

Summer 2017

Presence of Social Presence during Disasters

Alivelu Mukkamala

IT University of Copenhagen Copenhagen S, Denmark, ammu@itu.dk

Roman Beck

IT University of Copenhagen Copenhagen S, Denmark, romb@itu.dk

Follow this and additional works at: <http://aisel.aisnet.org/pacis2017>

Recommended Citation

Mukkamala, Alivelu and Beck, Roman, "Presence of Social Presence during Disasters" (2017). *PACIS 2017 Proceedings*. 135.
<http://aisel.aisnet.org/pacis2017/135>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2017 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Presence of Social Presence during Disasters

Completed Research Paper

Alivelu Mukkamala
IT University of Copenhagen
Copenhagen S, Denmark
ammu@itu.dk

Roman Beck
IT University of Copenhagen
Copenhagen S, Denmark
romb@itu.dk

Abstract

During emergencies, affected people use social media platforms for interaction and collaboration. Social media is used to ask for help, provide moral support, and to help each other, without direct face-to-face interactions. From a social presence point of view, we analyzed Twitter messages to understand how people cooperate and collaborate with each other during heavy rains and subsequent floods in Chennai, India. We conducted a manual content analysis to build social presence classifiers comprising intimacy and immediacy concepts which we used to train a machine learning approach to subsequently analyze the whole dataset of 1.65 million tweets. The results showed that the majority of the immediacy tweets are conveying the needs and urgencies of affected people requesting for help. We argue that during disasters, the online social presence creates a sense of responsibility and common identity among the social media users to participate in relief activities.

Keywords: Social media, social presence, disasters, and disaster management.

Introduction

In recent years, social media has significantly changed the way we interact on the Internet. It has achieved popularity because it allows the creation and exchange of end-user-generated content, quick information dissemination, and most importantly, publicly available content (Kaplan and Haenlein 2010). Social media platforms such as Facebook or Twitter are even used in situations of emergency either to communicate or to share information with their followers. Through social media, people can collaboratively engage in disaster response situations during an emergency such as when hurricane Sandy hit the east coast of U.S (Kryvasheyev et al. 2016), during earthquakes in Haiti (Sarcevic et al. 2012; Zook et al. 2010), China (Qu et al. 2011), and Japan (Toriumi et al. 2013), during wildfires in southern California (Sutton et al. 2008), grassfires in Oklahoma (Vieweg et al. 2010b), during floods in U.S (Starbird et al. 2010), Australia (Cheong and Cheong 2011), and Pakistan (Murthy and Longwell 2013), and more recently during floods in Chennai, India (Pandey 2015).

One social media platform gained specifically importance in emergencies for information dissemination, and that is Twitter. It enables people to share opinions, experiences, and information through short text messages. On Twitter, people involve themselves in conversations by posting messages, following others or retweeting others' messages. Most importantly, anyone can follow any other without having a mutual acquaintance. Overall, it facilitates interactions and conversations among its users (Boyd et al. 2010).

Although people use Twitter in their personal life for entertainment purposes, it is evident that at times of disasters people are using Twitter to share situation updates and to provide emotional support on a large scale (Verma et al. 2011; Vieweg et al. 2010a). Twitter has gained a lot of attention from research communities, especially its usage during times of disasters, to understand how information is generated (Starbird et al. 2010), how different types of messages are shared by people (Qu et al. 2011), and how credible the information is (Oh et al. 2010). Thus, Twitter is changing the traditional communication practices during emergencies because of real-time user generated content which is enhancing the collective collaboration (Vieweg et al. 2008) through building social capital.

Considering the media-related component of social media, based on the degree of social presence, different social media applications such as blogs, Wikipedia, YouTube, second life, and social networking sites were classified on a continuum between low and high social presence. Based on it, social networking sites fall into the “medium” category because users can connect and communicate with others through text messages and can also upload videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). During disasters, despite the lack of face-to-face interactions, people ask for help, provide moral support, and even help each other on social media platforms. It is important to explore what makes people to perceive and feel for the others and lend their hand, by reading the short text messages. So far research focus was to understand the social presence on online learning (Gunawardena and Zittle 1997; Tu and McIsaac 2002) and on social media (Al-Ghaith 2015; Xu et al. 2012) by applying survey strategies or interviews. Especially during disasters, when people seek help and support on social media, other online users feel intimacy and immediacy for those affected people and provide support by sharing the relevant information online and actively participate in relief activities. Without the feeling of social presence neither people who are in need ask for help nor people come forward to lend their hand.

Emergency management agencies (EMA) such as FEMA or the Red Cross can exploit real-time information from self-organized communities on social media during disasters. It is also envisioned that due to the potential usefulness of social media, EMA will increase social media use (Kaewkitipong et al. 2016). It is quiet beneficial for EMA to collaborate with self-organized communities to implement the relief measures effectively. In doing so, it is important to understand what makes people build relationships, cooperate, and collaborate with each other to face the threatening conditions during times of emergencies. In this regard, using the theory of social presence as a methodological framework, this paper shows how seemingly ephemeral and hastily written text messages on Twitter can create a feeling of intimacy and immediacy which are an important aspect to share information at community level, which will be useful to disaster management agencies.

In order to grasp the severity of the situation and act collaboratively to organize and to participate in emergency response activities, people primarily need to perceive, sense and empathize with others on social media, which is similar to the related concepts of social presence (Short et al. 1976). While there are a few studies which have used a questionnaire-based approach in their social presence research on social media (Al-Ghaith 2015; Xu et al. 2012) there is no research so far that has explored social presence based on a content analysis of tweets during disasters. The perceived social presence by people on social media is important. Thus, our research question is:

How social presence can be detected through a content analysis of tweets and which role does it play for building relationships, cooperate, and collaborate during times of emergencies?

In order to understand how people express intimacy and immediacy as forms of social presence in times of disasters, we analyzed approximately 1.65 million tweets from a devastating flooding in Chennai, India, which took place in December 2015. We applied a supervised machine learning approach to analyze the content of the 1.65 million tweets. In doing so, we will also illustrate how machine learning can be applied to analyze large volumes of textual content for exploring theoretical concepts such as social presence.

The remainder of this paper is structured as follows: Section two provides the theory of social presence. The main focus in section three is about the methodology we applied which is of three-fold: 1) operationalizing social presence in social media, 2) conducting manual content analyses to develop training dataset for message classification, 3) training and using a Naïve Bayes machine-learning approach to classify our dataset. In section four, the empirical analysis and results will be presented and subsequently, we conclude the paper with discussion, limitations, and future research.

Literature Background

Social presence

Social presence can be traced back to telecommunications research in the 1970s (Lowenthal 2009; Short et al. 1976) where it was viewed as a media characteristic (Kehrwald 2008). Social presence can be defined as “the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships” (Short et al. 1976). In the communication context, the degree of salience indicates the perceived feeling or significance of the other person being present in the interaction (Kehrwald 2008). The quality of a communication medium plays an important role as it can determine the way people interact and communicate. Hence, from this perspective, the degree of

social presence of a communication medium is assessed based on how well a medium can transmit information of non-verbal cues, facial expressions, posture and attire (Gunawardena 1995). It is believed that each communication medium is different in its degree of social presence, for example, audio vs. video (Lowenthal 2009). Social presence is typically associated mainly with two concepts; namely intimacy (Argyle and Dean 1965) and immediacy (Wiener and Mehrabian 1968). The former concept describes how people act and come close during social interactions while the latter concept focuses on interpersonal communication and communicative behavior (Short et al. 1976).

Equilibrium of intimacy develops between any pair of people, where the joint function of mutual exchange of smile, conversation, eye contact, or physical distance occurs. People alter their behavior to maintain the intimacy whenever one of the functions changes (Argyle and Dean 1965). However, immediacy is a measure of the psychological distance, which an individual places between himself and his target audience (Wiener and Mehrabian 1968). The selected communicative behavior of an individual leads to physical or psychological closeness in interpersonal communication (Wiener and Mehrabian 1968; Woods and Baker 2004). As an example, the use of television enhances intimacy to a greater degree than radio (Short et al. 1976). Along the same lines, it can be argued that a person can create an impression of formal or informal attitude while speaking with someone on the phone. In other words, a person can convey immediacy or non-immediacy through verbal and non-verbal communication while speaking with someone on the phone (Aragon 2003; Cobb 2009; Gunawardena and Zittle 1997; Tu 2002). Therefore, social presence can be viewed as an attribute of the media in question as well as that of the communicators and their presence in a sequence of interactions (Biocca et al. 2003; Gunawardena and Zittle 1997).

Over time, with the rise of computer-mediated communication (CMC), social presence theory altered and gained importance in online learning disciplines (Lowenthal 2009). CMC facilitates social interactions through text-based content. However, from the social presence theory point of view, a text-based CMC could be relatively low in social presence when compared to face-to-face interactions. Although the new technologies are more effective in information processing, transmission, and user experience, it is unclear yet “how the social meaning of interactions is affected in the absence of nonverbal cues when communicators substitute text-based electronic messaging for face-to-face encounters” (Walther et al. 2005, p. 1). However, when it comes to text-based media such as e-mail or chat, which are considered richer than face-to-face conversations (Walther 1992). Moreover, mediated communications play an important role to perceive the other person as real while communicating (Gunawardena 1995). Thus, the theory of social presence that originates from media studies has been often applied to examine interactions between students and teachers in the context of online learning (Tu and McIsaac 2002). Most importantly, in online learning, online social presence is conveyed through the messages sent by the online participants and the interpretation of those messages by others. The visible activities, such as posting messages, replying and responding to others, and participating in the activities of the group contain the cues of social presence of the individuals who send them and who receive them (Kehrwald 2008). This confirms the fact that despite the lack of existence of non-verbal cues in online environments, individuals grasp cues through language, style, and other cues to build relationships (Walther et al. 2005). Thus, the theory of social presence can be used to explain and understand how people interact within online learning environments while there is still a lack of understanding on how to properly detect and measure social presence in social media environments. In consequence, social presence has been defined in different ways in online learning research in the past (Kehrwald 2008; Lowenthal 2009). It is a known fact that social networking sites fall into the “medium” category because of the richness of different attributes such as textual content, videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). While social presence has been used in research on social networks before, it is unclear which role social presence plays in times of emergencies, e.g., to stimulate relief activities. However, previous studies mentioned that online users’ interactions with other users and their engagement is directly related to social presence (Lim et al. 2015) and that it plays an important role in fulfilling social connection needs in online environments (Han et al. 2015).

Operationalization of Social Presence

On Twitter, a user can “follow” and “followed” by anyone without mutual acquaintance. A user can “retweet” (RT) and also “like” (favorite) some one’s status updates. Users can form groups using “list” and can use “mention” to send public messages. The hashtags, with prefix “#” empowers users to connect with a group and enhances coordination among users for a common cause and also increases the search ability (Starbird et al. 2015). For example, at the times of emergency situations, a victim of a disaster or a person who is witnessing it can send a message for help or update his/her status by

posting a message related to the disaster. Thus, a user can tweet a message for help or explain the situation. People who want to help can forward the same message as RT@ and also can reply using @xxx to provide help.

Social Presence

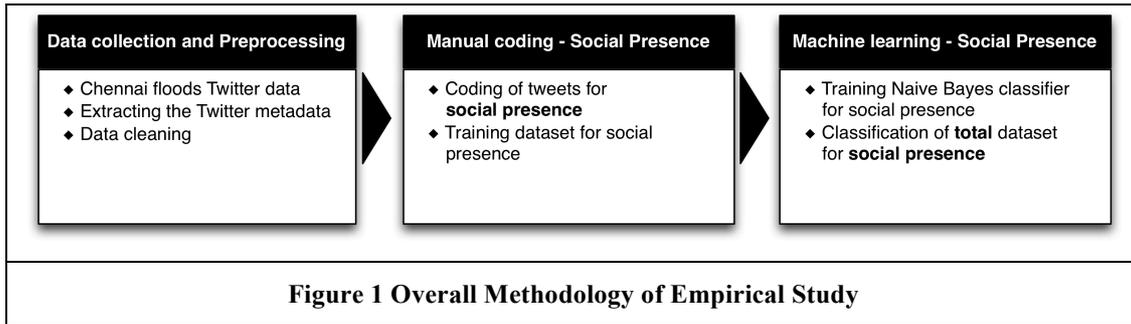
All the above features and the characteristics of Twitter are facilitating interactions and conversations with others. In general, if we perceive Twitter and its features as media-related component, then the features create a sense of intimacy and immediacy based on its content. However, a tweet could also create a feeling of intimacy and immediacy based on the textual content. The online social presence is conveyed through the messages sent by the online participants and the interpretation of those messages by others (Kehrwald 2008).

In order to operationalize the concept of “intimacy”, we analyze it from an angle of affective intimacy that explains how people express their feeling of closeness through liking and emotional bonding by providing moral support (Hu et al. 2004; Tolstedt and Stokes 1983). In other words, at times of emergency situations, people post messages to express moral support to combat the situation, or in situations where people go through the same bad conditions, provide a feeling of closeness, e.g., through providing information about road blocks during earthquakes or floods, to create awareness. Although immediacy means perceived psychological distance people feel while communicating with others (Mehrabian 1967; Walther 1992), immediacy is expressed also through situations that give rise for a sense of urgency or excitement and involve people instantly in the action. In general, people perceive psychological closeness towards immediate family members or friends but whenever the whole community faces an emergency, people feel psychological closeness towards affected people such as neighbors or the community itself. Therefore, people come forward to provide help in times of emergency situations as they perceive the closeness, which creates a sense of urgency and importance to act immediately to the people who are in vulnerable conditions such as in need for shelter or food in life threatening situations. Some of the example tweets for social presence reflecting intimacy and immediacy are provided in Table 1.

Table 1 Overview of the Social Presence concepts and their operationalization		
categories	Description	Tweet Content
Intimacy	Feeling closeness: sharing road closure info, asking for help on behalf of others	Is there any way to provide any form of support monetary and supplies? #ChennaiRains
	Moral support: stand by people, providing hope for best	Hats off to the fighting spirit of Chennai! A salute all those volunteers who have been helping relentlessly! #staystrongchennai #chenna
Immediacy	Urgent action is needed: different types of rescue requests	any doctors in mudichur area? One pregnant lady in labour...no access to boat. Here is the Contact: 9940203871#chennairains
	Sharing information: provide shelter, food and help	#food available at #tnagar gurdwara 9094790989#ChennaiRainsHelp #ChennaiMicro #ChennaiVolunteer https://t.co/JD1BAdWXSe

Disaster Description and Applied Methodology

Chennai, a city in southern India, received a devastatingly high amount of rainfall in December 2015. Especially in the first few days of December, the rainfall intensified and Chennai received 34 times more than the normal daily amount of rain (Misra 2015). According to reports, the flooding caused not only severe economic damages but also severely disrupted the whole city infrastructure such as homes, hospitals, roads, railway tracks and the international airport. During the flooding, people in Chennai used social media to coordinate and to cooperate in relief activities. For example, it was used to rescue people who were stranded in floods, for food distribution, to provide shelters, and also to reach out to people who needed help (Pandey 2015). For our analytics, we focus on user-generated content we harvested from Twitter.



In this section, we will illustrate the methodology to build the classifier and how we analyzed social presence through Twitter messages to have the sense of identity in emergency situations. Specifically, we conducted a manual coding approach as well as machine learning as shown in Figure 1. In order to classify tweet messages for social presence. In the following, we will discuss the depicted phases.

Data Collection and Preprocessing

We used the social media data collection tool Radian 6 to collect the Twitter messages. In case of disasters, hashtags are created during or soon after the incident unfolds to share and communicate information regarding disasters. Hence, we used the hashtags #TNflood, #chennaiRains, #chennafloods #chennaiRainsHelp, #IndiaWithChennai and #chennaiMicro to extract related tweets. The timeline for the collected data was from November 30th to December 16th, 2015 with a total dataset consisting of 1.65 million tweets posted by 209,644 unique users as shown in Table 2. The Radian 6 tool provided some Twitter attributes such as tweet ID, author, content, and followers count, but it does not provide metadata such as retweet status, retweet count, and original tweetId for retweets. Hence, we again downloaded the whole dataset via the open Twitter API using tweet-Ids with the help of a custom tool and reanalyzed the dataset to segregate the original tweets from the retweets based on the retweet status information. As part of the data pre-processing, we deleted the tweets containing some of the mentioned hashtags, but which fall outside the time period mentioned above. Moreover, we also extracted the original tweet-Id for retweets by processing the Twitter data.

Twitter messages (tweets)	Absolute numbers	Percentage
Total tweets	1,658,220	100%
Retweets	1,226,098	73.94%
Original tweets of Retweets	141,941	08.56%
Tweets never got retweeted	290,181	17.50%
Mean Retweet Ratio	8.64	
Total Unique Twitter Users	209,644	

Most of our dataset consists of “retweets” as shown in Table 2 which constitutes around 74% of the total dataset. Moreover, only 8.5% tweets are original tweets that got retweeted (73.94%) many times, at an average retweet rate of 8.64.

Manual Coding and Content Analysis

We have conducted a manual content analysis in two phases as shown in Figure 1 to analyze social presence. A content analysis helps researchers to have their own context of inquiry and constructs to make the texts more meaningful. Through this approach, one can make replicable, reliable, and valid inferences from data on an aggregate level that opens an avenue to understand trends, patterns, and differences (Krippendorff 1989; Lombard et al. 2002). For our approach, we followed a directed content analysis where we have drawn the coding scheme from existing theory (Hsieh and Shannon 2005; Risius et al. 2015; Risius and Beck 2014).

To establish principles along the entire process of our manual content analysis and to ensure obtained measures and results to be more valid and reliable, we followed Morris' 5-step process (Morris 1994) for our content analysis. This approach not only guided us through a step-wise iterative research process but also made the whole process more transparent. In general, the unit of analysis can be a word or a sentence. In the first step, we decided to take the whole tweet content as unit of analysis, as a tweet can be objectively identifiable by coders (Rourke et al. 2001). In the second step, drawing on existing theory, we developed the categories as coding scheme for the data classification. The categories for social presence are 1) intimacy (in), 2) immediacy (im), and 3) none (n). We introduced none label in categories to filter out the tweets that don't belong to the categories. To ensure validity, both researchers who upfront have intensively worked on the theoretical background discussed extensively what constitutes the categories and what does not. The third step basically enhances the coders' familiarity with the coding scheme and also acts as a training session for independent coders. Therefore, one of the coders coded a sample of 100 tweets in both phases, and later on both coders together analyzed the results to exclude the subjective bias and discrepancies.

In the fourth step, after reaching a consensus, both coders independently coded a randomly selected sample of approximately 500 (580) tweets for intimacy and immediacy. Afterwards, the results were compared and both coders discussed about the tweets to clear the discrepancies about the concepts. After an iteration of coding, the inter coder agreement matrix for the texts of social presence coded by the two different coders is presented in Table 3. The integer values in Table 3 represent the number of texts coded by each coder under different categories, whereas the decimal values shown in parentheses represent proportions of the categorized texts to the total coded texts. The inter coder agreement value can be simply measured by considering the agreement rates between different coders, which can be calculated by the proportion of the total number of agreed texts by the total number of texts coded. For example, in case of social presence, the inter-coder agreement value is $(38 + 83 + 431) / 580 = 0.95$. However, this proportion does not account for discrepancies that might have occurred by chance i.e. where the coder's agreement for the texts might have expected on the basis of chance.

In order to eliminate this limitation and to ensure the validity of inter-coder agreement, we have calculated Cohen's Kappa value (Cohen 1960; Stemler 2001). The Cohen's Kappa value varies from 0 to 1, where a value of 1 indicates perfectly reliable agreement and the value 0 indicates that there is no agreement between the coders other than what is expected on the basis of chance. The calculation of Cohen's Kappa involves two variables: proportion of texts where both coders agree (inter-coder agreement as calculated previously) and proportion of texts for which agreement happened by chance. The proportion of agreement by chance can be calculated by multiplying the proportions of marginal totals of each category and then summing them up. For example, in case of social presence, the proportion of agreement by chance = $(0.08 * 0.08) + (0.16 * 0.16) + (0.76 * 0.76) = 0.62$. Finally, Cohen's Kappa value for social presence can be calculated by using the standard formula as: $(0.95 - 0.62) / (1 - 0.62) = 0.868$.

		Coder 1			Marginal Totals
		Intimacy	Immediacy	None	
Coder 2	Intimacy	38 (0.06)	7 (0.01)	2 (0.01)	47 (0.08)
	Immediacy	7 (0.01)	83 (0.14)	6 (0.01)	96 (0.16)
	None	1 (0.01)	5 (0.01)	431 (0.74)	437 (0.76)
Marginal Totals		46 (0.08)	95 (0.16)	439 (0.76)	580 (1.00)

In general, an inter-coder agreement of 0.40 to 0.80 is considered as a good indicator of valid agreement between coders (Stemler 2001). In the final step of content analysis, one of the researchers processed randomly selected 5000 tweets for social presence in the subsequent phase according to categorization of coding scheme to get the trained dataset. Thus, in this phase we conducted a manual coding of randomly selected tweets from the whole dataset to prepare a training dataset for social presence (intimacy and immediacy). Using the trained dataset of social presence, we applied machine learning to classify the whole dataset for social presence.

Tweet Classification using Naïve Bayes Classifier

We have adopted a supervised machine learning approach to classify the tweet content of the dataset. Text classification can be defined as a process that comprises of assigning a predefined category of labels to new texts or documents based on a probabilistic measure of likelihood using a training set of labeled texts (Yang 1999). We have used a Naïve Bayes classifier, which is a probabilistic classifier that will estimate the probability of a given text based on the joint probabilities of words and categories using a bag of words approach. The naïve part of the classifier is that it assumes that the conditional probability of a word given a category is independent from the conditional probabilities of other words given in that category. The Naïve Bayes assumption makes the classifier far more efficient and practical than the exponential complexity of other classifiers and also it works quite well for the text classification with a fair amount of accuracy and therefore it stands as one of the most used techniques for text classification (Yang and Liu 1999; Zhang and Li 2007). For classification of tweets using a Naïve Bayes classifier, we used the Natural Language Toolkit (NLTK) (Bird 2006) and Python as programming language to train the classifier using the training datasets containing manually coded tweets as explained in the previous section.

The classification of tweets was conducted for social presence containing labels as intimacy and immediacy. A training set of 5000 manually coded tweets was used, where 80% of the training set was used to train the classifier and the rest 20% of the tweets were used to test the accuracy of the classifier. The text classification is performed in several iterations and within each iteration adopting different strategies (Narayanan et al. 2013) to enhance the accuracy of the classifier. Even though, initially we started with a strategy of using all the words from the training set, but in subsequent iterations we employed strategies such as stop words removal, bigram association measures, feature selection and others to find out a suitable strategy that will suit to our text corpus and yields the best accuracy for Naïve Bayes classifier. Finally, we adopted bigram association measures with removal of stop words and proper nouns (such as names of the people, places etc.) which yielded the best accuracy out of all the strategies for the text classification.

In general, performance of a machine learning algorithm can be described by four measures: precision, recall, F-measure, and accuracy (Powers 2011; Yang 1999). These measures are built over the statistical variables True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). True and False Positives refer to the number of predicted values that were correctly identified and incorrectly identified, whereas True and False Negatives refer to the predicted values that are correctly rejected and incorrectly rejected respectively. Building on these variables, *precision* is defined as the ratio of predicted positive values that are correctly identified as real positive values i.e. $TP/(TP+FP)$. Similarly, *recall* (also known as true positive rate) is defined as a ratio of correctly predicted positive values over all positive values which is $TP/(TP+FN)$. Moreover, F-measure or F-score is a trade-off between precision and recall and it is defined as a single measure which is the harmonic mean of the precision and recall. All these measures precision, recall and F-measure provide performance information at the level of labels, whereas *accuracy* of the classifier provides information about overall performance of the classifier, which can be defined as $(TP + TN) / (TP+FP+TN+FN)$.

Results and Analysis

We will present results and analysis of the text classification of tweet content of twitter data.

Model	Labels	Precision	Recall	F-Measure	Accuracy
Social Presence	Intimacy	0.158	0.466	0.236	0.805
	Immediacy	0.626	0.520	0.568	
	None	0.943	0.858	0.898	

The performance measures about text classification of tweet content using Naïve Bayes classifier are presented in Table 4. The overall accuracy of the classifier is fairly high (i.e. around 80% of the prediction are accurate for correct predictions). In terms of precision and recall, *Immediacy* received fairly good values than *intimacy*. In case of F-measure, a value around 0.6-0.7 indicates a fairly better performance and the F-measure values for all labels/categories indicate reasonably good performance except for the categories: *intimacy*.

Model	Labels	Most informative features/words
Social Presence	Intimacy	elderly, child, glad, responsive, struck, requesting, praying, help, bravery, aged, catastrophic, wife, together, ...
	Immediacy	tonight, labour, pregnant, accommodate, rides, blankets, mall, ambulance, parcels, surgery, packets, inside, needed, evacuate, hot, boat, distribute, immediately
	None	media, india, victims, well, news, money, hope, government, chennaifloods, day, helpchennai, traffic

Furthermore, we also extracted the most informative features (i.e. the words with high probabilities) that the classifier used to classify the tweet messages into categories of social presence. A few of the top of such words from each category are gathered and presented in Table 5 for illustration of the nature of words that were used by the classifier. An obvious distinction between the words of intimacy and immediacy is that the words in the immediacy category are more related to the needs and urgencies of the effected people, such as *blankets*, *ambulance*, *pregnant*, and so on, whereas the words in the intimacy category are more related to expressing concerns and moral support. The words in both none categories are more related to opinions and criticism as the words indicate *media*, *news*, *government*, *citizens*, *India* and so on.

In the classification of tweets for social presence, we have applied the classifier on the whole dataset, with the results depicted in Table 6. Out of 1.65 million tweets, 37% of tweets are classified as social presence containing intimacy and immediacy categories, whereas the remaining 63% of tweets belong to the none category. This is consistent with previous research that states that during disasters people post and share different types of messages containing suggestions, comments, criticism, etc., (Qu et al. 2011; Vieweg 2012) and also discussions about either media or government, frustrations or anger which could be not categorized as social presence. However, when it comes to retweets which constitute 80% of the dataset, the proportion of social presence retweets is higher with around 42% of retweets belong to intimacy and immediacy, as elaborated in Table 6.

Social Presence	Total tweets	Percentage	Retweets	Percentage
Intimacy	285,292	17.20%	226,697	18.49%
Immediacy	335,877	20.26%	291,516	23.78%
None	1,037,051	62.54%	707,885	57.73%
Total	1,658,220	100%	1,226,098	100%

The reason for having more retweets in the dataset is that retweet “RT” allow users to mention the original tweet author’s name and also acts as re-broadcasting or forwarding original tweets. Moreover, originators of those retweets are in general tweeting from the affected zone (Starbird and Palen 2010) and through this informal communication, people want to pass on the important information of disastrous events such as shelter information, coordination activities, and personal experiences that are useful to other affected people (Starbird and Palen 2010; Vieweg et al. 2010a).

Discussion and Conclusion

Emergent support groups evolve immediately after a disaster or after a crisis unfolds and it has been identified that these groups collectively work together to cope with the situation (Drabek and McEntire 2003). Nowadays, during emergency situations, people are interacting and communicating rather intensively on social media. In this research, we shed light on how the online social presence creates a sense of responsibility and common identity during disasters among the social media users. According to the social presence theory, people perceive intimacy and immediacy on mediated communications (Short et al. 1976). One of the reason behind the active participation of people on online social media during emergencies could be that they perceive the online social presence through the messages sent by the online participants and the interpretation of those messages by others. In order to detect the social presence based on message content apply the theory of social presence on

social media users, we analyzed the Twitter messages and segregated them into based on intimacy and immediacy categories. First, we operationalized the concepts and then conducted manual content analysis to prepare the training sets for the classifier. Since the dataset is huge (around 1.65 million), therefore, we adopted a machine learning technique to automatically categorize the data into intimacy, immediacy and none categories. The results indicate that during emergencies, people are drawn to Twitter to fulfill their social need for connections as they feel the presence of others. Moreover, people, who feel higher levels of social presence continue to use and interact more on Twitter (Han et al. 2015). Thus, social presence plays a significant role hence despite the lack of face-to-face communication, people feel a sense of bonding due to the intimacy and immediacy felt for each other and therefore offer assistance by simply reading and retweeting the tweets. As explained previously, intimacy tweets are more towards standing by the people while showing moral support and coming forward to provide help on online, while immediacy tweets especially make people to perceive the vulnerable situation so people start actively taking part in relief activities.

Research has also revealed that during emergency situations, local people often retweet to ensure that emergency relevant information (Starbird and Palen 2010) is forwarded to others through collective action and collective behavior. Our total dataset consists of around 80% of retweets and more importantly we noticed that our analyzed dataset of social presence also contains a large number of retweets. The results showed that, the majority of the immediacy tweets conveying the needs and urgencies of affected people requesting for help. Especially in Chennai, people estimated the severity of the situation and gathered information that lead to the active participation and orchestration of needs and urgencies for affected individuals. According to our findings, even though a number of tweets fall into the “n” category, it is evident that social presence is important factor on social media to reach out people. Our results support the theory that online participants perceive others, experience a feeling of closeness, and when situations give rise for a sense of urgency people collectively get involved in the grassroots activities as well. For example, people offered accommodation, supplied and provided needs and necessities like food, water, clothes and activity participated in volunteer activities. Large private places like cinema theaters and wedding halls were opened for shelters. Another interesting observation is that people even shared their personal mobile numbers online to allow those in need to get in contact with them. Most importantly the hashtag #chennaiimicro was introduced by one of the active volunteers and requested others online to use that particular hashtag for food supply only (Pandey 2015). We argue that social presence which is conveyed through a message content especially immediacy is highly valuable in emergency situations for emergence support groups to combat the situation.

Our results are consistent with previous findings (Al-Ghaith 2015; Xu et al. 2012), where social presence has been found to have a positive impact on usage of social media, as we also observed that social media has been actively used for communication and coordination during times of disasters. Moreover, in line with previous research results, affected individuals distribute coordination and online collective action (Vieweg et al. 2008) by relying on relevant information while verifying it and later on structure the information in a meaningful way (Starbird 2013). It is also apparent from our research results that people coordinate actively support activities through social media by retweeting, as indicated by the fact that 74% of our dataset consists of retweets.

Our research contributions are of twofold. Firstly, our research contributes to the theory of social presence in particular, and to social media analytics in general. So far, the theory of social presence was predominantly used in online learning and the research focus on social presence on social media was minimal and by conducting surveys only. Most importantly, on online learning, it was mentioned that messages have cues which evokes certain feelings but none measured them from content point of view. Unlike the previous studies, which were mostly survey based methods, our work focused more from message content point of view. Thus, we are among the first to analyze tweets using social presence as theoretical lens. Furthermore, our analysis presents new insights how social presence is formed during disasters through the use of social media messages.

Our findings have also some practical implications. We argue that information in immediacy tweets which reflect the needs and urgencies of affected people is important and valuable to emergency management agencies to reach out and save people’s lives. We were able to illustrate how immediacy tweets can be automatically extracted using a machine learning approach. For emergency management agencies, those tweets contain valuable information to coordinate their support activity in close to real time.

Limitations and Future Research

Our study has certain limitations based on our data collection method. We exclusively used hashtags (#) to collect the data. This could have resulted in missing other relevant tweets in our dataset. We applied a machine learning approach with a relatively good accuracy for one of our classifier concepts (immediacy). However, the accuracy for the intimacy is rather comparatively low. During disasters, often social media data (e.g. twitter data) is produced in huge volumes, therefore our focus is to analyze the whole dataset instead of considering a small sample of the huge dataset. So we applied a technique that helps to analyze huge volumes of textual content for exploring theoretical concept such as social presence. However, there is also a certain likelihood that tweets fall in both categories (immediacy and intimacy), but the machine learning approach forces the tweets in one category only. Furthermore, the dataset used comprises only tweets from a flooding disaster and thus our findings may not be generalizable to other types of disasters. As part of our future research, we will analyze the tweets which have fallen into the “n” category. In the research at hand, we limited ourselves to only carrying out a content analysis of tweets from a social presence point of view. However there are other Twitter features that have yet to be explored through our research lens. These features, which include mentions, replies, and follower and following counts, may also exhibit social presence and collective intelligence which have yet to be explored.

References

- Al-Ghaith, W. 2015. "Understanding Social Network Usage: Impact of Co-Presence, Intimacy, and Immediacy," *International Journal of Advanced Computer Science and Applications (IJACSA)* (6:8), p. 12.
- Aragon, S. R. 2003. "Creating Social Presence in Online Environments," *New directions for adult and continuing education* (2003:100), pp. 57-68.
- Argyle, M., and Dean, J. 1965. "Eye-Contact, Distance and Affiliation," *Sociometry* (28:3), pp. 289-304.
- Biocca, F., Harms, C., and Burgoon, J. K. 2003. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence* (12:5), pp. 456-480.
- Bird, S. 2006. "Nltk: The Natural Language Toolkit," *Proceedings of the COLING/ACL on Interactive presentation sessions: Association for Computational Linguistics*, pp. 69-72.
- Boyd, D., Golder, S., and Lotan, G. 2010. "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," *System Sciences (HICSS), 2010 43rd Hawaii International Conference on: IEEE*, pp. 1-10.
- Cheong, F., and Cheong, C. 2011. "Social Media Data Mining: A Social Network Analysis of Tweets During the 2010-2011 Australian Floods," *In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*, pp. 46-46.
- Cobb, S. C. 2009. "Social Presence and Online Learning: A Current View from a Research Perspective," *Journal of Interactive Online Learning* (8:3), pp. 241-254.
- Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20:1), pp. 37-46.
- Drabek, T. E., and McEntire, D. A. 2003. "Emergent Phenomena and the Sociology of Disaster: Lessons, Trends and Opportunities from the Research Literature," *Disaster Prevention and Management: An International Journal* (12:2), pp. 97-112.
- Gunawardena, C. N. 1995. "Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences," *International journal of educational telecommunications* (1:2), pp. 147-166.
- Gunawardena, C. N., and Zittle, F. J. 1997. "Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment," *American journal of distance education* (11:3), pp. 8-26.
- Han, S., Min, J., and Lee, H. 2015. "Antecedents of Social Presence and Gratification of Social Connection Needs in Sns: A Study of Twitter Users and Their Mobile and Non-Mobile Usage," *International Journal of Information Management* (35:4), pp. 459-471.
- Hsieh, H.-F., and Shannon, S. E. 2005. "Three Approaches to Qualitative Content Analysis," *Qualitative health research* (15:9), pp. 1277-1288.
- Hu, Y., Wood, J. F., Smith, V., and Westbrook, N. 2004. "Friendships through Im: Examining the Relationship between Instant Messaging and Intimacy," *Journal of Computer-Mediated Communication* (10:1), pp. 00-00.

- Kaewkitipong, L., Chen, C. C., and Ractham, P. 2016. "A Community-Based Approach to Sharing Knowledge before, During, and after Crisis Events: A Case Study from Thailand," *Computers in Human Behavior* (54), pp. 653-666.
- Kaplan, A. M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business horizons* (53:1), pp. 59-68.
- Kehrwald, B. 2008. "Understanding Social Presence in Text-Based Online Learning Environments," *Distance Education* (29:1), pp. 89-106.
- Krippendorff, K. 1989. "Content Analysis," in *International Encyclopedia of Communication*, G.G. E. Barnouw, W. Schramm, T. L. Worth, & L. Gross (ed.). pp. 403-407.
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. 2016. "Rapid Assessment of Disaster Damage Using Social Media Activity," *Science Advances* (2:3), p. e1500779.
- Lim, J. S., Hwang, Y., Kim, S., and Biocca, F. A. 2015. "How Social Media Engagement Leads to Sports Channel Loyalty: Mediating Roles of Social Presence and Channel Commitment," *Computers in Human Behavior* (46), pp. 158-167.
- Lombard, M., Snyder-Duch, J., and Bracken, C. C. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability," *Human communication research* (28:4), pp. 587-604.
- Lowenthal, P. R. 2009. "The Evolution and Influence of Social Presence Theory on Online Learning," *Social Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, p. 113.
- Mehrabian, A. 1967. "Orientation Behaviors and Nonverbal Attitude Communication," *Journal of communication* (17:4), pp. 324-332.
- Misra, T. 2015. "Chennai Floods: A Perfect Storm of Climate Change and Failed Planning." 2016, from <http://www.citylab.com/weather/2015/12/chennai-flood-climate-change-urban-planning-cop21-paris/419349/>
- Morris, R. 1994. "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations," *Journal of Management* (20:4), pp. 903-931.
- Murthy, D., and Longwell, S. A. 2013. "Twitter and Disasters: The Uses of Twitter During the 2010 Pakistan Floods," *Information, Communication & Society* (16:6), pp. 837-855.
- Narayanan, V., Arora, I., and Bhatia, A. 2013. "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model," in *Intelligent Data Engineering and Automated Learning—Ideal 2013*. Springer, pp. 194-201.
- Oh, O., Kwon, K. H., and Rao, H. R. 2010. "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter During the Haiti Earthquake 2010," *ICIS*, p. 231.
- Pandey, V. 2015. "How Chennai Residents Are Helping Flood Affected Via Social Media." from <http://www.bbc.com/news/world-asia-india-34981360>
- Powers, D. M. 2011. "Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation," 2229-3981, School of Informatics and Engineering, Flinders University, Adelaide, Australia.
- Qu, Y., Huang, C., Zhang, P., and Zhang, J. 2011. "Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake," In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 25-34.
- Risius, M., Akolk, F., and Beck, R. 2015. "Differential Emotions and the Stock Market—the Case of Company-Specific Trading," In: *Proceedings of the European Conference on Information Systems*.
- Risius, M., and Beck, R. 2014. "You Reap What You Sow? How Knowledge Exchange Effectiveness Is Affected by Different Types of Communication in Enterprise Social Media," *System Sciences (HICSS), 2014 47th Hawaii International Conference on: IEEE*, pp. 540-549.
- Rourke, L., Anderson, T., Garrison, D. R., and Archer, W. 2001. "Methodological Issues in the Content Analysis of Computer Conference Transcripts," *International Journal of Artificial Intelligence in Education (IJAIED)* (12), pp. 8-22.
- Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., and Anderson, K. 2012. "Beacons of Hope in Decentralized Coordination: Learning from on-the-Ground Medical Twitterers During the 2010 Haiti Earthquake," *Proceedings of the ACM 2012 conference on computer supported cooperative work: ACM*, pp. 47-56.
- Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*.
- Starbird, K. 2013. "Delivering Patients to Sacré Coeur: Collective Intelligence in Digital Volunteer Communities," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: ACM*, pp. 801-810.

- Starbird, K., Dailey, D., Walker, A. H., Leschine, T. M., Pavia, R., and Bostrom, A. 2015. "Social Media, Public Participation, and the 2010 Bp Deepwater Horizon Oil Spill," *Human and Ecological Risk Assessment: An International Journal* (21:3), pp. 605-630.
- Starbird, K., and Palen, L. 2010. *Pass It On?: Retweeting in Mass Emergency*. International Community on Information Systems for Crisis Response and Management.
- Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. 2010. "Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information," *In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*, pp. 241-250.
- Stemler, S. 2001. "An Overview of Content Analysis," *Practical assessment, research & evaluation* (7:17), pp. 137-146.
- Sutton, J., Palen, L., and Shklovski, I. 2008. "Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires," *In: Proceedings of the 5th International ISCRAM Conference*: Washington, DC, pp. 624-632.
- Tolstedt, B. E., and Stokes, J. P. 1983. "Relation of Verbal, Affective, and Physical Intimacy to Marital Satisfaction," *Journal of Counseling Psychology* (30:4), p. 573.
- Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S., and Noda, I. 2013. "Information Sharing on Twitter During the 2011 Catastrophic Earthquake," *In: Proceedings of the 22nd international conference on World Wide Web companion*, pp. 1025-1028.
- Tu, C.-H. 2002. "The Impacts of Text-Based CMC on Online Social Presence," *The journal of interactive online learning* (1:2), pp. 1-24.
- Tu, C.-H., and McIsaac, M. 2002. "The Relationship of Social Presence and Interaction in Online Classes," *The American journal of distance education* (16:3), pp. 131-150.
- Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., and Anderson, K. M. 2011. "Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency," *ICWSM: Citeseer*.
- Vieweg, S., Hughes, A., Starbird, K., and Palen, L. 2010a. "Supporting Situational Awareness During Emergencies Using Microblogged Information," *SIGCHI 2010*.
- Vieweg, S., Hughes, A. L. A., Starbird, K., and Palen, L. 2010b. "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," *In: Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*: ACM, pp. 1079-1088.
- Vieweg, S., Palen, L., Liu, S. B., Hughes, A. L., and Sutton, J. 2008. "Collective Intelligence in Disaster: An Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shootings," *Proceedings of the information systems for crisis response and management conference (ISCRAM)*.
- Vieweg, S. E. 2012. "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications." University of Colorado at Boulder.
- Walther, J. B. 1992. "Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective," *Communication research* (19:1), pp. 52-90.
- Walther, J. B., Loh, T., and Granka, L. 2005. "Let Me Count the Ways the Interchange of Verbal and Nonverbal Cues in Computer-Mediated and Face-to-Face Affinity," *Journal of language and social psychology* (24:1), pp. 36-65.
- Wiener, M., and Mehrabian, A. 1968. *Language within Language: Immediacy, a Channel in Verbal Communication*. Ardent Media.
- Woods, R. H., and Baker, J. D. 2004. "Interaction and Immediacy in Online Learning," *The International Review of Research in Open and Distributed Learning* (5:2).
- Xu, C., Ryan, S., Prybutok, V., and Wen, C. 2012. "It Is Not for Fun: An Examination of Social Network Site Usage," *Information & Management* (49:5), pp. 210-217.
- Yang, Y. 1999. "An Evaluation of Statistical Approaches to Text Categorization," *Information retrieval* (1:1-2), pp. 69-90.
- Yang, Y., and Liu, X. 1999. "A Re-Examination of Text Categorization Methods," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, pp. 42-49.
- Zhang, H., and Li, D. 2007. "Naive Bayes Text Classifier," *Granular Computing, 2007. GRC 2007. IEEE International Conference on*: IEEE, pp. 708-708.
- Zook, M., Graham, M., Shelton, T., and Gorman, S. 2010. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake," *World Medical & Health Policy* (2:2), pp. 7-33.