Evaluating Disaster Time-line from Social Media with Wavelet Analysis

Amrita Anam
University of Maryland
Baltimore County
amrita1@umbc.edu

Aryya Gangopadhyay
University of Maryland
Baltimore County
gangopad@umbc.edu

Nirmalya Roy
University of Maryland
Baltimore County
nroy@umbc.edu

Abstract—For over a decade, social media has proved to be a functional and convenient data source in the Internet of things. Social platforms such as Facebook, Twitter, Instagram, and Reddit have their own styles and purposes. Twitter, among them, has become the most popular platform in the research community due to its nature of attracting people to write brief posts about current and unexpected events (e.g., natural disasters). The immense popularity of such sites has opened a new horizon in 'social sensing' to manage disaster response. Sensing through social media platforms can be used to track and analyze natural disasters and evaluate the overall response (e.g., resource allocation, relief, cost and damage estimation). In this paper, we propose a two-step methodology: *i)* wavelet analysis and *ii)* predictive modeling to track the progression of a disaster aftermath and predict the time-line. We demonstrate that wavelet features can preserve text semantics and predict the total duration for localized small scale disasters. The experimental results and observations on two real data traces (flash flood in Cummins Falls state park and Arizona swimming hole) showcase that the wavelet features can predict disaster time-line with an error lower than 20% with less than 50% of the data when compared to ground truth.

Index Terms—Flash Flood, Disaster Response, Wavelet Analysis, Social Media

I. Introduction

In recent years, the frequent occurrence of disasters has caused devastation that is beyond imagination. Last year alone, an alarming number of hurricanes, earthquakes and floods [1], [2], [3], [4] stunned the world. Irrespective of the cause of a disaster, it is vital for the authorities to be prepared for an emergency situation and act to the best of their ability. The outdated, traditional approaches of response operations are falling short. Therefore, a need has emerged to build data driven, automated, robust and resilient communication platforms to help the authority make efficient decisions. This once overreaching ambition is now achievable by establishing human-in-the-loop cyber-physical systems with the help of social media because of its ubiquitous connectivity with the rest of the world. The power of social media lays across many different domains that provide social good. Social media can help accelerate the detection and monitoring of an emergency event in real time. During hurricane Harvey, people used social media to draw attention to rescue requests, finding safe shelters and relief inquiries [5]. The appropriate sensing through social-media can help in estimating the damages and recovery costs,

reconstructing national infrastructures, creating social awareness and providing just-in-time aid [7], [8], [9]. As mentioned in the federal research document [6], social media can be used systematically to conduct emergency communications, spread awareness, and receive information from the victims. In fact, FEMA has been using social media to send and receive information on disasters since 2011 by curating and processing them manually [6]. This leads to the research question of how to automatically process real-time information from different social sensing platforms and use them to manage disaster response.

Wavelet analysis is applied to find patterns or anomalies with respect to both time and frequency of a signal. Wavelets are oscillations that start and end with an amplitude of zero and oscillates in between [10]. Conversations following a disaster on social media follow the same pattern which manipulates the occurrence of words specific to an incident (e.g., 'water' during a flood and 'wind' during a storm). Throughout the event, the words oscillate and follow a sharp decrease when the event ends. In this paper, we demonstrate: *i)* a scalable framework that creates a temporal representation of context from Twitter similar to physical sensor data, and *ii)* an algorithm to estimate the rate of change of uncertainty and time-line of disaster response using wavelet entropy. In the process, we make the assumption that every physical event has an identifiable context that is reflected in social media by a list of unique words which we will be referring to as 'context words'. Each context word can be represented as time-dependent signals, corresponding to a frequency spectrum, and amplitudes that vary with time. This representation of sensing and processing the context words as signals is robust and scalable. Our observations reveal that temporal attributes depict the progression of storyline whereas, the frequency spectrum indicates the concentration and diffusion of a topic during an event.

Our key contributions are summarized in the following points:

1) We present a scalable method to create a feature from wavelet entropy to assess change in ongoing events
2) We propose an iterative nonlinear interpolation model utilizing the new feature to predict the trajectory and time-line of the response at a very early stage of a disaster

3) We illustrate that wavelets can preserve text semantics without having to maneuver supervised natural language models

4) We substantiate the assumptions by investigating two recent local flash floods in the USA – Cummins Falls state park and a swimming hole in Arizona

II. Data

We identified two flash flood events in two different locations with very similar response time. The tweets were collected using query terms specific to the two events (e.g., "cummins falls + flash flood", "cummins falls + flood", "arizona + swimming hole", "arizona + flood", etc). During the flash flood in Cummins Falls State Park, 40 people were trapped behind the falls, a 73 year-old woman and a 45-year-old woman were reported missing. Their bodies were found after three days of frantic rescue operations. The flash flood in Arizona hit a swimming hole at the Tonto national forest. A group of 15 people was in the swimming hole when the flood hit the popular recreation area killing nine people from the same family. The rescue operation was carried out by 40-50 professionals. Snapshots of two Twitter posts of the flash floods are shown in Fig. 1.

We collected Twitter data using the REST API. In Twitter, the re-tweets provide a shared perspective and spread of information, whereas, individual tweets provide unique perspectives and a diverse set of information. If a group of
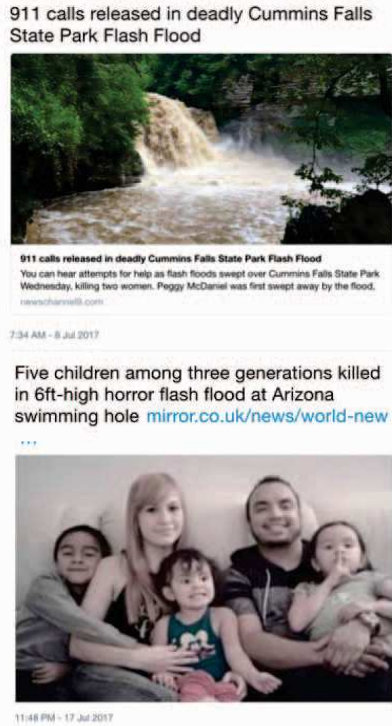

Figure 1: Snapshots of tweets about flash floods in Cummins Falls state park (top) and Arizona swimming hole (bottom).
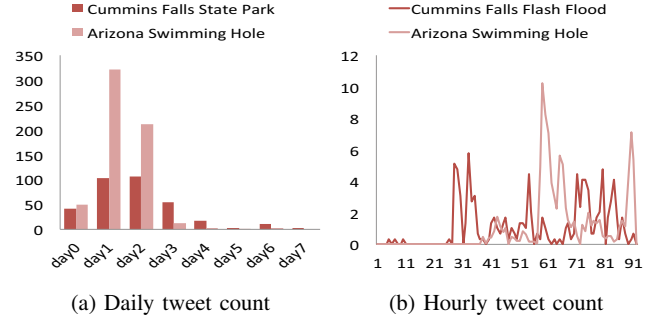

Figure 2: Temporal pattern of social media posts

people individually tweet about the same physical event, then they are legitimizing that the event has really happened. To get more individual perspectives and avoid redundancy, we removed weather alerts and re-tweets of news broadcasts. The number of unique tweets for Cummins Falls Flash Flood is 338 and Arizona Swimming Hole is 599. Since these are small-scale local events that mostly spurred a reaction near the origin of the event, there are not many unique tweets. The tweets were collected before, during and after the event. The tweet count from both the flash floods for every day and every hour is shown in Fig. 2a and Fig. 2b respectively. Both the distributions in Fig. 2a are unimodal because there is only one day with the most significant incident, the flash flood itself. The distributions may have different characteristics for different scale and type of events. Note that the daily count spikes on the *day 1* right after the flash flood and slowly diminishes. The hourly count in Fig. 2b varies with the time-of-day making the graphs multimodal. Particularly, most of the tweets occur between evening and late night. It depicts that people tweet at their convenience and the news takes a few hours to spread [12].

III. Background and Related Work

Researchers have been studying the effectiveness of social media in acquiring and analyzing real-time information for almost a decade. In [12] and [13], the authors explained information diffusion, influence score, and propagation of news by analyzing people's tendency to tweet. In [14], [15], and [16], the authors demonstrated that social media can be used in all phases of disaster management, such as monitoring causes and effects, broadcasting alerts, helping in search and rescue, and relief management. The wordwide popularity of social media among the emergency management organizations to disseminate their message to the communities and first responders was discussed in [6], [9], [8]. Event detection has been a very popular field in social media analytics [19], [20], [21]. Multiple tools and systems have been designed to track and monitor real-life incidents on Twitter. TwitterMonitor is a web-based system that detects emerging topics and trends by extracting and grouping bursty keywords from Twitter streams. [18]. TweetTracker was built in collaboration with Humanitarian Aid and Disaster Relief (HADR) to track targeted

disasters on Twitter that can guide first responders in gaining awareness regarding an emergency situation [17]. The authors of [22] fused the events detected from Twitter and Instagram which performs better than their individual detetction. In [23], invented a system, StoryLine that can identify and track events and sub-events from Twitter by tracing the change in entropy and information gain of keyword pairs. These tools and platforms utilize natural language features (e.g., keywords, word associations, probabilities, position in a sentence, etc.) to assess social media posts which require prior knowledge.

Some previous research works show that wavelet analysis on social media can be used to analyze content dynamically without any prior knowledge. We took inspiration from the authors of [24] who used discrete wavelet transform and wavelet entropy for event detection. The authors of [26] created time-varied signals of clicks, hash-tags, and phrases and converted them to wavelets and developed a clustering algorithm to find the temporal pattern of content. In [27], the authors explored the diversity of content where diversity is defined by the change in entropy in a different spectrum of the wavelets created from different contents. In [28], a wavelet based pattern matching algorithm was proposed for user recommendation. In [25], the authors used wavelet transform in detecting events of different scales by analyzing temporal and spatial features together. Some of the challenges of these approaches include merging the text processing with wavelet transform and the complexity of wavelet analysis on massive data sets. To tackle this challenge, we opted for a discrete representation of the context, binned with a fixed duration to create signals similar to sensors. In addition, instead of applying CWT on the full signal, we developed an overlapping, sliding window approach that processes a fixed number of bins at a time. The advantages of our approach are: *i)* it makes the model scalable, *ii)* it captures a continuous change in the data, and *iii)* it does not require any prior knowledge or training.

IV. Text to Signal Processing

During a natural disaster, social media, much like physical sensors, constantly provide real-time information. Taking advantage of this behavior, we converted the unstructured text into a list of words, each represented by numeric vectors of fixed length. Every value in the numeric vector is associated with a temporal stage. The new representation has unique words as entries and time bins as features. The tweets from the two events are processed separately. The architecture that we followed to create the word signals from the social media posts is presented in Fig. 3.

A. Text preprocessing

In the preprocessing stage, we tokenize the tweets into unigrams. We filter out stop words and query words (i.e., 'Cummins Falls', 'Arizona', 'flash flood', etc.). In this paper, we refer to tweets from an event by T_{event} (Cummins Falls as T_{cf} and Arizona as T_{az}) and the set of context words by U_{event} (U_{cf} and U_{az}). A T_{event} is associated with only one
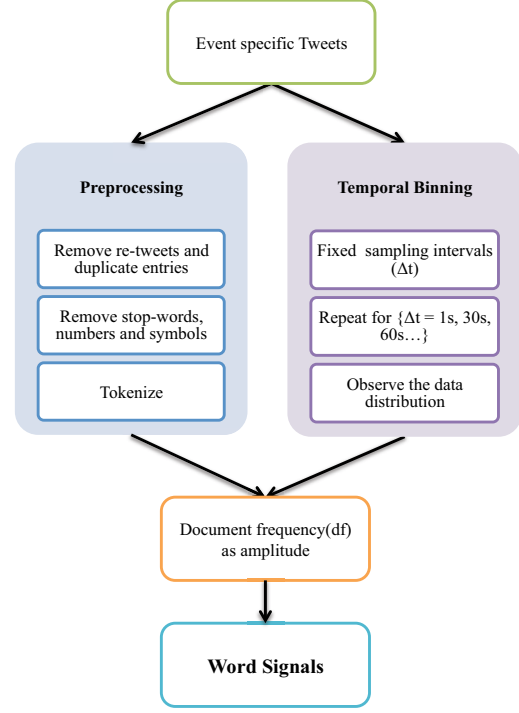

Figure 3: The conversion of text to time dependent signals

U_{event}. We denote the total number of unique words in an event by m, $U_{event} = u_1, u_2, ..., u_m$.

B. Temporal Binning

The temporal granularity of Twitter is one second which means that data can be collected for every second. However, for small scale local events, the data flow for every second created from a T_{event} is sparse with little variation. Therefore, we discretize the occurance of the context words into n bins with a fixed duration Δt where $\Delta t >= 1$. $\Delta t = 1$ is the continuous and $\Delta t > 1$ is the discrete representation of the context words. We refer to this representation of context words as 'word signal'. We used document frequency (df) as the amplitude of the word signals [24]. If the tweets for a disaster have m unique words, then the df of the ith word for the jth bin is, $df(u_{ij})$ = Number of tweets in bin j that contain u_i. Let, t_0 denote the start time of a word signal, then the next bin is $t_1 = t_0 + \Delta t$ and amplitude is the df of the word between
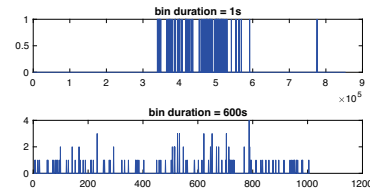

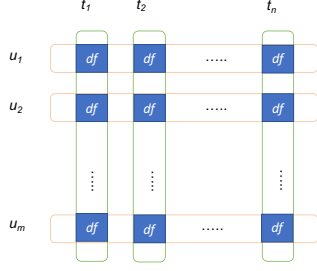Figure 4: Signal representation of the word "woman"

Figure 5: Temporal representation of context

Figure 7: Wavelet coefficients of the word signal "woman"

t_0 to t_1. The continuous ($\Delta t = 1s$) and discrete ($\Delta t = 600s$) representations of a word signal from the Cumins Falls Flash Flood is shown in Fig. 4. The continuous representation has less variability than the discrete representation. the data structure of the word signals is given in Fig. 5. By calculating df for n bins we create a signal for the word u. After, repeating this step for m words, we get an mxn matrix, S where m = number of signals and n = number of bins and $s_{ij} = df(u_{ij})$ as shown below.

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \dots & s_{1n} \\ s_{21} & s_{22} & s_{23} & \dots & s_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & s_{m3} & \dots & s_{mn} \end{bmatrix}$$

We observed the data distribution and sparsity of the signals by varying Δt from one second to one hour. For both the events, the data appears to be very sparse. We observed the log-density distribution of all the word signals starting from one second to one hour. Fig. 6 shows the sparsity of the data for different values of Δt. The overall variance is very low for the lower values of Δt. By manually comparing the incidents in the the two data sets, we found that Δt = 600 seconds captures the variation in the data flow without losing information. The document frequencies of words by time articulates this story and its progression quite accurately. We did not normalize df with the number of documents per bin for the following reasons: *i)* in a single platform, signal parameters are evaluated on the same scale *ii)*, reduces overall
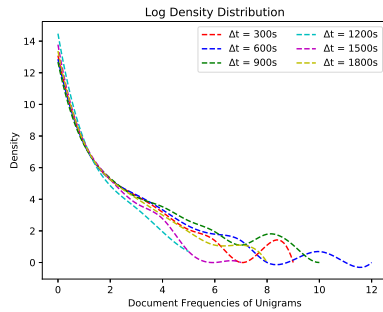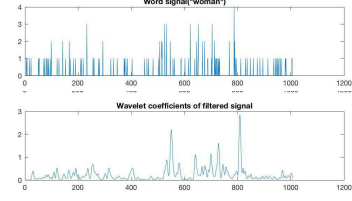
variance in the data and *iii)* creates an unbalanced comparison of importance between terms in different bins.

V. WAVELET ANALYSIS

After the context words and converted to word signals, we perform wavelet analysis to create new features.

A. Continuous Wavelet Transform

The wavelet analysis consists of two parts: *i)* applying Continuous Wavelet Transform (CWT) on the word signals and *ii)* calculating the normalized Shannon Wavelet Entropy (SWE) i.e., H-measure of the wavelet coefficinets. We used Morse wavelet as the mother wavelet [29] with the symmetry parameter $\gamma = 3$. The Morse mother wavelet converts one dimensional real or complex vector into a time-frequency matrix of complex coefficients. We split the discrete word signals into chronological windows where each window contain six adjacent bins and apply CWT on each window. To maintain the continuity of the wavelets, we overlap the last bin of a window with the next window. Since CWT is the most computationally expensive part of our framework, the windowing approach addresses the problem by fixing the input size of one wavelet transform. The complexity of the wavelet transform increases linearly with the number of word signals m and the number of windows. In this paper, we represent window length by Δw and number of bins in a window by p, $p = \Delta w/\Delta t$. The dimensions of the CWT for each iteration is, $\mathbb{R}^p$ or $\mathbb{C}^p \Rightarrow \mathbb{C}^{p \times q}$ where p is the number of bins in a window and q is the total number of frequencies extracted from the input signal during the wavelet transform. Fig. 7 illustrates

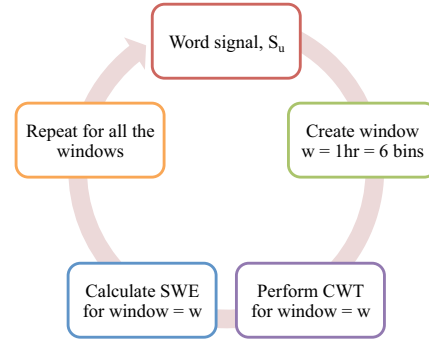Figure 6: Sparsity of the Signals
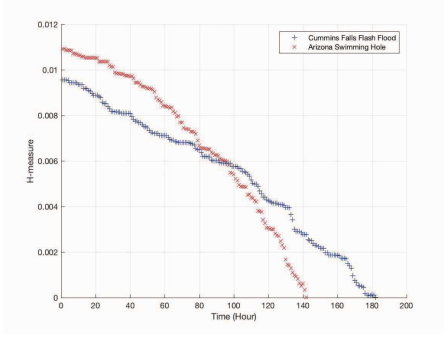

Figure 8: Wavelet Analysis Process

Figure 9: The total H-measures of the two events

the wavelet coefficients of the word signal with $\Delta t = 600$ in Fig. 4.

B. Trajectory of Uncertainty

After applying CWT, we calculate H-measure of the wavelet coefficients for each window. The process is repeated for all the words. The method of calculating H-measure of a word signal is shown in Fig. 8 and the equations are given in Appendix. B. We define the trajectory of an event by linearly summing over H-measure of all the word signals, $H_{total} = H_1 + H_2 + ... + H_m$ where $H_1, H_2, ..., H_m$ are the H-measures of word signals $s_1, s_2, ..., s_m$ respectively. The trajectory of uncertainty of the two events are presented in Fig. 9. By observing the trail of H-measures of the events, we can assert that the uncertainty is at the peak when the event starts and diminishes depending on the following aftermath (i.e., rescue, damage and casualty). The trajectory depicts how a real event is projected in the social media. The diminishing pattern of the trajectory mirrors a quadratic nature where the height of the total H-measure depends on the initial reaction of the physical event in the social media and the descent depends on the response of the event and its aftermath.

VI. PREDICTION OF EVENT CHARACTERISTICS

As one of the many possible applications, we have designed an iterative nonlinear interpolation model using the total H-measure as the input feature. The model predicts the trajectory and total duration of the event from the moment it is detected. We make predictions every hour by processing all the context words during that hour. The algorithm is designed to suit batch processing technique on streaming data. The total number of data points (H-measure) is N where N represents the number of windows. For the kth iteration, the model predicts the future $N - k$ data points from the past k input data points by fitting a quadratic diminishing curve. Since the number of data points increases with every iteration, so does the prediction accuracy of the model. Algorithm 1. shows the function of the polynomial interpolation and Algorithm 2 shows the steps of the iterative predictive model. We define the start of the event when the H-measure is at the peak and start the first iteration from the next hour.

Algorithm 1 Fit H-measure to Learn Event Parameters

INPUT: x, y co-ordinates with independent variable x and depedndent variable $y = f(x)$
OUTPUT: $\bar{y}$, fitted values, $\beta = \beta_0, \beta_1, \beta_2$

    function FITTRAJECTORY:(x, y)
        while all β values have not converged **do**
            Fit $\bar{y} = \beta_0 + \beta_1(x - \beta_2)^2$
        end while
    end function

Return: $\bar{y}, \beta$

Algorithm 2 Predict Event Trajectory and Duration

1: $N \leftarrow$ Total number of data points
2: $t \leftarrow$ Time sequence of the event
3: $h \leftarrow H$-measures of the event, $h = f(t)$
4: $k \leftarrow$ Number of data points to fit the regression
5: $D \leftarrow$ Total duration of the event
6: **for** $i = 2 : t[end]$ **do**
7:    **Call:** FITTRAJECTORY(x, y) to fit $(t, h)[1 : i]$
8:    **Learn:** event coefficients $\beta = \beta_0, \beta_1, \beta_2$
9:    **Predict:** $\bar{h}[i : end]$ with learned $\beta = \beta_0, \beta_1, \beta_2$
10:   **Derive:** event duration, $\bar{D} = \sqrt{\frac{-\beta_0}{\beta_1}} + \beta_2$
11:   **Calculate:** error, $\frac{1}{N} \sum_{i=1}^{N} (\bar{y} - y)^2$
12: **end for**

To evaluate the predictive model, we calculated the mean squared error (MSE), $MSE = \frac{1}{N} \sum_{k=1}^{N} (\bar{y} - y)^2$ between predicted H-measures with the original H-measures for every iteration. The performance of the predictive model for the two data sets is depicted in Fig. 10 and Fig. 11. The sub-graphs Fig. 10a and Fig. 11a display the H-measures of all the context words. The randomly scattered points at the beginning of the graphs correspond to the entropy of the words that are extracted from the tweets collected before the event started. When the event starts, the H-measures rise abruptly and then slowly decay with time. The sub-graphs Fig. 10b and Fig. 11b demonstrate the iterative prediction of the H-measures for different values of k. With each iteration, more information is known which makes the learning algorithm perform better and the predicted trajectory moves closer to the original (entropy shown in solid line). Fig. 10c and Fig. 11c show the MSE values of the prediction with every iteration, projected on top of the original H-measures. The MSE values are quite high initially, but they gradually decrease as the event progresses. The decrease in MSE with increasing percentage of k is given in Table. I. It shows that for Arizona Flash flood, the MSE drops to 15% from 45% only with 10% increase in the data, however for Cummins Falls flash flood, the model takes 50% of data for the MSE to drop to 18% which indicates the percentage of data needed for 85% accuracy vary for different events. This gives the confidence that even with less than 20% of the data points we can achieve up to 85% accuracy which can help inform the responsible authorities to make a
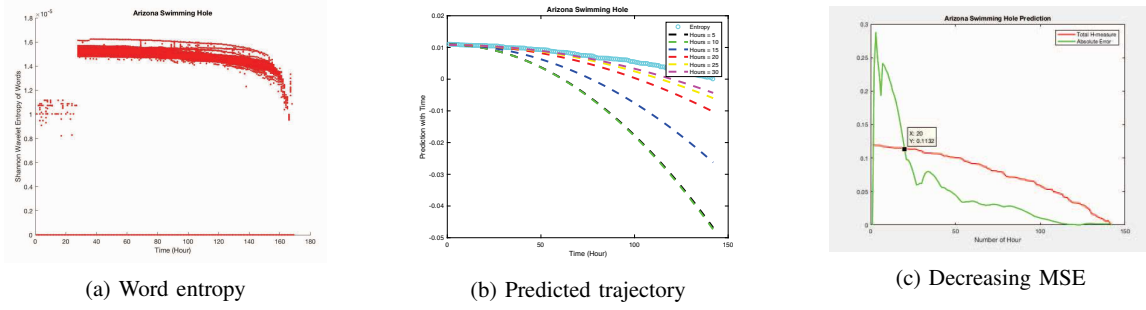
(a) Word entropy (b) Predicted trajectory (c) Decreasing MSE

Figure 10: Arizone Swimming Hole - Prediction results

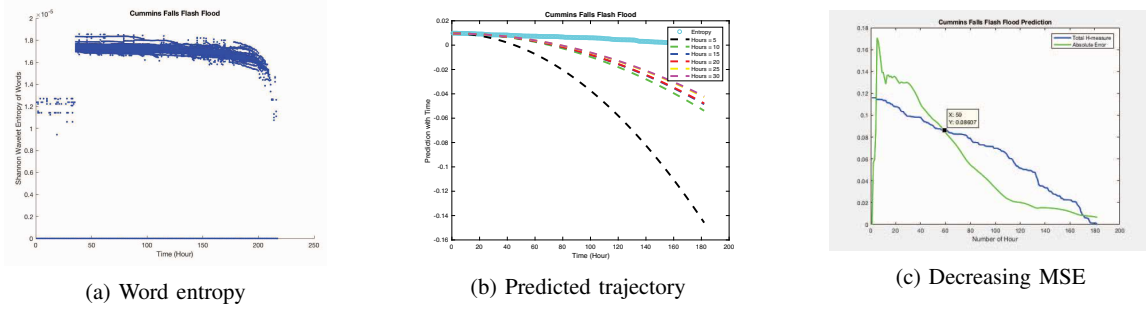(a) Word entropy (b) Predicted trajectory (c) Decreasing MSE

Figure 11: Cummins Falls State Park - Prediction results

Table I: Percentage error of event duration prediction

Data%	MSE % (Ucf)	MSE% (Uaz)
10%	59.37%	45.88%
20%	50.44%	15.60%
30%	39.40%	14.05%
40%	27.89%	8.70%
50%	18.76%	7.64%
60%	10.04%	6.70%
70%	7.45%	2.83%
80%	6.51%	0.55%
90%	4.48%	0.53%
100%	2.92%	0.44%

better decision with their disaster management effort. Using this method we can make predictions during the events using wavelets without any prior knowledge and adjust the estimates with time.

VII. SEMANTIC ANALYSIS WITH WAVELETS

In this section, we illustrate, that it is actually possible to explain semantics with wavelets by comparing the time and the frequency values of the high magnitude coefficients to the actual event. The word signals vary in both time and frequency. When people post on social media reacting to a physical event, they use words that reflect the incidents, the news, and their emotions. Instead of processing social media posts based on only time or frequency of their occurrence, it is more meaningful to process both together to capture the dynamic movement of the context. Fig. 12 shows the scalograms of the wavelets of some of the most frequent words from the tweets of Cummins Falls flash flood. The scalograms display the magnitude of the complex coefficients of the wavelets with time in the x-axis and frequency in the y-axis. The white dashed line represents the cone of influence that marks where edge effects become significant. The color scale depicts the energy distribution of the coefficients throughout the event. The blue region is caused by the low magnitude values of the coefficients and the illuminated region is caused by the high magnitude values of the coefficients. The blue region seems to be dominating due to the sparsity of the input signals. The word "people" is spread out on both time and frequency axes because the word is used in multiple context throughout the event (e.g., "40 people were stranded on the park'", "people were searching", etc.). On the other hand, the word "woman" only appears in the context of the elderly woman who got washed away. The wavelet coefficients corresponding to the word "woman" has high magnitude value at the beginning of the scalogram which is consistent with the actual event because the woman got washed away during the flood and her body was found after two days of search. The word "search" and "missing" are also spread out but "missing" is brighter because it appears more in the tweets than "search". The word "rescue" dominates a few time intervals because there are tweets about the rescue process throughout the event. However, "rescued" has its high magnitude coefficients towards the end because
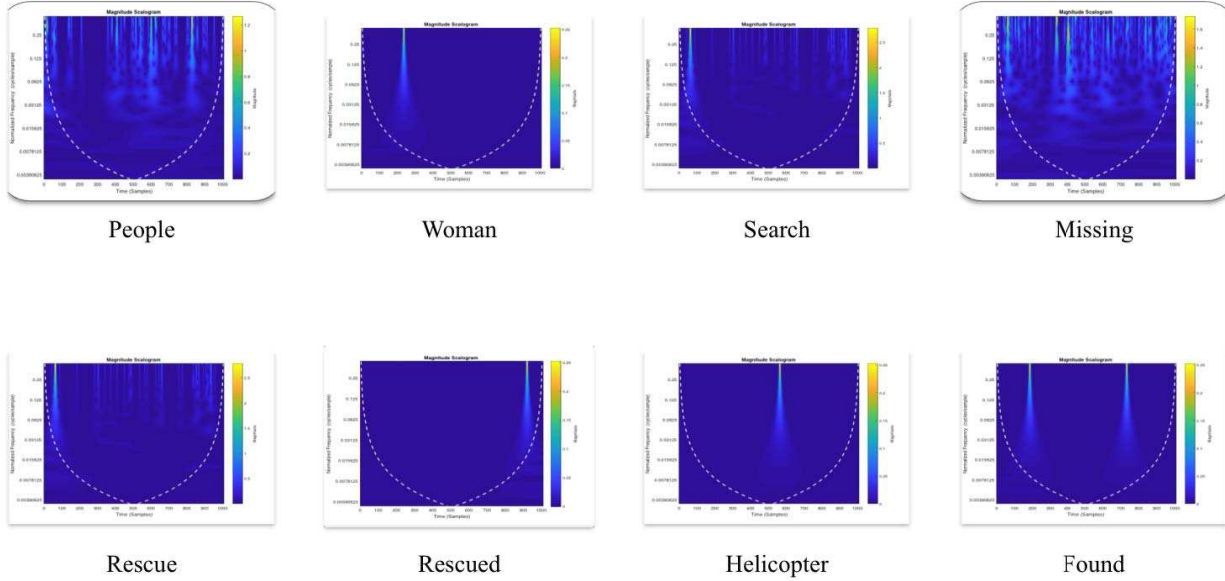
Figure 12: Scalograms of word-wavelets

the process had ended by then. The word "helicopter" has high magnitude coefficients at the middle of the scalogram when the rescue teams were searching for the missing woman's body. The word "found" has two strong appearances. We manually looked into it and saw that in the beginning, the conversations were about "woman not being found" which changed later to her body being "found".

VIII. CONCLUSION

Social sensing opens a whole new world of disaster management where the information goes well beyond the structured realm of physical sensors. To be able to access this data freely is a reward on its own. It should be fully explored into building applications and services that benefit the source community. The success of the domain lies in exploring fresh, creative and efficient methods to analyze the data. In every phase of a natural disaster, the responsible authority continuously makes estimates and predictions before and during the disasters. However, when an unexpected natural disaster takes place, it is impossible for the authorities to make predictions and preparations before the event. Even with the most accurate predictions and preparations, the aftermath can be devastating which is why it is vital to build models that can adjust as the event progresses. The experiment presented in this paper shows that, with wavelet analysis of social media data, predictions can be made about the aftermath during the event.

The prediction of the event trajectory and duration will help the authorities make preparations and allocate resources based on the projected duration. The framework can also be used to track aftermath such as, damage status, victim sentiments during a disaster, which can improve the management of

the future disasters. This method is not confined to any language or platform. This can effectively be extended to other social media and news articles to analyze the major topics of conversation during natural disasters. The algorithm can be built into an end-to-end tool that tracks social media data and provides feedback to the authority and first responders. Since the method is focused on traditional signal processing, it can be fused with other cyber-physical systems by combining physical sensor data with social media data. Apart from the evolution of text, the methodology shows an interesting observation that wavelets can be an efficient alternative to traditional methods in explaining and analyzing text semantics. Wavelet analysis can be used to process conventional natural language parameters (e.g., context, word co-occurrences, synonyms, antonyms, etc.) that will go beyond the scope of social media and disaster management. We are currently working on advancing this project in multiple directions including identifying sub-events, extending to other social media and news articles and building an end-to-end system that tracks social media data and provides feedback to the authority and first responders.

REFERENCES

[1] J. L. Beven, "Tropical storm harvey discussion number 2." August 2017, [Online; 17 Aug. 2017, 5:00 PM AST]. [Online]. Available: http://www.nhc.noaa.gov/archive/2017/al09/al092017.discus.002.shtml.

[2] E. S. Blake, "Tropical storm irma discussion number 2." August 2017, [Online; 30 Aug. 2017, 5:00 PM AST]. [Online]. Available: http://www.nhc.noaa.gov/archive/2017/al11/al112017.discus.002.shtml.

[3] R. Pasch, "Tropical storm maria discussion number 4." September 2017, [Online; 17 Sept. 2017, 5:00 AM AST]. [Online]. Available: http://www.nhc.noaa.gov/archive/2017/al15/al152017.discus.004.shtml.

[4] "Man captures dramatic rescue at mexico school on video," September 2017. [Online]. Available: https://www.cbsnews.com/news/mexico-earthquake-school-rescue-children-caught-on-video-enrique-rebsamen/.

[5] M. Rhodan, "'please send help.' hurricane harvey victims turn to twitter and facebook."

[6] B. R. Lindsay, "Social media and disasters: Current uses, future options, and policy considerations," 2011.

[7] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, 2012.

[8] B. Stollberg and T. De Groeve, "The use of social media within the global disaster alert and coordination system (gdacs)," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 703–706.

[9] M. A. Cameron, R. Power, B. Robinson, and J. Yin, "Emergency situation awareness from twitter for crisis management," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 695–698.

[10] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[11] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 851–860.

[12] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 591–600.

[13] R. Bandari, S. Asur, and B. A. Huberman, "The pulse of news in social media: Forecasting popularity." *ICWSM*, vol. 12, pp. 26–33, 2012.

[14] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intelligent Systems*, vol. 26, no. 3, pp. 10–14, 2011.

[15] D. E. Alexander, "Social media in disaster risk reduction and crisis management," *Science and Engineering Ethics*, vol. 20, no. 3, pp. 717–733, 2014.

[16] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster-related messages in social media." in *ISCRAM*, 2013.

[17] S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu, "Tweettracker: An analysis tool for humanitarian and disaster relief." in *ICWSM*, 2011.

[18] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010, pp. 1155–1158.

[19] A. Ritter, O. Etzioni, S. Clark *et al.*, "Open domain event extraction from twitter," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1104–1112.

[20] C. Li, A. Sun, and A. Datta, "Twevent: segment-based event detection from tweets," in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 155–164.

[21] H. Abdelhaq, C. Sengstock, and M. Gertz, "Eventweet: Online localized event detection from twitter," *Proceedings of the VLDB Endowment*, vol. 6, no. 12, pp. 1326–1329, 2013.

[22] S. Wang, P. Giridhar, L. Kaplan, and T. Abdelzaher, "Unsupervised event tracking by integrating twitter and instagram," in *Proceedings of the 2nd International Workshop on Social Sensing*. ACM, 2017, pp. 81–86.

[23] S. Wang, P. Giridhar, H. Wang, L. Kaplan, T. Pham, A. Yener, and T. Abdelzaher, "Storyline: Unsupervised urban geo-event demultiplexing on social media without location information."

[24] J. Weng and B.-S. Lee, "Event detection in twitter." *ICWSM*, vol. 11, pp. 401–408, 2011.

[25] X. Dong, D. Mavroeidis, F. Calabrese, and P. Frossard, "Multiscale event detection in social media," *Data Mining and Knowledge Discovery*, vol. 29, no. 5, pp. 1374–1405, 2015.

[26] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 177–186.

[27] M. De Choudhury, S. Counts, and M. Czerwinski, "Find me the right content! diversity-based sampling of social media spaces for topic-centric search." in *ICWSM*, 2011.

[28] G. Arru, D. Feltoni Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti, "Signal-based user recommendation on twitter," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 941–944.

[29] S. C. Olhede and A. T. Walden, "Generalized morse wavelets," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2661–2670, 2002.

APPENDIX A
CONTINUOUS WAVELET TRANSFORM

The formal definition of CWT of a continuous time-function $x(t)$,

$$X_w = \int_{-\infty}^{\infty} x(t) \bar{\psi}(t) dt \tag{1}$$

where $\bar{\psi}(t)$ is either a real or a complex function, known as the mother wavelet. The mother wavelet works as the source function to generate a group of child wavelets, which are then scaled and shifted versions of the mother wavelet. The child wavelet at scale a with a shift b is given by,

$$\psi_{a,b}(t) = \frac{1}{|a|^{1/2}} \psi\left(\frac{t-b}{a}\right) \tag{2}$$

where $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$.

APPENDIX B
SHANNON WAVELET ENTROPY

SWE of a T_{event} can be calculated by the following steps [24].

Let, j, k be the scale and time components of a wavelet and $C_j(k)$ be the wavelet coefficient at time k and scale j, then the wavelet energy E_j can be calculated by,

$$E_j = \sum_k |C_j(k)|^2 \tag{3}$$

The total energy can be calculated by,

$$E_{total} = \sum_j E_j \tag{4}$$

The Relative Wavelet Energy (RWE) at scale j can be retrieved by,

$$\rho_j = \frac{E_j}{E_{total}} \tag{5}$$

RWE represents the distribution of wavelet energy across different scales. We can calculate SWE of a signal s at for a window w by summing over the entropy of the RWE values of all the scales.

$$SWE(s_w) = -\sum_j \rho_j \cdot log\rho_j \tag{6}$$

The normalized SWE is known as H-measure,

$$H = SWE(s)/SWE_{max} \tag{7}$$

Total H-measure for an event,

$$H_{total} = \sum_1^m H, \text{ where } m \in U_{event} \tag{8}$$