A Preliminary Study of Tweet Summarization using Information Extraction

Wei Xu, Ralph Grishman, Adam Meyers
Computer Science Department
New York University
New York, NY 10003, USA
{xuwei,grishman,meyers}@cs.nyu.edu

Alan Ritter
Computer Science and Engineering
University of Washington
Seattle, WA 98125, USA
aritter@cs.washington.edu

Abstract

Although the ideal length of summaries differs greatly from topic to topic on Twitter, previous work has only generated summaries of a pre-fixed length. In this paper, we propose an event-graph based method using information extraction techniques that is able to create summaries of variable length for different topics. In particular, we extend the Pagerank-like ranking algorithm from previous work to partition event graphs and thereby detect fine-grained aspects of the event to be summarized. Our preliminary results show that summaries created by our method are more concise and news-worthy than SumBasic according to human judges. We also provide a brief survey of datasets and evaluation design used in previous work to highlight the need of developing a standard evaluation for automatic tweet summarization task.

1 Introduction

Tweets contain a wide variety of useful information from many perspectives about important events taking place in the world. The huge number of messages, many containing irrelevant and redundant information, quickly leads to a situation of information overload. This motivates the need for automatic summarization systems which can select a few messages for presentation to a user which cover the most important information relating to the event without redundancy and filter out irrelevant and personal information that is not of interest beyond the user's immediate social network.

Although there is much recent work focusing on the task of multi-tweet summarization (Becker et al., 2011; Inouye and Kalita, 2011; Zubiaga et al., 2012; Liu et al., 2011a; Takamura et al., 2011; Harabagiu and Hickl, 2011; Wei et al., 2012), most previous work relies only on surface lexical clues, redundancy and social network specific signals (e.g. user relationship), and little work has considered taking limited advantage of information extraction techniques (Harabagiu and Hickl, 2011) in generative models. Because of the noise and redundancy in social media posts, the performance of off-the-shelf news-trained natural language process systems is degraded while simple term frequency is proven powerful for summarizing tweets (Inouye and Kalita, 2011). A natural and interesting research question is whether it is beneficial to extract named entities and events in the tweets as has been shown for classic multi-document summarization (Li et al., 2006). Recent progress on building NLP tools for Twitter (Ritter et al., 2011; Gimpel et al., 2011; Liu et al., 2011b; Ritter et al., 2012; Liu et al., 2012) makes it possible to investigate an approach to summarizing Twitter events which is based on Information Extraction techniques.

We investigate a graph-based approach which leverages named entities, event phrases and their connections across tweets. A similar idea has been studied by Li et al. (2006) to rank the salience of event concepts in summarizing news articles. However, the extreme redundancy and simplicity of tweets allows us to explicitly split the event graph into subcomponents that cover various aspects of the initial event to be summarized to create comprehen-

Proceedings of the Workshop on Language in Social Media (LASM 2013), pages 20–29,
Atlanta, Georgia, June 13 2013.

Work	Dataset (size of each cluster)	System Output	Evaluation Metrics
Inouye and Kalita (2011)	trending topics (approximately 1500 tweets)	4 tweets	ROUGE and human (overall quality comparing to human summary)
Sharifi et al. (2010)	same as above	1 tweet	same as above
Rosa et al. (2011)	segmented hashtag topics by LDA and k-means clustering (average 410 tweets)	1, 5, 10 tweets	Precision@k (relevance to topic)
Harabagiu and Hickl (2011)	real-word event topics (a minimum of 2500 tweets)	top tweets until a limit of 250 words was reached	human (coverage and coherence)
Liu et al. (2011a)	general topics and hashtag topics (average 1.7k tweets)	same lengths as of the human summary, vary for each topic (about 2 or 3 tweets)	ROUGE and human (content coverage, grammaticality, non-redundancy, referential clarity, focus)
Wei et al. (2012)	segmented hashtag topics according to burstiness (average 10k tweets)	10 tweets	ROUGE, Precison/Recall (good readability and rich content)
Takamura et al. (2011)	specific soccer games (2.8k - 5.2k tweets)	same lengths as the human summary, vary for each topic (26 - 41 tweets)	ROUGE (considering only content words)
Chakrabarti and Punera (2011)	specific football games (1.8k tweets)	10 - 70 tweets	Precision@k (relevance to topic)

Table 1: Summary of datasets and evaluation metrics used in several previous work on tweet summarization

sive and non-redundant summaries. Our work is the first to use a Pagerank-like algorithm for graph partitioning and ranking in the context of summarization, and the first to generate tweet summaries of variable length which is particularly important for tweet summarization. Unlike news articles, the amount of information in a set of topically clustered tweets varies greatly, from very repetitive to very discrete. For example, the tweets about one album release can be more or less paraphrases, while those about another album by a popular singer may involve rumors and release events etc. In the human study conducted by Inouye and Kalita (2011), annotators strongly prefer different numbers of tweets in a summary for different topics. However, most of the previous work produced summaries of a pre-fixed length and has no evaluation on conciseness. Liu et al. (2011a) and Takamura et al. (2011) also noticed the ideal length of summaries can be very different from topic to topic, and had to use the length of human reference summaries to decide the length of system outputs, information which is not available in practice. In contrast, we developed a system that is capable of detecting fine-grained sub-events and generating summaries with the proper number of representative tweets accordingly for different topics.

Our experimental results show that with information extraction it is possible to create more meaningful and concise summaries. Tweets that contain real-world events are usually more informative and readable. Event-based summarization is especially beneficial in this situation due to the fact that tweets are short and self-contained with simple discourse structure. The boundary of 140 characters makes it efficient to extract semi-structured events with shallow natural language processing techniques and re-

Tweets (Date Created)	Named Entity	Event Phrases	Date Mentioned
Nooooo.. Season premiere of Doctor Who is on Sept 1 world wide and we'll be at World Con (8/22/2012)	doctor who, world con	season, is on, premiere	sept 1 (9/1/2012)
guess what I DON'T get to do tomorrow! WATCH DOCTOR WHO (8/31/2012)	doctor who	watch	tomorrow (9/1/2012)
As I missed it on Saturday, I'm now catching up on Doctor Who (9/4/2012)	doctor who	missed, catching up	saturday (9/1/2012)
Rumour: Nokia could announce two WP8 devices on September 5 http://t.co/yZUwDFLV (via @mobigyaan)	nokia, wp8	announce	september 5 (9/5/2012)
Verizon and Motorola won't let Nokia have all the fun ; scheduling September 5th in New York http://t.co/qbBlYnSl (8/19/2012)	nokia, verizon, motorola, new york	scheduling	september 5th (9/5/2012)
Don't know if it's excitement or rooting for the underdog, but I am genuinely excited for Nokia come Sept 5: http://t.co/UhV5SUMP (8/7/2012)	nokia	rooting, excited	sept 5 (9/5/2012)

Table 2: Event-related information extracted from tweets

duces the complexity of the relationship (or no relationship) between events according to their co-occurrence, resulting in differences in constructing event graphs from previous work in news domain (Li et al., 2006).

2 Issues in Current Research on Tweet Summarization

The most serious problem in tweet summarization is that there is no standard dataset, and consequently no standard evaluation methodology. Although there are more than a dozen recent works on social media summarization, astonishingly, almost each research group used a different dataset and a different experiment setup. This is largely attributed to the difficulty of defining the right granularity of a topic in Twitter. In Table 1, we summarize the experiment designs of several selective works. Regardless of the differences, researchers generally agreed on :

- clustering tweets topically and temporally
- generating either a very short summary for a focused topic or a long summary for large-size clusters
- difficulty and necessity to generate summaries of variable length for different topics

Although the need of variable-length summaries have been raised in previous work, none has provide a good solution (Liu et al., 2011a; Takamura et al., 2011; Inouye and Kalita, 2011). In this paper, our focus is study the feasibility of generating concise summaries of variable length and improving meaningfulness by using information extraction techniques. We hope this study can provide new insights on the task and help in developing a standard evaluation in the future.

3 Approach

We first extract event information including named entities and event phrases from tweets and construct event graphs that represent the relationship between them. We then rank and partition the events using PageRank-like algorithms, and create summaries of variable length for different topics.

3.1 Event Extraction from Tweets

As a first step towards summarizing popular events discussed on Twitter, we need a way to identify events from Tweets. We utilize several natural language processing tools that specially developed for noisy text to extract text phrases that bear essential event information, including named entities (Ritter et al., 2011), event-referring phrases (Ritter et al.,

2012) and temporal expressions (Mani and Wilson, 2000). Both the named entity and event taggers utilize Conditional Random Fields models (Lafferty, 2001) trained on annotated data, while the temporal expression resolver uses a mix of hand-crafted and machine-learned rules. Example event information extracted from Tweets are presented in Table 2.

The self-contained nature of tweets allows efficient extraction of event information without deep analysis (e.g. co-reference resolution). On the other hand, individual tweets are also very terse, often lacking sufficient context to access the importance of events. It is crucial to exploit the highly redundancy in Twitter. Closely following previous work by Ritter et al. (2012), we group together sets of topically and temporally related tweets, which mention the same named entity and a temporal reference resolved to the same unique calendar date. We also employ a statistical significance test to measure strength of association between each named entity and date, and thereby identify important events discussed widely among users with a specific focus, such as the release of a new iPhone as opposed to individual users discussing everyday events involving their phones. By discarding frequent but insignificant events, we can produce more meaningful summaries about popular real-world events.

3.2 Event Graphs

Since tweets have simple discourse and are self-contained, it is a reasonable assumption that named entities and event phrases that co-occurred together in a single tweet are very likely related. Given a collection of tweets, we represent such connections by a weighted undirected graph :

- Nodes: named entities and event phrases are represented by nodes and treated indifferently.
- Edges: two nodes are connected by an undirected edge if they co-occurred in k tweets, and the weight of edge is k.

We find it helpful to merge named entities and event phrases that have lexical overlap if they are frequent but not the topic of the tweet cluster. For example, 'bbc', 'radio 1', 'bbc radio 1' are combined together in a set of tweets about a band. Figure 1 shows a very small toy example of event graph. In the experiments of this paper, we also exclude the edges with $k < 2$ to reduce noise in the data and calculation cost.

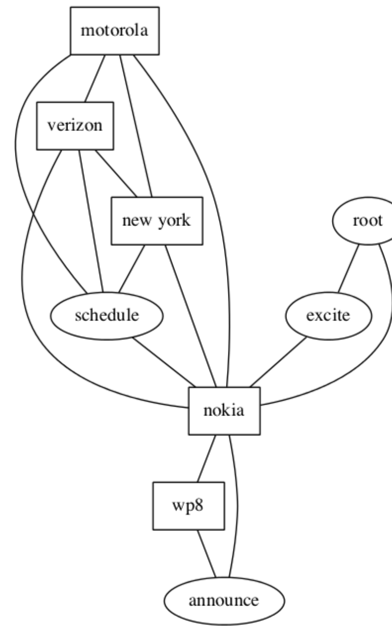

Figure 1: A toy event graph example built from the three sentences of the event 'Nokia - 9/5/2012' in Table 2

3.3 Event Ranking and Partitioning

Graph-based ranking algorithms are widely used in automatic summarization to decide salience of concepts or sentences based on global information recursively drawn from the entire graph. We adapt the PageRank-like algorithm used in TextRank (Mihalcea and Tarau, 2004) that takes into account edge weights when computing the score associated with a vertex in the graph.

Formally, let $G = (V, E)$ be a undirected graph with the set of vertices V and set of edges E, where E is a subset of $V \times V$. For a given vertex V_i, let $Ad(V_i)$ be the set of vertices that adjacent to it. The weight of the edge between V_i and V_j is denoted as w_{ij}, and $w_{ij} = w_{ji}$. The score of a vertex V_i is defined as follows:

$$S(V_i) = (1 - d) + d \times \sum_{V_j \in Ad(V_i)} \frac{w_{ij} \times S(V_j)}{\sum_{V_k \in Ad(V_j)} w_{jk}}$$

where d is a damping factor that is usually set to 0.85 (Brin and Page, 1998), and this is the value we are also using in our implementation.

Starting from arbitrary values assigned to each node in the graph, the computation iterates until convergence. Note that the final salience score of each node is not affected by the choice of the initial values assigned to each node in the graph, but rather the weights of edges.

In previous work computed scores are then used directly to select text fractions for summaries (Li et al., 2006). However, the redundancy and simplicity of tweets allow further exploration into sub-event detection by graph partitioning. The intuition is that the correlations between named entities and event phrases within same sub-events are much stronger than between sub-events. This phenomena is more obvious and clear in tweet than in news articles, where events are more diverse and complicated related to each other given lengthy context.

As theoretically studied in local partitioning problem (Andersen et al., 2006), a good partition of the graph can be obtained by separating high ranked vertices from low ranked vertices, if the nodes in the graph have ranks that are distinguishable. Utilizing a similar idea, we show that a simple greedy algorithm is efficient to find important sub-events and generate useful summaries in our tasks. As shown in Figure 2 and 3, the high ranked nodes (whose scores are greater than 1, the average score of all nodes in the graph) in tweet event graphs show the divisions within a topic. We search for strongly connected sub-graphs, as gauged by parameter α, from the highest ranked node to lower ranked ones.The proportion of tweets in a set that are related to a sub-event is then estimated according to the ratio between the sum of node scores in the sub-graph versus the entire graph. We select one tweet for each sub-event that best covers the related nodes with the highest sum of node scores normalized by length as summaries. By adding a cutoff (parameter β) on proportion of sub-event required to be included into summaries, we can produce summaries with the appropriate length according to the diversity of information in a set of tweets.

In Figure 2, 3 and 4, the named entity which is also the topic of tweet cluster is omitted since it is connected with every node in the event graph. The size of node represents the salience score, while the shorter, straighter and more vertical the edge is, the higher its weight. The nodes with rectangle shapes are named entities, while round shaped ones are event phrases. Note that in most cases, sub-events correspond to connected components in the event graph of high ranked nodes as in Figure 2 and 3. However, our simple greedy algorithm also allows multiple sub-events for a single connected component that can not be covered by one tweet in the summary. For example, in Figure 4, two sub-events $e_1 = \{sell, delete, start, payment\}$ and $e_2 = \{facebook, share\ user\ data, privacy\ policy, debut\}$ are chosen to accommodate the complex event.

Algorithm 1 Find important sub-events

Require: Ranked event graph $G = (V, E)$, the named entity V_0 which is the topic of event cluster, parameters α and β that can be set towards user preference over development data

Initialize the pool of high ranked nodes $\tilde{V} \leftarrow \{V_i | \forall V_i \in V, S(V_i) > 1\} - V_0$ and the total weight $W \leftarrow \sum_{V_i \in \tilde{V}} S(V_i)$
while $\tilde{V} \neq \emptyset$ **do**
  Pop the highest ranked node V_m from $\tilde{V}$
  Put V_m to a temporary sub-event $e \leftarrow \{V_m\}$
  for all V_n in $\tilde{V}$ **do**
    if $w_{mn}/w_{0m} > \alpha$ and $w_{0n}/w_{0m} > \alpha$ **then**
      $e \leftarrow e \cup \{V_n\}$
    end if
  end for
  $W_e \leftarrow \sum_{V_i \in e} S(V_i)$
  if $W_e/W > \beta$ **then**
    Successfully find a sub-event e
    Remove all nodes in e from $\tilde{V}$
  end if
end while

4 Experiments

4.1 Data

We gathered tweets over a 4-month period spanning November 2012 to February 2013 using the Twitter Streaming API. As described in more details in previous work on Twitter event extraction by Ritter et al. (2012), we grouped together all tweets which mention the same named entity (recognized using

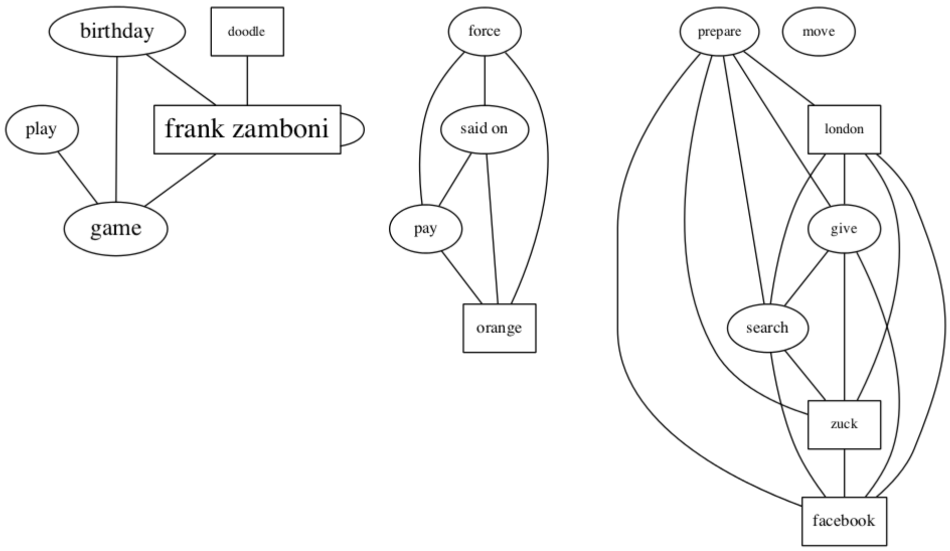

Figure 2: Event graph of ’Google - 1/16/2013’, an example of event cluster with multiple focuses

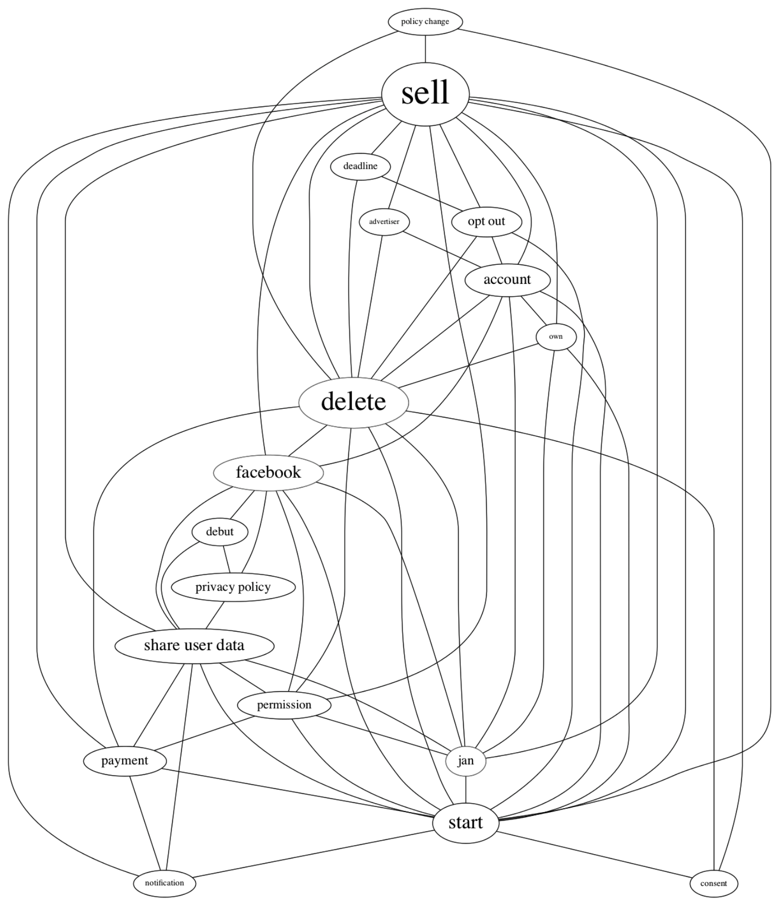

Figure 3: Event graph of ’Instagram - 1/16/2013’, an example of event cluster with a single but complex focus

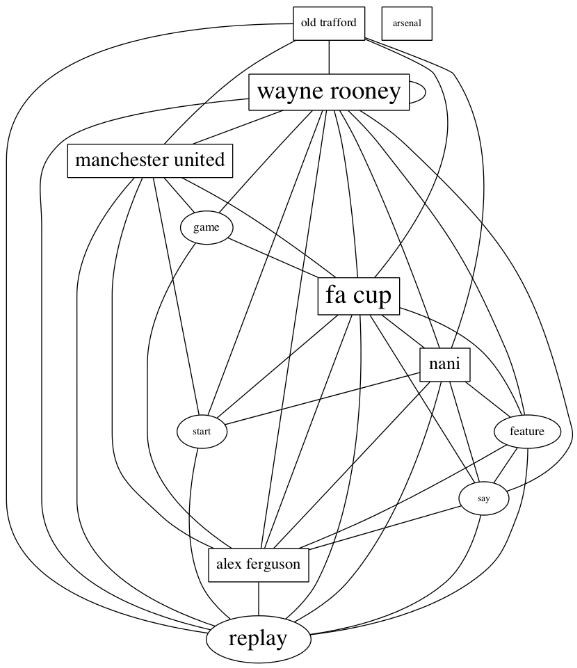

Figure 4: Event graph of 'West Ham - 1/16/2013', an example of event cluster with a single focus

a Twitter specific name entity tagger[1]) and a reference to the same unique calendar date (resolved using a temporal expression processor (Mani and Wilson, 2000)). Tweets published during the whole period are aggregated together to find top events that happen on each calendar day. We applied the G^2 test for statistical significance (Dunning, 1993) to rank the event clusters, considering the corpus frequency of the named entity, the number of times the date has been mentioned, and the number of tweets which mention both together. We randomly picked the events of one day for human evaluation, that is the day of January 16, 2013 with 38 events and an average of 465 tweets per event cluster.

For each cluster, our systems produce two versions of summaries, one with a fixed number (set to 3) of tweets and another one with a flexible number (vary from 1 to 4) of tweets. Both α and β are set to 0.1 in our implementation. All parameters are set experimentally over a small development dataset consisting of 10 events in Twitter data of September 2012.

[1]https://github.com/aritter/twitter_nlp

4.2 Baseline

SumBasic (Vanderwende et al., 2007) is a simple and effective summarization approach based on term frequency, which we use as our baseline. It uses word probabilities with an update function to avoid redundancy to select sentences or posts in a social media setting. It is shown to outperform three other well-known multi-document summarization methods, namely LexRank (Erkan and Radev, 2004), TextRank (Mihalcea and Tarau, 2004) and MEAD (Radev et al., 2004) on tweets in (Inouye and Kalita, 2011), possibly because that the relationship between tweets is much simpler than between sentences in news articles and can be well captured by simple frequency methods. The improvement over the LexRank model on tweets is gained by considering the number of retweets and influential users is another side-proof (Wei et al., 2012) of the effectiveness of frequency.

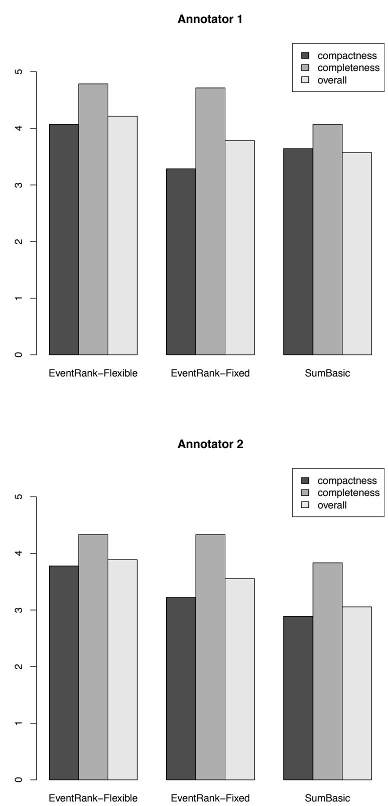

Figure 5: human judgments evaluating tweet summarization systems

Event	System	Summary
Google 1/16/2013	EventRank (Flexible)	- Google 's home page is a Zamboni game in celebration of Frank Zamboni 's birthday January 16 #GameOn - Today social , Tomorrow Google ! Facebook Has Publicly Redefined Itself As A Search Company http://t.co/dAevB2V0 via @sai - Orange says has it has forced Google to pay for traffic . The Head of the Orange said on Wednesday it had ... http://t.co/dOqAHhWi
	SumBasic	- Tomorrow's Google doodle is going to be a Zamboni! I may have to take a vacation day. - the game on google today reminds me of hockey #tooexcited #saturday - The fact that I was soooo involved in that google doodle game says something about this Wednesday #TGIW You should try it!
Instagram 1/16/2013	EventRank (Flexible)	- So Instagram can sell your pictures to advertisers without u knowing starting January 16th I'm bout to delete my instagram ! - Instagram debuts new privacy policy , set to share user data with Facebook beginning January 16
	SumBasic	- Instagram will have the rights to sell your photos to Advertisers as of jan 16 - Over for Instagram on January 16th - Instagram says it now has the right to sell your photos unless you delete your account by January 16th http://t.co/tsjic6yA
West Ham 1/16/2013	EventRank (Flexible)	- RT @Bassa_Mufc : Wayne Rooney and Nani will feature in the FA Cup replay with West Ham on Wednesday - Sir Alex Ferguson
	SumBasic	- Wayne Rooney could be back to face West Ham in next Wednesday's FA Cup replay at Old Trafford. #BPL - Tomorrow night come on West Ham lol - Nani's fit abd WILL play tomorrow against West Ham! Sir Alex confirmed :)

Table 3: Event-related information extracted from tweets

4.3 Preliminary Results

We performed a human evaluation in which two annotators were asked to rate the system on a five-point scale (1=very poor, 5=very good) for completeness and compactness. Completeness refers to how well the summary cover the important content in the tweets. Compactness refers to how much meaningful and non-redundant information is in the summary. Because the tweets were collected according to information extraction results and ranked by salience, the readability of summaries generated by different systems are generally very good. The top 38 events of January 16, 2013 are used as test set. The aggregate results of the human evaluation are displayed in Figure 5. Agreement between annotators measured using Pearson's Correlation Coefficient is 0.59, 0.62, 0.62 respectively for compactness, completeness and overall judgements.

Results suggest that the models described in this paper produce more satisfactory results as the baseline approaches. The improvement of EventRank-Flexible over SumBasic is significant (two-tailed $p < 0.05$) for all three metrics according to student's t test. Example summaries of the events in Figure 2, 3 and 4 are presented respectively in Table 3. The advantages of our method are the following: 1) it finds important facts of real-world events 2) it prefers tweets with good readability 3) it includes the right amount of information with diversity and without redundancy. For example, our system picked only one tweet about 'West Ham -1/16/2013' that convey the same message as the three tweets together of the baseline system. For another example, among the tweets about Google around 1/16/2013, users intensively talk about the Google doodle game with a very wide range of words creatively, giving word-based methods a hard time to pick up the diverse and essential event information that is less frequent.

5 Conclusions and Future Work

We present an initial study of feasibility to generate compact summaries of variable lengths for tweet summarization by extending a Pagerank-like algorithm to partition event graphs. The evaluation shows that information extraction techniques are helpful to generate news-worthy summaries of good readability from tweets.

In the future, we are interested in improving the approach and evaluation, studying automatic metrics to evaluate summarization of variable length and getting involved in developing a standard evaluation for tweet summarization tasks. We wonder whether other graph partitioning algorithms may improve the performance. We also consider extending this graph-based approach to disambiguate named entities or resolve event coreference in Twitter data. Another direction of future work is to extend the proposed approach to different data, for example, temporal-aware clustered tweets etc.

Acknowledgments

This research was supported in part by NSF grant IIS-0803481, ONR grant N00014-08-1-0431, and DARPA contract FA8750- 09-C-0179, and carried out at the University of Washington's Turing Center.

We thank Mausam and Oren Etzioni of University of Washington, Maria Pershina of New York University for their advice.

References

Reid Andersen, Fan Chung, and Kevin Lang. 2006. Local graph partitioning using pagerank vectors. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 475–486. IEEE.

Hila Becker, Mor Naaman, and Luis Gravano. 2011. Selecting quality twitter content for events. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11)*.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117.

Deepayan Chakrabarti and Kunal Punera. 2011. Event summarization using tweets. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 66–73.

Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1):61–74.

Günes Erkan and Dragomir R. Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res. (JAIR)*, 22:457–479.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *ACL*.

Sanda Harabagiu and Andrew Hickl. 2011. Relevance modeling for microblog summarization. In *Fifth International AAAI Conference on Weblogs and Social Media*.

David Inouye and Jugal K Kalita. 2011. Comparing twitter summarization algorithms for multiple post summaries. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 298–306. IEEE.

John Lafferty. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann.

Wenjie Li, Wei Xu, Chunfa Yuan, Mingli Wu, and Qin Lu. 2006. Extractive summarization using inter- and intra- event relevance. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, ACL-44, pages 369–376, Stroudsburg, PA, USA. Association for Computational Linguistics.

Fei Liu, Yang Liu, and Fuliang Weng. 2011a. Why is ÒsxswÓ trending? exploring multiple text sources for twitter topic summarization. *ACL HLT 2011*, page 66.

Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. 2011b. Recognizing named entities in tweets. In *ACL*.

Xiaohua Liu, Furu Wei, Ming Zhou, et al. 2012. Quickview: Nlp-based tweet search. In *Proceedings of the ACL 2012 System Demonstrations*, pages 13–18. Association for Computational Linguistics.

Inderjeet Mani and George Wilson. 2000. Robust temporal processing of news. In *Proceedings of the 38th An-*

nual Meeting on Association for Computational Linguistics, ACL '00, pages 69–76, Stroudsburg, PA, USA. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In *Proceedings of EMNLP*, volume 4, pages 404–411. Barcelona, Spain.

Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, et al. 2004. Mead-a platform for multidocument multilingual text summarization. In *Proceedings of LREC*, volume 2004.

Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named entity recognition in tweets: An experimental study.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *KDD*, pages 1104–1112. ACM.

Kevin Dela Rosa, Rushin Shah, Bo Lin, Anatole Gershman, and Robert Frederking. 2011. Topical clustering of tweets. *Proceedings of the ACM SIGIR: SWSM*.

Beaux Sharifi, Mark-Anthony Hutton, and Jugal K Kalita. 2010. Experiments in microblog summarization. In *Proc. of IEEE Second International Conference on Social Computing*.

Hiroya Takamura, Hikaru Yokono, and Manabu Okumura. 2011. Summarizing a document stream. *Advances in Information Retrieval*, pages 177–188.

Lucy Vanderwende, Hisami Suzuki, Chris Brockett, and Ani Nenkova. 2007. Beyond sumbasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing & Management*, 43(6):1606–1618.

Furu Wei, Ming Zhou, and Heung-Yeung Shum. 2012. Twitter topic summarization by ranking tweets using social influence and content quality. In *COLING*.

Arkaitz Zubiaga, Damiano Spina, Enrique Amigó, and Julio Gonzalo. 2012. Towards real-time summarization of scheduled events from twitter streams. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, pages 319–320. ACM.