Using natural language processing to extract health-related causality from Twitter messages

Son Doan
Medical Informatics
Kaiser Permanente Southern California
San Diego, CA
Son.Doan@kp.org

Elly W. Yang
Medical Informatics
Kaiser Permanente Southern California
San Diego, CA
Elly.W.Yang@kp.org

Sameer Tilak
Medical Informatics
Kaiser Permanente Southern California
San Diego, CA
Sameer.S.Tilak@kp.org

Manabu Torii
Medical Informatics
Kaiser Permanente Southern California
San Diego, CA
Manabu.Torii@kp.org

Abstract— **Twitter messages (tweets) contain various types of information, which include health-related information. Analysis of health-related tweets would help us understand health conditions and concerns encountered in our daily life. In this work, we evaluated an approach to extracting causal relations from tweets using natural language processing (NLP) techniques. We focused on three health-related topics: "stress", "insomnia", and "headache". We proposed a set of lexico-syntactic patterns based on dependency parser outputs to extract causal information. A large dataset consisting of 24 million tweets were used. The results show that our approach achieved an average precision between 74.59% and 92.27%. Analysis of extracted relations revealed interesting findings about health-related in Twitter.**

Keywords- Twitter, causal relationships, cause-effect, natural language processing (NLP)

I. INTRODUCTION

Twitter messages (tweets) are a unique public resource for monitoring health-related information, including, but not limited to, disease outbreaks, suicidal ideation and sleep issues [1]–[5]. Tweets provide diverse types of information, such as users' behaviors, lifestyles, thoughts, and experiences. Causal relations in tweets have been studied in the health domain for specific topics, such as adverse reactions caused by drugs [6], [7] or various factors causing stress and relaxation [8]. However, there has not been studies extensively investigate causal relation extraction from tweets in health-domain yet. In this study, we investigate if causes for a given health problem or concern can be extracted more generally from Twitter messages. We focus on three health-related topics: stress, insomnia, and headache.

Text mining from tweets poses various challenges [9]–[11]. One of the challenges in studying causal relations is the small fraction of relevant tweets that need to be accurately spotted in a large data collection. For the simplicity and clarity, in this study, we focus on causal relationships in explicit expressions, such as *"Excessive over thinking leads to insomnia"*. We do not consider implicit or uncertain relationships, such as "*cannot sleep #insomnia #overthinking*", where "overthinking" is not explicitly stated as the cause of insomnia.

II. METHODS

A. Dataset

We used a corpus of 24 million tweets, collected from four cities (New York, Los Angeles, San Francisco and San Diego) over 4-month period (Sep 30, 2013 and Feb 10, 2014). Twitter Streaming API was used to retrieve 1% of all the tweets from these cities during the time period. This corpus was previously used to study stress and relaxation tweets [8]. As the target "effects", we selected three terms: *stress*, *insomnia*, and *headache*.

B. Natural Language Processing (NLP) pipeline

The NLP pipeline for extracting causal relation is summarized as follows: First, the corpus is filtered using the target keywords. Next, a series of basic NLP components are applied: sentence splitter, Part-of-Speech (POS) tagger, and dependency parser. Finally, causal relations are identified based on syntactic relations generated by the dependency parser. We used CoreNLP package [12] (release version 3.8), a widely used Java library providing various NLP functionalities. The default settings and pre-trained models in the package were used for sentence splitter and POS tagger. For the parser, we selected Probabilistic Context-Free Grammar (PCFG) parser with the pre-trained English model in the package, which generates a constituent tree for an input sentence. Then, a tree is converted into a dependency graph using a CoreNLP library tool. A dependency graph consists of vertices representing tokens (words and punctuations) and edges representing dependency relations among tokens. Dependency relations are convenient for the purpose of extracting term relations in a sentence. Among several options provided for dependency graph generation in CoreNLP package, "Universal Dependencies" was used.

C. Cause-Effect Relation Extraction

We created a set of six general rules to identify cause-effect relationship from verb and noun phrase. Those rules are based on syntactic relations derived from a dependency graph generated by a dependency parser. We used CoreNLP Semgrex [13] to facilitates subgraph pattern matching over a dependency graph. For example, a Semgrex pattern "{}=subj <subj ({word: /cause/}=target >dobj {}=cause)" finds a match in a sentence

978-1-5386-6777-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICHI-W.2018.00031

"Stress caused my insomnia", where "Stress" is matched with the pattern "{}=subj" and "insomnia" is matched with the pattern "{}=cause." Using Semgrex, we extracted the triple <cause, relation, effect> from tweets, where effect is one of the three health-related topics of our focus: insomnia, stress and headache.

The final step is to extract causality from extracted cause-effect relations. To do so, we extracted the triple <cause, relation, effect>, where effect is one of the three health-related topics of our focus: insomnia, stress and headache.

III. RESULTS

We observed that the number of tweets containing specific health-related cause-effect relationships is small in comparison to the overall corpus. Specifically, the number of sentences matched by the rules is 501 from 29705 tweets for stress (1.6%), 72/3827 (1.8%) for insomnia, and 94/11252 (0.8%) for headache. The final causality extracted from the matched sentences are 41, 98 and 42 for insomnia, stress and headache, respectively.

To evaluate the precision of causal relation extraction, we compared system outputs with human annotations. Three human annotators [SD, EY, MT] annotated the system outputs in two ways: strict annotation and relaxed annotation. With strict annotation, extracted relations are considered correct only when an observed cause is clearly and explicitly stated. In relaxed annotation, negated or hypothetical statements are additionally considered as correct extraction. For example, a tweet *"Cell phone radiation can cause insomnia"* reports possibility, and was annotated as false positive in strict annotation, but true positive in relaxed annotation. Disagreement among annotators were resolved through discussion.

The precision is calculated by the number of true positives annotated by human annotators divided by the number of tweets system found. The micro-average is calculated by a sum of all true positives across all three categories divided by the total tweets reviewed.

Table I shows the precision when comparing system outputs to human annotations. It shows that the micro-average for strict and relaxed annotation is 74.59% and 92.27%, respectively. It also suggests that finding causal relationships for "headache" is more difficult than "insomnia" and "stress". The large variations of strict and relaxation evaluation (74.59% vs. 92.27%) indicates that hypothetical statements, negated expressions, and other such subtle expressions play important roles in Twitter messages.

IV. CONCLUSIONS

In this paper, we presented a general NLP approach to extracting cause-effect relationships from Twitter messages. While preliminary, the results on four months Twitter data showed favorable precision and revealed subtly different expressions for three target health-related topics. The proposed approach can be extended into other data such as research literature , and clinical notes. Further investigation is planned to extend our work.

TABLE I. PRECISION OF EXTRACTED CAUSAL RELATIONS WHEN COMPARING TO HUMAN ANNOTATORS.

	Strict evaluation	**Relax evaluation (exclude hypothetical and negation)**
Insomnia	73.81%	88.10%
Stress	82.65%	96.94%
Headache	56.10%	85.37%
Micro-average	74.59%	92.27%

REFERENCES

[1] N. Collier and S. Doan, "Syndromic classification of Twitter messages," in *Procs. of 4th ICST International Conference on eHealth*, 2011, pp. 186–195.

[2] M. J. Paul, M. Dredze, and D. Broniatowski, "Twitter improves influenza forecasting.," *PLoS Curr.*, vol. 6, Oct. 2014.

[3] H. Sueki, "The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan.," *J. Affect. Disord.*, vol. 170, pp. 155–60, Jan. 2015.

[4] D. J. McIver, J. B. Hawkins, R. Chunara, A. K. Chatterjee, A. Bhandari, T. P. Fitzgerald, S. H. Jain, and J. S. Brownstein, "Characterizing Sleep Issues Using Twitter.," *J. Med. Internet Res.*, vol. 17, no. 6, p. e140, Jun. 2015.

[5] A. Stefanidis, E. Vraga, G. Lamprianidis, J. Radzikowski, P. L. Delamater, K. H. Jacobsen, D. Pfoser, A. Croitoru, and A. Crooks, "Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts.," *JMIR public Heal. Surveill.*, vol. 3, no. 2, p. e22, Apr. 2017.

[6] A. Cocos, A. G. Fiks, and A. J. Masino, "Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts.," *J. Am. Med. Inform. Assoc.*, vol. 24, no. 4, pp. 813–821, Jul. 2017.

[7] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on twitter? Mining tweets for adverse drug reactions.," *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2014, pp. 924–33, 2014.

[8] S. Doan, A. Ritchart, N. Perry, J. D. Chaparro, and M. Conway, "How Do You #relax When You're #stressed? A Content Analysis and Infodemiology Study of Stress-Related Tweets.," *JMIR public Heal. Surveill.*, vol. 3, no. 2, p. e35, Jun. 2017.

[9] S. Doan, L. Ohno-Machado, and N. Collier, "Enhancing Twitter Data Analysis with Simple Semantic Filtering: Example in Tracking Influenza-Like Illnesses," in *IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB 2012)*, 2012, no. 1, pp. 62–71.

[10] A. Culotta, "Towards detecting influenza epidemics by analyzing Twitter messages," in *KDD Workshop on Social Media Analytics*, 2010, no. May, pp. 1–11.

[11] D. Mowery, H. Smith, T. Cheney, G. Stoddard, G. Coppersmith, C. Bryan, and M. Conway, "Understanding Depressive Symptoms and Psychosocial Stressors on Twitter: A Corpus-Based Study.," *J. Med. Internet Res.*, vol. 19, no. 2, p. e48, Feb. 2017.

[12] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014, pp. 55–60.

[13] N. Chambers, C. D. Manning, D. Cer, T. Grenager, D. Hall, C. Kiddon, B. MacCartney, M.-C. de Marneffe, D. Ramage, and E. Yeh, "Learning alignments and leveraging natural logic," *Proc. ACL-PASCAL Work. Textual Entailment Paraphrasing - RTE '07*, no. June, p. 165, 2007.