AIDR: Artificial Intelligence for Disaster Response

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

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

Ji Lucas
Qatar Computing
Research Institute
Doha, Qatar
jlucas@qf.org.qa

Patrick Meier
Qatar Computing
Research Institute
Doha, Qatar
patrick@irevolution.net

Sarah Vieweg
Qatar Computing
Research Institute
Doha, Qatar
svieweg@qf.org.qa

ABSTRACT

We present AIDR (Artificial Intelligence for Disaster Response), a platform designed to perform automatic classification of crisis-related microblog communications. AIDR enables humans and machines to work together to apply human intelligence to large-scale data at high speed.

The objective of AIDR is to classify messages that people post during disasters into a set of user-defined categories of information (e.g., "needs", "damage", etc.) For this purpose, the system continuously ingests data from Twitter, processes it (i.e., using machine learning classification techniques) and leverages human-participation (through crowdsourcing) in real-time. AIDR has been successfully tested to classify informative vs. non-informative tweets posted during the 2013 Pakistan Earthquake. Overall, we achieved a classification quality (measured using AUC) of 80%. AIDR is available at `http://aidr.qcri.org/`.

Categories and Subject Descriptors

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

Keywords

Stream processing; Crowdsourcing; Classification; Online Machine learning

1. INTRODUCTION

Information overload during disasters can be as paralyzing to humanitarian response as the absence of information. During disasters, microblogging platforms like Twitter receive an overwhelming amount of situation-sensitive information that people post in the form of textual messages, images, and videos. Despite the fact that social media streams contain a significant amount of noise, much research [9, 4] has shown that these same streams of information also include relevant, tactical information (e.g., regarding infrastructure damage, needs, donations). Because social media communications provide a rich trove of information, it is possible that even a small amount of relevant information can greatly enhance situational awareness and help responders and other concerned parties make more informed decision.

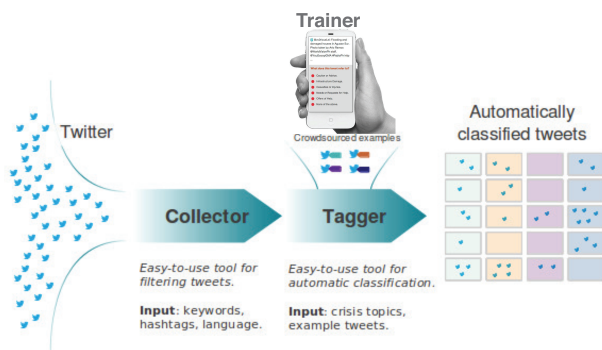

Figure 1: AIDR: overall approach

Finding tactical and actionable information in real time within a rapidly growing stack of information is challenging for many reasons. For instance, performing information extraction 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 pr news articles [6]. Moreover, research has shown that pre-trained classifiers significantly drop in classification accuracy when used in different but similar disasters [3]. This requires learning and training new classifiers using fresh training data every time a disaster strikes.

Considering the amount of information that flows on Twitter, it is challenging for emergency managers and other stakeholders to investigate each individual tweet in real-time to

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577034.

look for useful information. Therefore, our goal is to leverage different machine learning techniques (e.g., information classification, and extraction) to perform the job automatically. Moreover, we want humans (i.e. volunteers) to label part of the incoming data to be used for the training purposes of machine learning algorithms. Above all, the whole process must be ingesting, processing and producing only credible information in *real-time*, or with low latency [5].

The rest of the paper is organized as follows: In the next section, we describe domain challenges in crisis response. In section 3, we present an overview of AIDR from an end-user perspective, as well as an evaluation. Section 4 presents AIDR's architecture and implementation. A demonstration storyboard is described in section 5, followed by the conclusion in section 6.

2. DOMAIN CHALLENGES IN CRISIS RESPONSE

During disasters, social media messages provide real-time or low-latency situational awareness information that can enable crisis responders to be more effective in their relief efforts [8]. However, different emergency response agencies are interested in different types of messages. For instance, reports of damage to infrastructures should be directed to some agencies, while reports about shortages of water and/or food should be directed to others.[1]

Moreover, disaster response in its various types can be applied during the various phases of disaster such as preparation, response and recovery. During each phase, disaster responders require different information. In our previous work [1], we observed that social media response to disasters follows the same pattern, that is, messages posted on social media during the early phases of a disaster talk about caution & warnings, whereas messages posted during the later phases report infrastructure damage, casualties, donations required or available, etc.

Below, we discuss the roles of automatic computation, human computation, and the combination of the two in the processing of social media streams.

Role of machine intelligence: Traditional information processing cannot be employed in this model, as disaster responders cannot wait to collect information, and then curate and classify it offline. Instead, responders and other stakeholders require real-time insight and intelligence as the disaster unfolds. To this end, we aim to ingest and classify social media streams in real-time through automated means with the help of human intervention.

Role of human intelligence: When attempting to perform non-trivial tasks, machines alone are not capable of great accuracy. Human intervention is required to verify, teach, and/or correct the machine output [2]. Use of human intelligence fills the gap for the tasks that cannot be automated, for example, providing input labels (i.e., for initial training), correcting or validating the machine's output (i.e., for performance optimization) are among the types of human interventions. In AIDR, we aim to find a right balance so that the human intelligence can be used in an effective way.

Combined intelligence: Relying solely on humans to investigate each individual message is challenging due to the scale of information posted on Twitter, which goes beyond the processing capacity of humans. To this end, an automatic approach is required that can intelligently crowdsource messages to obtain training examples when needed, and additionally, the system should effectively use crowdsourcing workers both in terms of time (i.e., for volunteers) and cost (i.e., for paid workers).

[1]The United Nations organizes its agencies into clusters: `http://business.un.org/en/documents/6852`.

3. SYSTEMS OVERVIEW

The purpose of AIDR (Artificial Intelligence for Disaster Response),[2] is to filter and classify messages posted to social media during humanitarian crises in real time.

Specifically, AIDR collects crisis-related messages from Twitter[3] ("tweets"), asks a crowd to label a sub-set of those messages, and trains an automatic classifier based on the labels. It also improves the classifier as more labels become available. Automatic classification using pre-existing training data is not a satisfactory solution because although crises have elements in common, they also have specific aspects which make domain adaptation difficult. Crisis-specific labels lead to higher accuracy than labels from past disasters [3].

3.1 AIDR in action: end-user perspective

AIDR users begin by creating a collection process by entering a set of keywords or a geographical region that will be used to filter the Twitter stream, as shown in Figure 2(a). The user can monitor the collection status (e.g., total processed items, last processed item, time elapsed, etc.) using dashboard as shown in Figure 2 (b). Next, a crowd of annotators provide training examples: a system-selected message plus a human-assigned label, as shown in Figure 2(c), which are then used to train classifiers for incoming items, as shown in Figure 2(d).

Finally, an output of messages sorted into categories is generated, which can be collected and used to create crisis maps and other types of reports. An example consumer application is the current version of CrisisTracker,[4] which uses AIDR to enable users to slice the data by categories of interest, which vary by deployment scenario to include for instance eyewitness accounts, reports of violence, or reports of damage infrastructure.

3.2 Evaluation

AIDR was successfully tested during a recent earthquake in Pakistan in 2013. We set AIDR up to collect tweets using the hashtags (#Pakistan, #Awaran, #Balochistan, #earthquake, #ReliefPK) on September 25, 2013 at 20:20:09 AST[5] on a request of UN Office for the Coordination of Humanitarian Affairs (OCHA). Within a few hours, SBTF (Standby Task Force) [6] volunteers were asked to label whether a given tweet was informative (i.e., if the tweet reports infrastructure damage, casualties, donation offered or needed, etc.). They tagged about 1,000 tweets approximately within 6 hours. Though the prevalence of the negative class ("not informative") was high, the system was able to learn from ≈200 informative labeled tweets. In this setup, we achieved

[2]`http://aidr.qcri.org/`
[3]`http://twitter.com/`
[4]`http://ufn.virtues.fi/~jakob/yolanda/`
[5]Arabian Standard Time
[6]`http://blog.standbytaskforce.com/`

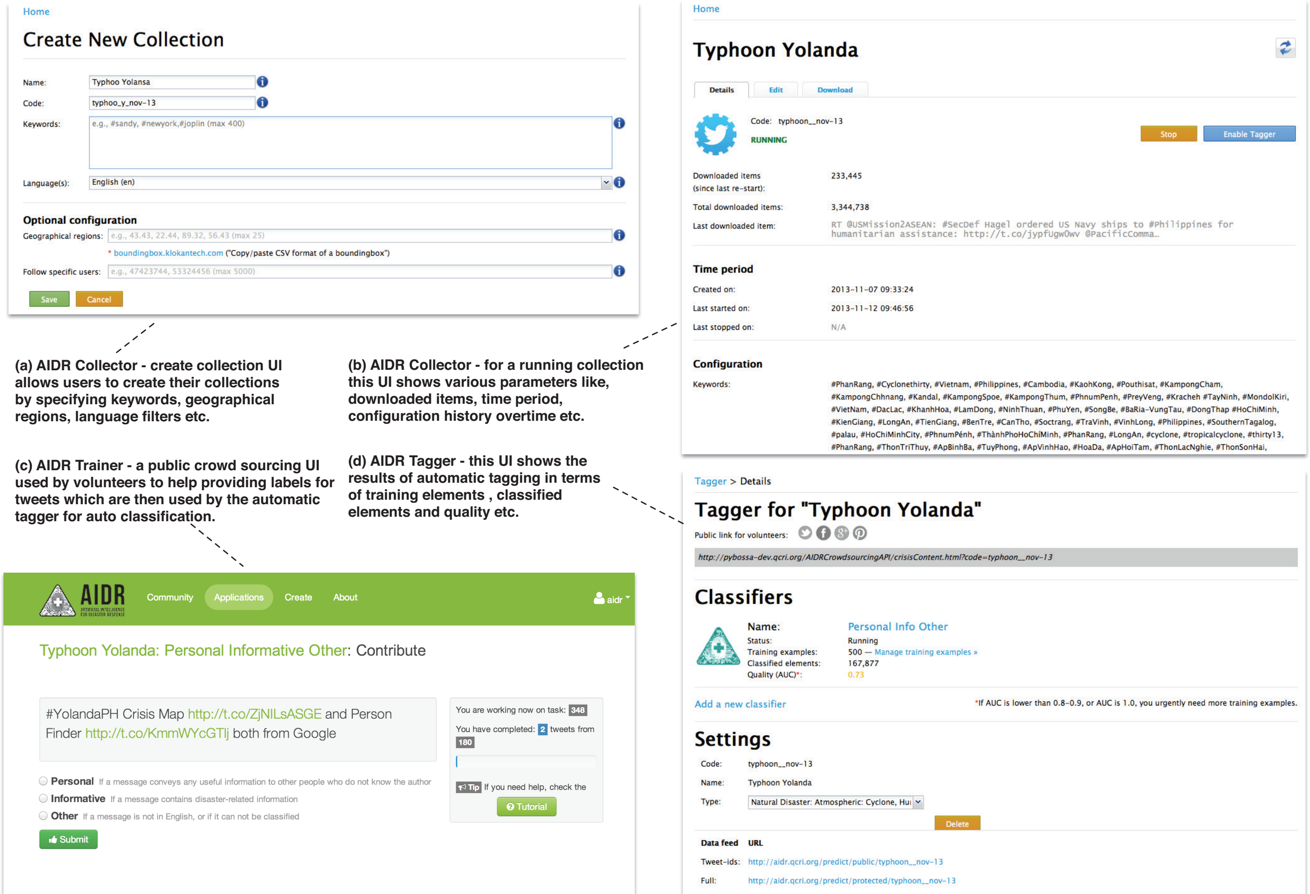

Figure 2: AIDR Screenshots: showing (a) collector (creation screens), (b) collector (monitoring screen), (c) trainer, and (d) tagger.

a maximum classification quality (in terms of AUC) up to 80%. AIDR success during the initial tests was featured by Wired UK[7] on 30 September 2013, and by CBC[8] on 18 December 2013.

4. ARCHITECTURE & IMPLEMENTATION

The general architecture of AIDR is shown in Figure 3. AIDR is a free software platform that can be run as a web application, or downloaded to create your own instance.[9] It consists of three core components; collector, tagger, and trainer. The collector performs edge adaptation [7], and is responsible for data collection. For instance, in our current setup it collects messages from Twitter using the Twitter streaming API. The collected tweets are then passed to the tagger for further processing. The tagger is responsible for the classification of each individual tweet. The tagger is comprised of three modules: feature extractor, learner, and classifier. First, the feature extractor receives a tweet, it extracts features (e.g., uni-grams and bi-grams), and passes it to the classifier. Second, the classifier's job is to assign one of the user-defined categories (e.g., donations, damage, casualties, etc.) to the tweet. To do so, the classifier uses the learner module, which requires sufficient training examples to learn about each user-defined category.

The training examples required by the system can be obtained either using internal web-based interface or by calling an external crowdsourcing platform. The former aims at enabling the collection owner to provide trusted training examples, whereas the latter collects training examples using public crowdsourcing with the help of volunteers. We assume that there is a fixed budget of crowdsourcing work, but even if that is not the case, we see this as a problem of cost effectiveness. To ensure quality, training examples are obtained in a way that maximizes marginal quality gains per human label. The maximization of quality gains per label is done by performing intelligent task generation by selecting a small set of messages to be labeled by humans. Details on AIDR crowdsourcing part and task generation strategies are discussed in detail in our additional research [2].

The output of AIDR (i.e., classified tweets) can be accessed through output adapters, which are exposed as an API. To show real-time classified items on a map or any other visualization widget, one can use AIDR's live stream output adapter. Moreover, to fulfill various visualization demands, AIDR includes APIs to retrieve the k-latest items or to subscribe to a live data feed.

---
[7] http://www.webcitation.org/6N9iZuG1E
[8] http://fw.to/QMOLqnl
[9] https://github.com/qcri-social/AIDR

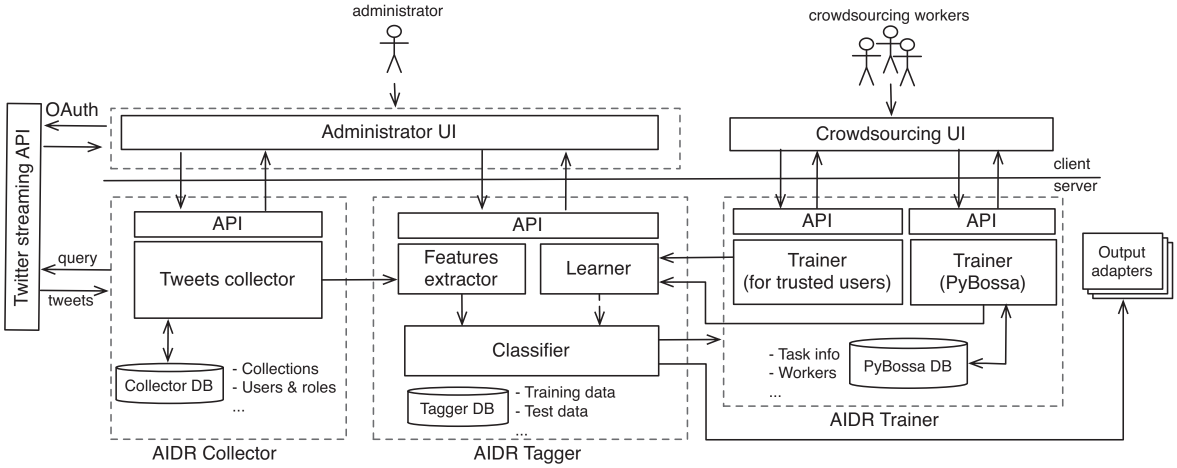

Figure 3: AIDR architecture shows Collector, Trainer, and Tagger

Implementation: AIDR comprised of a client-side and three server-side applications. Mainly, the client-side application has been developed using the Sencha ExtJS framework[10], and the server-side implementation is developed using Java and the Springs 3.0 framework for the main application logic. We use PyBossa for the crowdsourcing processing purposes, and REDIS[11] for the communication flows. AIDR is an open-source platform, and its source code is available at this repository[12].

5. DEMONSTRATION

A live demo will be presented starting from an introduction of the crisis computing domain and motivation behind the development of AIDR platform. A guided walk-through of the platform will be presented to introduce how different components of AIDR work. After demonstrating how to create collections, perform training, and enable an automatic classification process, we ask our reader to try the tool and create their own collection and perform classification without using any knowledge of machine learning.

6. CONCLUSIONS

Social media platforms like Twitter receive an overwhelming amount of situational awareness information. For emergency response, real-time disaster insights are important. Finding actionable and tactical information in real-time poses serious challenges. Effective coordination of human and machine intelligence can improve disaster response efforts. In this paper, we have described AIDR, a platform to classify Twitter messages into a set of user-defined situational awareness categories in real-time. The platform combines human and machine intelligence to obtain labels of a subset of messages and trains an automatic classifier to classify further posts. The platform uses active learning approach to select potential messages to tag, and learns continuously to increase classification accuracy when new training examples are available.

[10] `http://www.sencha.com/products/extjs/`
[11] `http://redis.io/`
[12] `https://github.com/qcri-social/AIDR`

7. REFERENCES

[1] S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, and C. Castillo. Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *Proc. of ISCRAM*, Baden-Baden, Germany, 2013.

[2] M. Imran, C. Castillo, J. Lucas, M. Patrick, and J. Rogstadius. Coordinating human and machine intelligence to classify microblog communications in crises. *Proc. of ISCRAM*, 2014.

[3] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. Practical extraction of disaster-relevant information from social media. In *Proc. of Workshop on Social Media Data for Disaster Management*, WWW '13 Companion, pages 1021–1024. ACM/IW3C2, 2013.

[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] M. Imran, I. Lykourentzou, and C. Castillo. Engineering crowdsourced stream processing systems. *arXiv preprint arXiv:1310.5463*, 2013.

[6] 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 *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 721–730. ACM, 2012.

[7] D. Turaga, H. Andrade, B. Gedik, C. Venkatramani, O. Verscheure, J. D. Harris, J. Cox, W. Szewczyk, and P. Jones. Design principles for developing stream processing applications. *Software: Practice and Experience*, 40(12):1073–1104, 2010.

[8] S. Vieweg. Microblogged contributions to the emergency arena: Discovery, interpretation and implications. In *Proc. of CSCW*, February 2010.

[9] S. E. Vieweg. Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications. 2012.