] Caregiver statistics: Demographics. https://www.caregiver.org/caregiver-statistics-demographics/. Accessed: 2016-05-23.

[3] Healthy aging. http://www.cdc.gov/aging/caregiving/. Accessed: 2016-05-23.

[4] Kathy Lee, Ankit Agrawal, and Alok Choudhary. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1474–1477. ACM, 2013.

978-1-5090-6117-4/16 $31.00 © 2016 IEEE
DOI 10.1109/ICHI.2016.78