Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions

Kate Starbird
University of Colorado, Boulder
starbird@colorado.edu

Grace Muzny
University of Washington, Seattle
muznyg@cs.washington.edu

Leysia Palen
University of Colorado, Boulder
palen@cs.colorado.edu

ABSTRACT

Social media tools, including the microblogging platform Twitter, have been appropriated during mass disruption events by those affected as well as the digitally-convergent crowd. Though tweets sent by those local to an event could be a resource both for responders and those affected, most Twitter activity during mass disruption events is generated by the *remote* crowd. Tweets from the remote crowd can be seen as noise that must be filtered, but another perspective considers crowd activity as a filtering and recommendation mechanism. This paper tests the hypothesis that crowd behavior can serve as a collaborative filter for identifying people tweeting from the ground during a mass disruption event. We test two models for classifying on-the-ground Twitterers, finding that machine learning techniques using a Support Vector Machine with asymmetric soft margins can be effective in identifying those *likely to be on the ground* during a mass disruption event.

Keywords

Crisis informatics, human computation, machine learning, mass disruption, microblogging, political protest, support vector machine

INTRODUCTION

Access to social media has created "sites" of interaction where people digitally converge during disasters (Palen & Liu, 2007) and other mass disruption events, including political protests (Grossman, 2009; Lotan et al., 2011). Digital convergers generate massive amounts of data—e.g. millions of tweets were sent referencing the 2010 Haiti earthquake during the early aftermath of that event (Anderson & Schram, 2011). Though a portion of these data come directly from the ground in the form of citizen reports from affected people or relayed by "proxy" accounts (Sarcevic et al., 2012), a majority of these communications are *derivative*—that is, information in the form of reposts or pointers to information available elsewhere (Starbird et al., 2010).

Derivative information is so abundant that it is commonly viewed as a form of *noise* that must be filtered out to arrive at the *signal* of good data. An alternative perspective is to view derivative information as a valuable part of the information ecosystem that can be treated as meta-data, or *information about information* that might be able to provide a road map for navigating the noisy information space, even as it simultaneously contributes to that noise. Starbird & Palen (2010) claim that one feature of derivative information—specifically the repost or *retweet* on the Twitter platform—can be viewed as a recommendation mechanism. This paper extends that view, examining the retweet along with other features of crowd behavior on Twitter, including *following* and *list-making*, as individual actions of recommendation that collectively shape the information space.

Collaborative filtering is a technique for extracting meaning from the aggregate behavior of a large number of users. In previous research, we quantified differences in how the Twitter crowd recommends locals and non-locals during a mass disruption event—the political protests in Egypt in 2011 (Starbird & Palen, 2012). That paper concluded that crowd behavior and individual behavior could serve as a collaborative filter for identifying people tweeting from the ground during a mass disruption event. The aim of the work reported in this paper is to test this hypothesis using models from machine learning techniques.

BACKGROUND

This paper examines the use of social media during the Occupy Wall Street (OWS) political protest as it began in New York City in September 2011. We characterize this protest as a *mass disruption event*—an event affecting a large number of people that causes disruption to normal social routines. Examples of mass disruption events include natural disasters, acts of terrorism, mass emergencies, extreme weather events and political protests. The emerging field of *crisis informatics* (Hagar & Haythornthwaite, 2005; Palen et al., 2010) is building a foundation for understanding online behavior and other aspects of social computing during mass disruption events. Crisis informatics can be defined broadly as the study of the social, technical and informational concerns of large-scale emergency response; it considers the interactions and concerns of formal responders as well as members of the public (Palen et al., 2010).

Social Media and Mass Disruption Events

Social media platforms have consistently been appropriated by current members and adopted by new users during and after mass disruption events (Hughes and Palen, 2009; Shklovski et al, 2010). Users turn to these platforms to participate in a wide range of information activities, e.g. to share information (Starbird et al., 2010; Heverin & Zach, 2010)—an activity also known as citizen journalism (Gillmor, 2004), to seek information about the status of people (Qu et al., 2011) or property (Shklovski et al, 2008), to gather and synthesize information (Qu et al., 2009), to seek or offer assistance (Palen & Liu, 2007; Vieweg et al., 2010; Qu et al., 2011; Starbird & Palen, 2011; Mark et al., 2012), and to coordinate action (Qu et al., 2009; Starbird & Palen, 2011; Sarcevic et al., 2012). These activities represent a digital-age equivalent to the informational convergence behavior long known to occur in the wake of disaster events (Fritz & Mathewson, 1957; Dynes 1970).

Twitter and Crisis

Starting not long after its inception, the microblogging platform Twitter has consistently been appropriated for use during mass disruption events by those affected (Messina, 2007), digital volunteers (Starbird & Palen, 2011), and emergency response organizations (Sarcevic, et al., 2012). Its appeal comes from its short message, broadcast, public nature: most posts can be seen by anyone which means that interactions are not "walled" away to a restricted group. As such, we find that Twitter is a place where information converges from across the Internet, serving often as a way-finding resource to places where additional interactions are happening. As a result, it is a valuable research site because it helps organize the otherwise boundless space of the Internet.

Specifically, Twitter is a socially-networked social media platform that allows users to broadcast 140-character messages (tweets) to their followers, and to receive broadcasted messages from users they choose to follow. Twitter users (Twitterers) maintain an account profile with information they provide including account name, user name, account description and location. Users can also make Twitter lists of other Twitterers, grouped by conversation topic or some other user-defined classification, and publicize these lists to other users to "follow." Additionally, all tweets broadcast from public accounts and all profile information associated with these accounts, including follower and following counts and usernames as well as user-provided information, are available for public search through application programming interfaces (APIs) made available by Twitter, a feature that permits large-scale information sharing and diffusion during mass disruption events.

Throughout Twitter's young life, users have introduced and adopted linguistic conventions to adapt the platform to support their needs. These include the retweet mechanism (`RT @username`) to permit message forwarding with upstream author attribution (boyd et al., 2010) and the hashtag (`#keyword`) to support information search and group formation (Messina, 2007).

Why Identifying "Locals" Among Social Media Users is Important

Though only a small portion of tweets contain information from local Twitterers coming into the space for the first time (Starbird et al., 2010), this information can be a valuable resource for emergency responders, event planners, affected people, journalists, and the digitally converging crowd. People who are on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space. Additionally, locals may have knowledge about geographic or cultural features of the affected area that could be useful to those responding from outside the area.

Existing research has recognized a distinction between local and non-local social media users. In their investigation of how situational awareness information appears in tweets, Vieweg et al. (2010) focused on a sample of accounts that were local to the event. Starbird and Palen (2010) found that people were more likely to

retweet accounts of those who were local to the event, even when local accounts only made up a small portion of event-related tweets. Lotan et al. (2011) argue that during the protests in Tunisia, activists tweeting from the ground provided a valuable information source for journalists, who would in turn often retweet activists' posts.

In the context of political disruptions, the value of locating information coming from the ground and identifying individuals who are on the ground is less straightforward. Whereas during crisis events there is typically one "side" in the response and the great majority of people are working to help those affected, during political disruption there are often two or more "sides": the protesters, the protested, and law enforcement groups. In these cases, it could be argued that identifying those on the ground could be detrimental to their cause. Along these lines, Burns & Eltham (2009) note that during the Iran Election protests in June 2009, social media may have been used most effectively by pro-government forces in their efforts to crush opposition protests, even putting protestors in more danger by allowing them to be more easily identified.

Though accepting some risk in identifying on-the-ground participants in this way, this research contends that the great majority of Twitter users during mass disruption events, even political protests, understand that their communications are public and want their accounts be identified with the protest and their voices to be heard. In the case of the Egypt protests and the OWS protests, there is evidence that individuals on the ground wanted the world to know they were there. In some cases, they seemed to try to garner media attention and foster solidarity with their cause. In both events, the platform was also used by on-the-ground Twitterers to send out requests for basic assistance:

> `@occupiedcairo: Food and Medicine needed at Tahrir. URGENT #jan25`
>
> `@jeffrae: We could really use a generator down here at Zuccotii Park. Can anyone help? #occupyWallStreet #takewallst #Sept17`

We also see evidence that digital volunteers supportive of the Occupy Wall Street cause were manually attempting to identify on-the-ground Twitterers by creating public Twitter lists of those accounts. These tweets illustrate how Twitterers advertised these lists, and also asked for help finding more on-the-ground accounts:

> `@CassProphet: Follow on-scene @AACina @Jeffrae @DhaniBagels @Korgasm_ @brettchamberlin #TakeWallStreet #OurWallStreet #OccupyWallStreet #yeswecamp`
>
> `@djjohnso: Live tweeter curated list for #OccupyWallStreet #BeatTheBell #TakeWallStreet @djjohnso/occupywallstreetlive`
>
> `@djjohnso: We have 20 livetweeters for this list. Are there others? @djjohnso/occupywallstreetlive #takewallstreet #OurWallStreet #needsoftheoccupiers`

Collaborative Filtering

Twitterers who are on the ground during mass disruptions can be valuable sources of information during these events, but there remains the challenge of identifying these local Twitterers within the vast and noisy information space. This research explores the possibility of using the noise to find the signal, employing collaborative filtering techniques to identify people who are on the ground during mass disruption events.

Collaborative filtering is a technique for using the individual and collective actions, both explicit and implicit, of a large number of people within an interaction space to filter information produced by that same group (Malone et al., 2009). Mendoza et al. (2010) report evidence that the social media community can collaboratively act to identify bad information. Studying the propagation of rumors through the Twitterverse in the wake of the Chile Earthquake in 2010, they found that tweets containing false information were more likely to be challenged by other Twitterers. Kwak et al. (2010) assert that social context, which consists of social interactions within social media (friend relationships, group membership, lists), can work as a collaborative filter to identify the value of information. Starbird and Palen (2010) describe how Twitterers use the retweet as a recommendation mechanism during crisis events.

Crowd Work during Egypt Protests

In our previous work (Starbird & Palen, 2012) we reported that certain characteristics of crowd behavior could act as a collaborative filter for identifying people tweeting from the ground during a mass disruption event. Through an empirical investigation of crowd behavior, we noted significant differences between how the crowd acts to recommend individuals who are present on the ground during a mass disruption event versus those who are not.

That study focused on Twitter use during the political unrest in Egypt in the first half of February 2011. For that investigation we worked from the assumption that retweets can be used as a recommendation mechanism. Using Twitter data collected using popular protest-related hashtags, we then identified the 1000 most highly retweeted

accounts. Applying content analysis to a sample of those accounts, we coded each as to whether the account owner was on the ground in protests in Cairo or not, finding nearly 30% of highly retweeted Twitterers were physically present at those protest events. Then, using statistical analysis, we found that 1) the total number of times an account was retweeted as well as 2) the number of *different* tweets for which an account was retweeted were positively correlated with being on the ground at those protests. Additionally, when the initial number of followers was high (early in the protest time frame), users were less likely to be on the ground—in other words, for accounts that already had a lot of followers, the high retweeted status was more likely correlated with high original follower count and less likely to be a result of the account being recommended due to being on the ground.

These findings suggest that the Twitter crowd can act—through the retweet mechanism and other social and interactive behaviors—as a filter for information coming from the ground during mass disruption events, a hypothesis we aim to further test here through a machine learning model.

Machine Learning to Build Location-Classification Models From Social Network Data

Machine learning algorithms are a form of artificial intelligence that can be used for a variety of applications, including natural language processing, search engine algorithms, spam detection software, and collaborative filters. Classifiers, an important subset of machine learning techniques, are programs that use training data to classify other data into specified categories. This training data "teaches" the classifier what is characteristic of each category of data. Once it has been trained, a classifier attempts to label data points provided in a set of validation data. These labels are then verified against the correct values and the classifier's accuracy is determined. Within machine learning there are several different approaches to classification, including Bayesian classifiers, neural networks, and Support Vector Machines (SVM), each of which excels at classifying certain types of data.

Machine learning algorithms have been employed by researchers to classify Twitter messages and profiles. For example, Verma et al. (2011) used such classifiers to identify tweets sent during crisis events that contain situational awareness information. Their work was based on natural language processing (NLP) strategies and focused on textual content of tweets. Hecht et al. (2011) tested the viability of using a machine learning classifier to determine the location of general Twitter profiles—not related to a specific event or other context—by examining the implicit location information contained within tweet text. Their algorithm attempted to assign profiles to the correct country or state and demonstrated decreasing accuracy with increased geographic focus. For mass disruption events, location identification may need to achieve a finer level of granularity, i.e. by city, neighborhood or even city-block. In the work reported here, instead of focusing on textual content within tweets or Twitter profiles to determine proximity to an event, we examine features of crowd recommendation and other social context (Kwak et al., 2010).

METHODS

We collected data during a short but high-activity period of the 2011 New York City Occupy Wall Street (OWS) protests, hand-coded each Twitterer according to location criteria informed by prior research, and applied a machine learning algorithm based on crowd recommendation and user behavior to identify on-the-ground Twitterers. We next explain these steps in detail.

Event Description: Occupy Wall Street, Early Protest (September 15 – September 21, 2011)

The "Occupy" movement is an ongoing[1] political protest and mass demonstration occurring in multiple cities around the globe. The original protests, loosely organized by Adbusters[2], were meant to mimic the Arab Spring demonstrations, uprisings that took place throughout the Arab World during 2011 (Pepitone, 2011). The original focus of the movement was the OWS protest in New York City, which was designed to have protesters camp out on Wall Street in the financial district to show their anger with economic disparities, unemployment, political corruption by financial interests, and the bank bailouts of 2008 and 2009. OWS was scheduled to begin on the afternoon September 17 and continue indefinitely. Highly publicized by its organizers, the protest's initial aim to occupy Wall Street itself was quickly thwarted by police, who closed off entrances to Wall Street. Protesters soon instead established themselves in Zuccotti Park, approximately two and half blocks away. In this

[1] Ongoing as of November 16, 2011.

[2] Link to Adbusters' site supporting protests: http://www.adbusters.org/campaigns/occupywallstreet

research, we focus on the New York geographical site of the OWS protest. Though several cities hosted Occupy protests, during the initial days of the event the New York protest was the only one that witnessed a large group of protestors, estimated in the hundreds, maintaining a constant presence at the protest site (Pepitone, 2011).

Twitter Data Collection

We began data collection at 11am EST on September 17, the first day of the protests, using the Twitter Streaming API to capture tweets forward-in-time and the Twitter Search API to collect tweets back-in-time.

Term(s)	Search API Window	Streaming API Window
`#occupywallstreet #dayofrage`	Sep 15 1pm – Sep 17 11am	Sep 17 11am – Sep 20 6:45pm
`#takewallstreet #sep17 #sept17`	Sep 15 1pm – Sep 17 1:45pm	Sep 17 1:45pm – Sep 20 6:45pm
`#ourwallstreet`	Sep 18 9:38am – Sept 18 10:05am	Sep 18 10:05am – Sep 20 6:45pm

Table 1. Collection Windows by Twitter API for each Hashtag Term

Initially, we captured tweets that contained either `#dayofrage` or `#occupywallstreet`, two terms that explicitly referenced the event. During this early period of the protest, several Twitterers claimed that tweets with protest hashtags were being censored and prevented from creating a "trending topic" on Twitter. In an attempt to counteract this perceived censoring, protesters, organizers and other Twitterer-supporters introduced new hashtags for people to use. This gave rise to a more elaborate strategy for data collection. For instance, `#ourwallstreet` did not receive widespread use until mid-morning on September 18. Additionally, over time we recognized that popular hashtags were omitted from our initial search. To deal with shifts in dominant hashtags, we added new search keywords after the initial collection began, conducting both a backward capture using the Twitter Search API and a forward-in-time filter with the Twitter Streaming API. The new terms then remained in the data collection protocol. This way, all terms were collected over the whole time period. Table 1 describes the resulting collection windows for each term or set of terms. This combined search resulted in 270,508 tweets from 53,296 Twitterers—what we call the *Keyword-Search-and-Filter* dataset.

Tweets collected by filtering from the Streaming API contain metadata that includes information from the author's Twitter profile, including number of followers, number of friends, number of lists the author is on, description, and location. To examine changes in profile information over time, we took a snapshot for each Twitterer in our *Keyword-Search-and-Filter* dataset of their profile information at the time we captured (with our filter search) their first and last protest related tweet. This analysis required us to limit our investigation to Twitterers for whom we collected at least two tweets (using the Streaming API), resulting in 23,847 users.

Next, we created a 10% sample of these users for further analysis. Because tweet-volume per user has a heavy-tailed distribution favoring low volume users, we used a tweet-based sampling strategy to flatten the distribution and sample more heavily among higher volume users. This set of 2,385 users is the *Twitterer-Sample* dataset.

To examine recommendation behavior related to the retweet mechanism, we tracked the propagation of all tweets that originated within an account in our *Twitterer-Sample*, counting the number of times each tweet was retweeted within the larger *Keyword-Search-and-Filter* dataset. For every Twitterer in the *Twitterer-Sample*, we also calculated the total number of times their account was retweeted, and the number of different tweets of theirs that were retweeted by others within the set.

Content Analysis – Identifying On-the-ground Tweeters

To create training data for the machine learning classifier, we needed to determine for every Twitterer whether they were on the ground at the New York City protests at any time during our collection window and whether or not they tweeted first-hand information about the event. Though Twitterers can designate a location for their accounts, research shows that this self-reported location does not always include valid geographic data (Hecht et al., 2011). During mass disruption events, self-reported location can also be inaccurate due to physical movement of the Twitterer (e.g. in cases of evacuation or convergence) or purposeful misinformation—e.g. many remote Twitter users changed their profile location to Iran during the Iran Election protests in 2009 (Reinikainen, 2009). Geolocation metadata may be valid and accurate, but only a small fraction of tweets have geolocation information. In our *Keyword-Search-and-Filter* dataset, only 124 of 53,296 Twitterers (0.23%) had geolocation metadata on any of their tweets. For these reasons, we could not rely on self-reported location information or tweet metadata to generate enough classification data for our study.

We used manual content analysis to make classifications for Twitterers, beginning with an investigation of all of their OWS keyword tweets—captured either by our Search API or Streaming API searches. If we could not make a determination from there, we then went to account owners' Twitter profile pages and read *all* of their tweets—those that contained protest keywords and those that did not.

Because the events were broadcast live through the Global Revolution livestream video feed[3], many Twitterers were tweeting real-time information from the ground without being physically present at the event. Conversely, there were a few Twitterers who we determined to be on the ground at the event but who were not tweeting information about the protest beyond assertions of being there, going there or having been there. Since our goal is to identify new information coming from on-the-ground sources, we classified Twitterers in the *Twitterer-Sample* into two groups: A) those who were on the ground and tweeting information from the ground, and B) those who were not on the ground or were not tweeting information about the protests from the ground.

Of the 2,385 Twitterers in our sample, 106 were found to be on the ground and tweeting information from the ground (A), 2270 were classified as not on the ground or tweeting information from the ground (B). Determinations could not be made for 9 Twitterers and these were excluded from the remainder of the study.

Location	**Total # of Twitterers**	**% of Total for SVM Classification**
Total	2385	100%
Ground & Tweeting Ground Info (***Group A***)	106	4.46%
Not Ground or Not Tweeting Ground Info (***Group B***)	2270	95.54%
Unknown – Excluded	9	NA

Table 2. Location Classification for *Twitterer-Sample*

USING A SUPPORT VECTOR MACHINE TO CLASSIFY TWITTERERS TWEETING FROM THE GROUND

For generating the machine learning model, we used a Support Vector Machine (SVM) using library LIBSVM (Chang & Lin, 2011) because this technique has been shown to work well on high-dimensional, noisy data (Schölkopf, 2004). The SVM can also classify data points with features that fall over a numerical range rather than being restricted to a binary assignment. To validate the accuracy and precision of the SVM, we used a stratified 10-fold cross-validation technique. Cross validation maximizes the use of data—each data point gets used for both validation and for training. Stratification of the folds reduces bias and variance as compared to unstratified k-fold cross-validation because each fold contains an equal concentration of data within each classification category (Kohavi, 1995). In our case, stratification maintains a constant ratio of on-the-ground Twitters to not-on-the-ground Twitterers across all folds. Once the SVM classifies each data point, we get an accuracy of the classifier based on the ratio of correct classifications to the number of instances in the dataset.

Feature Selection

We used the findings from Starbird & Palen (2012) to guide feature selection and the distillation of user profiles into quantifiable values. These features fell into two basic categories: *flat profile features* and *recommendation features*. Flat profile features are measures of social context represented by metadata in a user's profile—number of statuses, number of followers at the beginning of the event, number of tweets during the event, etc. Recommendation features included numbers associated with how the rest of the Twitter crowd interacted with the user—follower growth during the event, number of times the user was retweeted, etc.

Text in the user-specified location field was omitted for reasons explained earlier. We did determine whether text in the location field had *changed* over the course of the event, translating this information into a user behavior—or flat profile—feature. Table 3 lists each feature we used for our studies, along with statistical summaries for each feature by classification group.

[3] http://www.livestream.com/globalrevolution

Feature		Mean	Median	Standard Deviation	Min	Max
*Follower growth	*Group A*	181.3	20	658.6	-2	4167
	Group B	15.5	3	60.15	-90	1111
*Follower growth as % of initial #		1.20	0.0847	7.70	-0.025	75.0
		0.55	0.0126	7.50	-1.0	286.0
*Follower growth / friend growth		28.40	2.05	90.5	-2.0	701.2
		4.21	0.61	40.83	-32.0	1111
*Listed growth		6.24	1.0	17.11	-1	114
		0.54	0	2.01	-4	44
*Listed growth as % of initial #		0.33	0.032	0.61	-0.125	3.5
		0.09	0	0.44	-1	8
*Times retweeted (log)		4.24	4.16	1.99	0	9.5
		1.72	1.39	1.82	0	8.4
*Times RTed (log) / Initial followers (log)		0.89	0.78	0.77	0	6.2
		0.42	0.25	0.71	0	8.6
*Times retweeted (log) / # of tweets (log)		0.99	0.97	0.38	0	2.1
		0.46	0.42	0.48	0	4.7
*# different tweets retweeted / # of event tweets		0.32	0.27	0.21	0	1.0
		0.13	0.06	0.17	0	1.0
Statuses count (log)		7.12	7.22	2.40	0	11.5
		7.37	7.79	2.41	0	12.3
Initial followers count (log)		5.52	5.52	2.14	0	11.2
		5.04	5.25	2.11	0	11.5
Initial friends count (log)		5.45	5.51	1.56	0.69	8.9
		5.37	5.62	1.83	0	11.0
# of RTs as % of stream		0.44	0.46	0.23	0	0.9
		0.69	0.79	0.31	0	1.0
# of tweets for event (log)		4.17	4.21	1.20	1.39	6.8
		3.21	3.21	1.43	0.69	7.0
Description changed during event		0.25				
		0.21				
Location changed during event		0.057				
		0.068				

Table 3. Feature Summary by Classification Group.
Asterisks (*) denote Recommendation Features

Asymmetric Soft Margins

Because the ratio of on-the-ground Twitterers to not on-the-ground Twitterers is very low (see Table 2), our data set is considered *unbalanced*. Unbalanced data sets are challenging to classify with many machine learning techniques because the classifier maximizes overall accuracy by labeling most points as belonging to the category that is in the majority (Ben-Hur & Weston, 2010). For our problem, this means that the algorithm will tend to classify all or most of the data as not on-the-ground, which will result in high overall accuracy, but will identify very few on-the-ground Twitterers.

In a preliminary study we conducted in preparation for this research, using a simple SVM—one that did not take into account the unbalanced nature of the data set—the algorithm achieved an overall accuracy of 95.6%, but only classified 4.7% of on-the-ground Twitterer correctly. This led us to a second and better strategy for our goals and unbalanced data, which is to sacrifice overall accuracy and attempt instead to identify accounts that are likely to be on the ground. One way to do this is through the use of asymmetric soft-margins (Ben-Hur & Weston, 2010). SVMs can assign different misclassification costs—or *soft margins*—to each classification category. In our preliminary work, these soft margins were the same for both categories of data, a technique that works well for balanced data sets where maximum overall accuracy is ideal. When the data set is unbalanced, the SVM may lean towards sacrificing correct classification of the minority group in order to ensure higher correct classification overall.

To introduce different misclassification costs to each classification category, we enforce asymmetric soft-margins to effectively bias the SVM towards on-the-ground Twitterers. Asymmetric soft margins are implemented through constants calculated based upon the ratio of local to non-local Twitter accounts in the data set (Ben-Hur & Weston, 2010):

$$\frac{C_+}{C_-} = \frac{n_-}{n_+}$$

n_+	2270	C_+	1
n_-	106	C_-	21.415

Table 4. Asymmetric Soft Margins Constants

In this equation, the *n* values are the number of points in the data set belonging to each classification category and the *C* values are the asymmetric soft-margin constants.

Results

Table 5 shows the results achieved by this model. Even though overall accuracy (77.2%) was much lower than could have been achieved without asymmetric soft margins, *the SVM with asymmetric soft margins was far more successful at classifying on-the-ground Twitterers correctly*. The correct classification rate of locals we obtained was 67.9%, with a standard deviation of 19.3%. This is high in comparison to the percent of the data set that represents on-the-ground Twitterers (4.46%).

Fold	Total Accuracy	# On-ground Tweeters	# Correctly-Classified On-Ground Tweeters		# Not On-Ground Tweeters	# Correctly-Classified Not On-Ground	
1	74.0%	11	9	81.2%	227	167	73.6%
2	77.3%	11	10	90.1%	227	174	76.7%
3	78.2%	11	4	36.4%	227	182	80.2%
4	76.5%	11	8	72.7%	227	174	76.7%
5	74.4%	11	7	63.6%	227	170	74.9%
6	75.6%	11	9	81.2%	227	171	75.3%
7	76.4%	10	6	60%	227	175	77.1%
8	80.2%	10	4	40%	227	186	81.9%
9	82.7%	10	6	60%	227	190	83.7%
10	76.7%	10	9	90%	227	173	76.2%
Total	**77.2%**	**106**	**72**	**67.9%**	**2270**	**1762**	**77.6%**

Table 5. Results for Study: SVM with Asymmetric Soft Margins

The resulting dataset filtered by this model contains a far higher ratio of locals to non-locals. Among the 580 users the SVM identified as local, 72 were actually local, which triples the signal-to-noise ratio for on-the-ground Twitterers—from 0.047 in the original dataset to 0.142.

CONCLUSIONS AND FUTURE WORK

Building from existing research that claims that the connected crowd acts as a recommendation mechanism for information coming from the ground during mass disruption events (Starbird & Palen, 2012), this research demonstrates that leveraging crowd behavior in conjunction with Twitter profile information can work as a collaborative filter for identifying on-the-ground Twitterers. Using a Support Vector Machine with asymmetric soft margins, we were able to generate a model that correctly classified 68% of on-the-ground Twitterers, tripling to signal-to-noise ratio within our dataset.

In critically examining these techniques, it is important to appreciate that they do not generate a perfectly accurate classifier. This approach is not designed to be a standalone model for user classification. By isolating features of crowd recommendation and user behavior, this research demonstrates the utility of including these types of features in classification strategies. Ideally, classification algorithms should include features like the ones we identified here in combination with features related to the textual content of tweets and Twitter profiles. Additionally, in the context of mass disruption, where veracity of information is vital, machine-only computational solutions are not ideal, and the information resulting from this or any filtering technique must be further combined with human judgment to assess its accuracy.

This "limitation" fits well within an information space that is witnessing the rise of digital volunteer communities (Standby Task Force[4]; Humanity Road[5]; Starbird & Palen, 2011; Starbird, 2012) who monitor multiple data sources, including social media, looking to identify and amplify new information coming from the ground. Earlier in this paper, we described how a remote Twitterer worked to create and publicize a list of on-the-ground Twitterers during the first few days of the OWS protest. In related work, Starbird (2012) reports how, during response efforts, a virtual volunteer organization assigns multiple volunteers to the task of "media monitoring," an activity that includes identifying and creating lists of on-the-ground Twitterers. For volunteers like these, the use of techniques that increase the signal to noise ratio in the data has the potential to drastically reduce the amount of work they must do. The model that we have outlined does not result in perfect classification, but it does increase this signal-to-noise ratio substantially—tripling it in fact.

These findings also suggest some frightening implications—especially in the case of political protest—that people using social media during mass disruption events cannot hide. In some cases and for some people, this might not be a problem, but the possibility of discovery makes it harder for people to operate publicly during events where evasiveness bought through delay of identification might be important. The crowd does not shelter these people, but instead gives them away.

Importantly, this research demonstrates that we can use empirical studies like Starbird & Palen (2012) to inform the novel application of machine learning approaches and other computational strategies for extracting useful information from social media interactions. By isolating features of crowd recommendation and user behavior from previous empirical findings, this research demonstrates the utility of including these types of crowd behavioral features in classification strategies.

ACKNOWLEDGMENTS

We thank reviewers of this paper for helpful feedback. We thank members of Project EPIC at the University of Colorado Boulder for their ongoing support. We are grateful to the US National Science Foundation, which funded this research through a Graduate Research Fellowship; grant IIS-0546315 and IIS-0910586

REFERENCES

1. Anderson, K., & Schram, A. (2011). Design and Implementation of a Data Analytics Infrastructure In Support of Crisis Informatics Research. *ICSE 2011*, 21-28 May 2011, Honolulu, Hawaii.
2. Ben-Hur, A. & Weston, J. (2010). A User's Guide to Support Vector Machines, *Methods in Molecular Biology*, 609, 223-239.
3. boyd, d., Golder, S. & Lotan, G. (2010) Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: *HICSS-43 2010* (forthcoming).
4. Burns, A. & Eltham, B. (2009). Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 Election Crisis. In *Communications Policy & Research Forum*.
5. Chang, C. & Lin, C. (2011). LIBSVM : A Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, 2(27), 1-27.
6. Dynes, RR. (1970). Organized Behavior in Disaster. Heath: Lexington, MA.
7. Fritz, C. E., & Mathewson, J. H. (1957). *Convergence behavior in disasters: A problem in social control*. Washington, DC: National Academy of Sciences.
8. Gillmor, D. (2004). We the Media: The Rise of Citizen Journalists. *National Civic Review*, F 2004: 58-63.
9. Grossman, L. (2009). Iran's protests: Why Twitter is the medium of the movement. *Time*, 2009. Retrieved May 22, 2011. http:// www.time.com/time/world/article/ 0,8599,1905125,00.html
10. Hagar, C. & Haythornthwaite C. (2005). Crisis, Farming & Community. *Jour. of Community Informatics, 1*(3), 41-52.
11. Hecht, B., Hong, L., Suh, B. & Chi, E. (2011). Tweets from Justin Bieber's Heart: The Dynamics of the "Location" Field in User Profiles. *Proc of CHI 2011*, 237-246.
12. Heverin, T., & Zach, L. (2010). Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in Seattle-Tacoma, Washington Area. Presented at *ISCRAM 2010*.
13. Hughes, A. & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. *International Journal of Emergency Management*, 6 (3/4), pp 248-260.
14. Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, *Proceedings International Joint Conference on Artificial Intelligence*, *2*(12), 1137-1143.

[4] http://blog.standbytaskforce.com/

[5] http://www.humanityroad.org/

15. Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is Twitter, a social network or a news media? *Intl. WWW Conference*, (Raleigh, NC, 2010), ACM, New York, NY, USA, 591-600.
16. Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I. & boyd, d. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *Intl Journal of Communications* 5, 1375-1405.
17. Malone, T. W., Laubacher, R. & Dellarocas, C. N., (2009). Harnessing Crowds: Mapping the Genome of Collective Intelligence (February 3, 2009). MIT Sloan Research Paper No. 4732-09.
18. Mark, G., Bagdouri, M., Palen, L., Martin, J., Al-Ani, B., & Anderson, K. (2012, forthcoming) Blogs as a Collective War Diary. To appear in *Proc. of CSCW 2012*, (Seattle, WA).
19. Mendoza, M. Poblete, B. & Castillo, C. (2010). Twitter under crisis: can we trust what we RT? In *Proceedings of the First Workshop on Social Media Analytics* (SOMA '10). ACM, New York, NY, USA, 71-79.
20. Messina, C. (2007) Groups for Twitter; or A Proposal for Twitter Tag Channels. Oct 22, 2007. Blog in: FactoryCity. URL: http://factoryjoe.com/blog/2007/10/22/twitter-hashtags-for-emergency-coordination-and-disaster-relief/
21. Noble, W.S. (2006). What is a Support Vector Machine? *Nature Biotechnology*, *24*(12), 1565-1567.
22. Palen, L., & Liu, S. B. (2007). Citizen communications in crisis: Anticipating a Future of ICT-Supported Participation. *Proc of the CHI 2007*. ACM, NY, USA, 727-736.
23. Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A vision for technology-mediated support for public participation and assistance in mass emergencies and disasters. In *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*. *ACM-BCS Visions of Computer Science*. British Computer Society, Swinton, UK, 1-12.
24. Pepitone, J. (2011). Hundreds of Protestors Descend to 'Occupy Wall Street'. *CNNMoney* (September 17, 2011). Retrieved November 17, 2011.
25. Qu, Y., Huang, C., Zhang, P. & Zhang, J. (2011). Microblogging after a major disaster in China: A case study of the 2010 Yushu Earthquake, *Proc of CSCW 2011*, (Hangzhou, China). ACM, NY, USA, 25-34.
26. Qu, Y., Wu, P.F., Wang, X. (2009). Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake, *42nd Hawaii International Conference on System Sciences* (*HICSS '09)*, 1-11.
27. Quinn, A. & Bederson, B. (2011). Human-Machine Hybrid Computation. Position Paper for the CHI 2011 Workshop on Crowdsourcing and Human Computation. *CHI 2011*.
28. Reinikainen, E. (2009). #iranelection cyberwar guide for beginners. Blog in *Networked Culture*. Retrieved Sept 21, 2011. Available at: http://reinikainen.co.uk/2009/06/iranelection-cyberwar-guide-for-beginners
29. Sarcevic, A, Palen, L, White, J, Starbird, K, Bagdouri, M & Anderson, K. (2012). "Beacons of Hope" in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake. *Proc of CSCW 2012*.
30. Schölkopf, B., Tsuda, K. and Vert, J.P. (2004) - *Kernel Methods in Computational Biology*. MIT Press series on Computational Molecular Biology. MIT Press.
31. Shklovski, I., Burke, M., Kraut, R. & Kiesler, S. (2010) Technology Adoption and Use in the Aftermath of Hurricane Katrina in New Orleans. *American Behavioral Scientist*.
32. Shklovski, I., Palen, L., & Sutton, J. (2008) Finding Community Through Information and Communication Technology in Disaster Events. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work (CSCW 2008)*, November, San Diego, pp. 127-136.
33. Starbird, K. (2012). What "Crowdsourcing" Obscures: Exposing the Dynamics of Connected Crowd Work During Disaster. *Collective Intelligence 2012*, Cambridge, MA. Forthcoming.
34. Starbird, K., Palen, L., Hughes, A. & Vieweg, S. (2010). Chatter on The Red: What hazards threat reveals about the social life of microblogged information. *Proc of CSCW 2010*. ACM, NY, USA, 241-250.
35. Starbird, K., & Palen, L. (2010). Pass It On?: Retweeting in Mass Emergencies. Presented at *ISCRAM 2010*.
36. Starbird, K., & Palen, L. (2011). 'Voluntweeters': Self-organizing by digital volunteers in times of crisis, *Proc of CHI 2011*. ACM, New York, NY, USA, 1071-1080.
37. Starbird, K. & Palen, L. (2012). (How) Will the Revolution be Retweeted?: Information Propagation in the 2011 Egyptian Uprising. *Proc of CSCW 2012*.
38. Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A., & Anderson, K. NLP to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. In the *Fifth Intl. AAAI Conf. on Weblogs and Social Media*, 17-21 July 2011, Barcelona, Spain.
39. Vieweg, S., Hughes, A., Starbird, K., & Palen, L. (2010). Micro-blogging during two natural hazards events: What Twitter may contribute to situational awareness, *Proc of CHI*, pp. 1079-88.