Volunteer-powered Automatic Classification of Social Media Messages for Public Health in AIDR

Muhammad Imran
Qatar Computing
Research Institute
Doha, Qatar
mimran@qf.org.qa

Carlos Castillo
Qatar Computing
Research Institute
Doha, Qatar
chato@acm.org

ABSTRACT

Microblogging platforms such as Twitter have become a valuable resource for disease surveillance and monitoring. Automatic classification can be used to detect disease-related messages and to sort them into meaningful categories. In this paper, we show how the AIDR (Artificial Intelligence for Disaster Response) platform can be used to harvest and perform analysis of tweets in real-time using supervised machine learning techniques. AIDR is a volunteer-powered online social media content classification platform that automatically learns from a set of human-annotated examples to classify tweets into user-defined categories. In addition, it automatically increases classification accuracy as new examples become available. AIDR can be operated through a web interface without the need to deal with the complexity of the machine learning methods used.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.2 [**Software Engineering**]: Design Tools and Techniques

Keywords

Stream processing, Crowdsourcing, Classification, Epidemics

1. INTRODUCTION

In recent years, methods for monitoring disease out-breaks using using machine learning techniques on Twitter data have been tested. These methods are typically based on content classification or clustering [6, 2], and have been validated by contrasting them with data from health agencies, e.g. by [1].

Real-time information processing to find specific content within rapidly growing stacks of information pose many challenges. For instance, applying content classification techniques on short bursts of text (e.g., on 140-character "tweets") is significantly more difficult than performing the same task on large documents such as blog posts or news articles [5]. Moreover, relying only on machines cannot guarantee accuracy, at some point human-intelligence is required to provide training examples to automatic algorithms, or for the tasks that can not be automated.

This paper shows how the AIDR[1] (Artificial Intelligence for Disaster Response) platform can be used to classify Twitter messages for public health monitoring. AIDR is an easy-to-use web-based platform designed to allow analysts to collect and classify microblog posts. More specifically, AIDR classifies tweets into a set of user-defined categories of information. The platform continuously ingests data, processes it (i.e., using machine learning classification techniques), and leverages human intelligence (with the help of volunteers), when required [3].

As a proof of concept, in the next section we show how AIDR can be used to rapidly set up a flu detector on Twitter. We break down various steps involved in setting up the platform for this task. Moreover, we show how volunteers can be involved, and with the help of active learning, help to increase classification accuracy as time passes.

2. MONITORING FLU-RELATED TWEETS USING AIDR

AIDR operates in three steps: (i) information collection, (ii) acquisition of human-tagged examples, and (iii) automatic classification of messages.

In this example, we first use AIDR to capture all messages that are posted on Twitter containing the keyword "flu" or having the hashtag "#fluseason" (other user-defined terms can be used). Certainly, this approach introduces noise in our collection, i.e., messages which contain our query terms, but do not help in enhancing situational awareness [7, 4] about flu. Next, we set up an automatic classifier to filter out and discard those messages. Figure 1 depicts AIDR's user interface, (a) the collector definition and (b) the crowdsourcing interface.

2.1 Collecting candidate messages

AIDR exposes a web-based interface to define a search query including a set of keywords and a language (e.g. English). Within 7 hours, after starting our AIDR collector, we were able to collect around 25K tweets. Looking at the collected tweets, we can clearly see that many do not enhance situational awareness about flu:

[1] `http://aidr.qcri.org`

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579279.

- *Dear sir dr, antibiotic is not candy and flu-ing man is not crying girls*
- *Oh hell ya got my exam rescheduled till Tuesday. Only up side about having the flu lol*
- *Mornin' , I might get hit upside the head for this but I hope this flu Mr.Taylor has last longer so we can have more subs*

While, for instance, the following messages do enhance situational awareness:

- *According to the California Department of Public Health, we are experiencing a severe flu season. Remember to... http://.../*
- *People urged to get vaccinated as coming flu epidemic predicted – The Connexion http://.../*

2.2 Acquiring human-tagged examples

The next step is to involve humans to tag a sub-set of messages. The AIDR user creates the categories (in our case "Enhances situational awareness about flu" and "Does not enhance situational awareness about flu"). Then, AIDR creates a crowdsourcing task in the Pybossa platform[2], which is an open-source crowdsourcing system. A team of volunteers can then proceed to do the tagging. In the default setting, AIDR requires 3 volunteers to tag a message, and messages are assigned with a category once 2 out of 3 volunteers agree. The resulting messages are passed back to AIDR.

2.3 Enabling automatic classification of messages

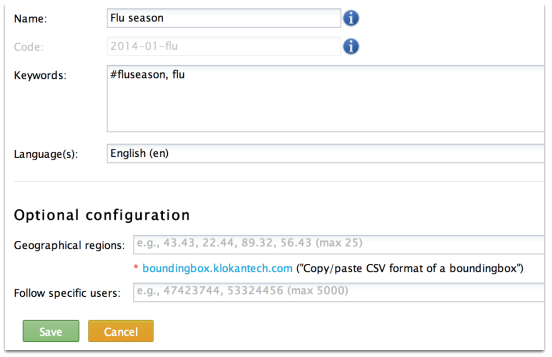
In our proof-of-concept, tagging was done by one of the authors of this paper. After receiving approximately 50 tagged messages, AIDR automatically creates a classifier. For testing purposes, AIDR considers 25% of the tagged examples as a test set, and rest of the examples as a training set. During these experiments, we achieved ≈75% classification accuracy (measured using AUC) using 200 labels, that also depends on the complexity of categories.

We remark that many other classification taxonomies can be used. For instance, once that messages enhancing situational awareness are identified, they can be further subdivided into e.g. those posted by agencies vs. those posted by individual users, those describing symptoms experienced vs. those describing treatment options used, and so on.

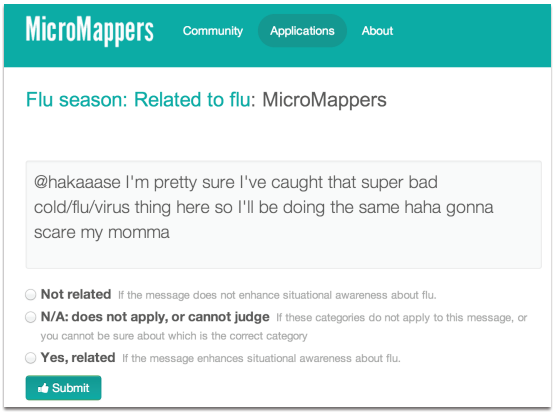
3. CONCLUSION

Microblog messages can provide valuable information for public health purposes. The volume and speed of the data means automatic methods are necessary to process it. The complexity and brevity of the messages means human intervention is usually necessary. Supervised machine learning methods powered by crowdsourcing labels provide an excellent combination of the best of human and machine intelligence. However, setting up them can be difficult for the non-expert. AIDR lowers the barrier of entry by providing a simple web interface where the analyst can quickly set-up a collection and automatic classification system. The short time needed to set-up this up, and the flexibility provided by this application, may lead to actionable information earlier than if this were to be set-up by other means.

[2]`http://pybossa.qcri.org/`



(a) Collector definition



(b) Crowdsourcing interface

Figure 1: User interface of AIDR.

4. REFERENCES

[1] T. Bodnar and M. Salathé. Validating models for disease detection using twitter. In *Proc. PHDA workshop*, 2013.

[2] C. D. Corley, D. J. Cook, A. R. Mikler, and K. P. Singh. Text and structural data mining of influenza mentions in web and social media. *Int. J. of Environmental Research and Public Health*, 7(2):596–615, 2010.

[3] M. Imran, C. Castillo, J. Lucas, M. Patrick, and V. Sarah. AIDR: Artificial intelligence for disaster response. In *Proc. WWW (Demos)*. ACM, 2014.

[4] M. Imran, S. M. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. Extracting information nuggets from disaster-related messages in social media. In *Proc. of ISCRAM*, Baden-Baden, Germany, 2013.

[5] C. Li, J. Weng, Q. He, Y. Yao, A. Datta, A. Sun, and B.-S. Lee. TwiNER: named entity recognition in targeted twitter stream. In *Proc. of SIGIR*, pages 721–730. ACM, 2012.

[6] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson. Identifying health-related topics on twitter. In *Social computing, behavioral-cultural modeling and prediction*, pages 18–25. Springer, 2011.

[7] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proc. of SIGCHI*, pages 1079–1088. ACM, 2010.