What Kind of #Communication is Twitter? A Psycholinguistic Perspective on Communication in Twitter for the Purpose of Emergency Coordination

Hemant Purohit  
Wright State University - Main Campus

Andrew Hampton  
Wright State University - Main Campus

Valerie L. Shalin  
Wright State University - Main Campus, valerie.shalin@wright.edu

Amit P. Sheth  
Wright State University - Main Campus, amit.sheth@wright.edu

John Flach  
Wright State University - Main Campus, john.flach@wright.edu
Follow this and additional works at: http://corescholar.libraries.wright.edu/knoesis

Part of the Bioinformatics Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, OS and Networks Commons, and the Science and Technology Studies Commons

Repository Citation
http://corescholar.libraries.wright.edu/knoesis/618

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu.
What kind of #communication is Twitter? A Psycholinguistic perspective on communication in Twitter for the purpose of emergency coordination

H. Purohit, A.J. Hampton, V. L. Shalin, A. Sheth & J. Flach
Ohio Center Of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University

Problem
The present research aims to detect coordinated citizen response within social media traffic to assist emergency response. We use domain-independent linguistic features as the first step in narrowing the candidate set of messages for domain-dependent and computationally intensive analysis.

Motivation
Properties of an exchange, including opening and closing phrases, anaphora, and deixis reveal conversational coordination, and hence the emergence of a new informal community conducting relevant coordinated activity.

Approach Overview
To examine the diagnosticity of conversational features in Social Media (Twitter), we require a systematic, independent approach to the classification of traffic as positive and negative instances of conversation. We approach this problem by sorting message traffic using platform properties that we expect to correlate with conversation: Reply, Retweet and Mention of another Twitter user. The absence of such properties suggests the absence of conversation. This approach is only heuristic. Platform properties such as Reply simply facilitate distribution, while Retweeting constitutes a kind of reply to the initial sender with unclear implications for the next set of recipients. Further, the absence of platform indicators does not preclude conversation. Because the corpus is imperfectly sorted, we expect less than perfect discrimination. However, the heuristic sorting promises empirically-based rules that transcend reliance on platform indicators.

Method
Data sets: A Twitter Search API crawler enables the collection of event specific postings using keywords (see below).

<table>
<thead>
<tr>
<th>Event</th>
<th>Duration</th>
<th>Tweets</th>
<th>Authors</th>
<th>Tweets Depicted</th>
<th>Tweets Mentioned</th>
<th>Tweets Mentioned in Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan Tsunami</td>
<td>2011-03-11</td>
<td>20</td>
<td>99953</td>
<td>20586</td>
<td>20586</td>
<td>54717</td>
</tr>
<tr>
<td>Earthquake</td>
<td>2010-08-12</td>
<td>57</td>
<td>96747</td>
<td>36840</td>
<td>55684</td>
<td>271725</td>
</tr>
<tr>
<td>Hurricane</td>
<td>2011-08-26</td>
<td>16</td>
<td>161971</td>
<td>63206</td>
<td>14603</td>
<td>72841</td>
</tr>
<tr>
<td>Cyclone</td>
<td>2011-08-26</td>
<td>6</td>
<td>99965</td>
<td>10960</td>
<td>3296</td>
<td>26057</td>
</tr>
<tr>
<td>Earthquake</td>
<td>2011-03-08</td>
<td>7</td>
<td>36562</td>
<td>2064</td>
<td>364</td>
<td>2149</td>
</tr>
</tbody>
</table>

Linguistic features: We examined the following hypotheses as potentially diagnostic of conversation: determiners (H1: the, H2: a, an), first and second person pronouns (H3: you, we, H4: my), other pronouns (H5: this, that, those, they, them, whom), dialogue management and speech acts (H6: thanks, yes, ok, sorry, hi, hello, bye, anyway, could you, can you, will you, how about) word counts (H7) and hedge words (H8: kinda, sorta, what do you mean, whatever).

Classifier & Feature Ranking: We performed classification modeling to establish the degree of conversationality shown by a potential conversation text sample, based on linguistic features corresponding to text samples. In order to build our classifier, we created training sets (to learn from the data) and testing sets (to test on the new data and make a more robust classifier) of the data samples. We performed a correlation based study by ranking the linguistic features that reflected significant alignment with the conversation class suggested by any of C1, C2 or C3 corporuses as compared to the non conversation class (NC). We used a chi-test for feature ranking.

Evaluation: We used 10-fold Cross Validation to establish the unbiased accuracy of conversation classifiers and feature ranking models, where the system generates partitions of samples of data into complementary subsets, performing the analysis on 9 subsets (the training set), and validating the analysis on the testing set. This process repeats 10 times, generating average accuracy and other statistics.

Results
The following table summarizes accuracy and the ROC values for the five events and the three platform indicators of conversation. ROC values for the three natural disasters examining reply-based tweets average .78, retweet based tweets (.77) and mention based tweets (.66). The general pattern holds for the other two events, with generally lower ROC values.

<table>
<thead>
<tr>
<th>Event</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>ROC Value (AUC)</th>
<th>ROC Value (AUC)</th>
<th>ROC Value (AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan Tsunami</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
<td>.70</td>
<td>.77</td>
<td>.77</td>
<td>.77</td>
</tr>
<tr>
<td>Earthquake</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
<td>.70</td>
<td>.77</td>
<td>.77</td>
<td>.77</td>
</tr>
<tr>
<td>Hurricane</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
<td>.70</td>
<td>.77</td>
<td>.77</td>
<td>.77</td>
</tr>
<tr>
<td>Cyclone</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
<td>.70</td>
<td>.77</td>
<td>.77</td>
<td>.77</td>
</tr>
<tr>
<td>Earthquake</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
<td>.70</td>
<td>.77</td>
<td>.77</td>
<td>.77</td>
</tr>
</tbody>
</table>

Discussion
- The size of non-conversation corpus is larger than any of the conversation corpus for all the events. This illustrates the need for distinguishing true conversations for coordination purpose in the huge message traffic from an event.
- Classifier accuracy values up to 77% and ROC area values up to 0.82 support the diagnostic role of linguistic indicators in conversation relative to non-conversation in general as well as social media conversation.
- Classifier accuracy is maximal for distinguishing reply-based tweets from non-conversation. This meets our intuitions about the relative mix of true conversations and various platform indicators. However, we attribute imperfect classification in part to the imperfect basis for the initial classification of conversation using platform indicators. Not all tweets in the reply-based corpus are true conversation. Furthermore some of the tweets initially classified as non-conversations may also reflect true conversations. Our analysis supports the need for conversation identifiers that are independent of platform indicators.
- Performance of the classifier in the case of first three (disaster) data sets is better than remaining generic data sets. We believe that the features of coordination in conversation correlate with the coordinated actionable activity.
- We note that performance in the case of individual conversation models on Haiti earthquake data set is not as good as in the case of the Japan earthquake or Hurricane Irene storm data sets among disaster type events.
- The top 4 features vary for the different types of conversation corpus and consistently for all types of events and mixed data sets. Users may have developed linguistic patterns while using a particular type of conversation (Reply, RT, Mention).
- Personal Pronouns, Relative Pronouns and Dialogue indicators play a major role for discriminating conversation types from non conversations, especially for Reply based exchange. Action oriented conversations establish referents for the important objects in a task, for efficient future conversation.

This work is sponsored by NSF grant#NSF-11111182, titled as ‘SOCS: Social Media Enhanced Organizational Sensemaking in emergency response’.

This page includes an extended discussion of some of the results.