Automatic Labelling of Topic Models Learned from Twitter by Summarisation

Amparo Elizabeth Cano Basave[†]   **Yulan He**[‡]   **Ruifeng Xu**[§]

[†] Knowledge Media Institute, Open University, UK

[‡] School of Engineering and Applied Science, Aston University, UK

[§] Key Laboratory of Network Oriented Intelligent Computation

Shenzhen Graduate School, Harbin Institute of Technology, China

`amparo.cano@open.ac.uk`, `y.he@cantab.net`, `xuruifeng@hitsz.edu.cn`

Abstract

Latent topics derived by topic models such as Latent Dirichlet Allocation (LDA) are the result of hidden thematic structures which provide further insights into the data. The automatic labelling of such topics derived from social media poses however new challenges since topics may characterise novel events happening in the real world. Existing automatic topic labelling approaches which depend on external knowledge sources become less applicable here since relevant articles/concepts of the extracted topics may not exist in external sources. In this paper we propose to address the problem of automatic labelling of latent topics learned from Twitter as a summarisation problem. We introduce a framework which apply summarisation algorithms to generate topic labels. These algorithms are independent of external sources and only rely on the identification of dominant terms in documents related to the latent topic. We compare the efficiency of existing state of the art summarisation algorithms. Our results suggest that summarisation algorithms generate better topic labels which capture event-related context compared to the top-n terms returned by LDA.

1   Introduction

Topic model based algorithms applied to social media data have become a mainstream technique in performing various tasks including sentiment analysis (He, 2012) and event detection (Zhao et al., 2012; Diao et al., 2012). However, one of the main challenges is the task of understanding the semantics of a topic. This task has been approached by investigating methodologies for identifying meaningful topics through semantic coher-

ence (Aletras and Stevenson, 2013; Mimno et al., 2011; Newman et al., 2010) and for characterising the semantic content of a topic through automatic labelling techniques (Hulpus et al., 2013; Lau et al., 2011; Mei et al., 2007). In this paper we focus on the latter.

Our research task of automatic labelling a topic consists on selecting a set of words that best describes the semantics of the terms involved in this topic. The most generic approach to automatic labelling has been to use as primitive labels the top-n words in a topic distribution learned by a topic model such as LDA (Griffiths and Steyvers, 2004; Blei et al., 2003). Such top words are usually ranked using the marginal probabilities $P(w_i|t_j)$ associated with each word w_i for a given topic t_j. This task can be illustrated by considering the following topic derived from social media related to Education:

> school protest student fee choic motherlod tuition teacher anger polic

where the top 10 words ranked by $P(w_i|t_j)$ for this topic are listed. Therefore the task is to find the top-n terms which are more representative of the given topic. In this example, the topic certainly relates to a student protest as revealed by the top 3 terms which can be used as a good label for this topic.

However previous work has shown that top terms are not enough for interpreting the coherent meaning of a topic (Mei et al., 2007). More recent approaches have explored the use of external sources (e.g. Wikipedia, WordNet) for supporting the automatic labelling of topics by deriving candidate labels by means of lexical (Lau et al., 2011; Magatti et al., 2009; Mei et al., 2007) or graph-based (Hulpus et al., 2013) algorithms applied on these sources.

Mei et al. (2007) proposed an unsupervised probabilistic methodology to automatically assign a label to a topic model. Their proposed approach

was defined as an optimisation problem involving the minimisation of the KL divergence between a given topic and the candidate labels while maximising the mutual information between these two word distributions. Lau et al. (2010) proposed to label topics by selecting top-n terms to label the overall topic based on different ranking mechanisms including pointwise mutual information and conditional probabilities.

Methods relying on external sources for automatic labelling of topics include the work by Magatti et al. (2009) which derived candidate topic labels for topics induced by LDA using the hierarchy obtained from the Google Directory service and expanded through the use of the OpenOffice English Thesaurus. Lau et al. (2011) generated label candidates for a topic based on top-ranking topic terms and titles of Wikipedia articles. They then built a Support Vector Regression (SVR) model for ranking the label candidates. More recently, Hulpus et al. (2013) proposed to make use of a structured data source (DBpedia) and employed graph centrality measures to generate semantic concept labels which can characterise the content of a topic.

Most previous topic labelling approaches focus on topics derived from well formatted and static documents. However in contrast to this type of content, the labelling of topics derived from tweets presents different challenges. In nature micropost content is sparse and present ill-formed words. Moreover, the use of Twitter as the "what's-happening-right now" tool, introduces new event-dependent relations between words which might not have a counter part in existing knowledge sources (e.g. Wikipedia). Our original interest in labelling topics stems from work in topic model based event extraction from social media, in particular from tweets (Shen et al., 2013; Diao et al., 2012). As opposed to previous approaches, the research presented in this paper addresses the labelling of topics exposing event-related content that might not have a counter part on existing external sources. Based on the observation that a short summary of a collection of documents can serve as a label characterising the collection, we propose to generate topic label candidates based on the summarisation of a topic's relevant documents. Our contributions are two-fold:

- We propose a novel approach for topics labelling that relies on term relevance of documents relating to a topic; and

- We show that summarisation algorithms, which are independent of extenal sources, can be used with success to label topics, presenting a higher perfomance than the top-n terms baseline.

2 Methodology

We propose to approach the topic labelling problem as a multi-document summarisation task. The following describes our proposed framework to characterise documents relevant to a topic.

2.1 Preliminaries

Given a set of documents the problem to be solved by topic modelling is the posterior inference of the variables, which determine the hidden thematic structures that best explain an observed set of documents. Focusing on the Latent Dirichlet Allocation (LDA) model (Blei et al., 2003; Griffiths and Steyvers, 2004), let $\mathcal{D}$ be a corpus of documents denoted as $\mathcal{D} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, .., \boldsymbol{d}_D\}$; where each document consists of a sequence of N_d words denoted by $\boldsymbol{d} = (w_1, w_2, .., w_{N_d})$; and each word in a document is an item from a vocabulary index of V different terms denoted by $\{1, 2, .., V\}$. Given D documents containing K topics expressed over V unique words, LDA generative process is described as follows:
- For each topic $k \in \{1, ...K\}$ draw $\phi_k \sim$ Dirichlet(β),
- For each document $d \in \{1..D\}$:
    ⋆ draw $\theta_d \sim$ Dirichlet(α);
    ⋆ For each word $n \in \{1..N_d\}$ in document d:
      ○ draw a topic $z_{d,n} \sim$ Multinomial(θ_d);
      ○ draw a word $w_{d,n} \sim$ Multinomial($\varphi_{z_{d,n}}$).

where ϕ_k is the word distribution for topic k, and θ_d is the distribution of topics in document d. Topics are interpreted using the top N terms ranked based on the marginal probability $p(w_i|t_j)$.

2.2 Automatic Labelling of Topic Models

Given K topics over the document collection $\mathcal{D}$, the topic labelling task consists on discovering a sequence of words for each topic $k \in \mathcal{K}$. We propose to generate topic label candidates by summarising topic relevant documents. Such documents can be derived using both the observed data from the corpus $\mathcal{D}$ and the inferred topic model variables. In particular, the prominent topic of a document d can be found by

$$k_d = \underset{k \in \mathcal{K}}{\arg\max}\, p(k|d) \qquad (1)$$

Therefore given a topic k, a set of C documents related to this topic can be obtained via equation 1.

Given the set of documents $\mathcal{C}$ relevant to topic k, we proposed to generate a label of a desired length x from the summarisation of $\mathcal{C}$.

2.3 Topic Labelling by Summarisation

We compare different summarisation algorithms based on their ability to provide a good label to a given topic. In particular we investigate the use of lexical features by comparing three different well-known multi-document summarisation algorithms against the top-n topic terms baseline. These algorithms include:

Sum Basic (`SB`) This is a frequency based summarisation algorithm (Nenkova and Vanderwende, 2005), which computes initial word probabilities for words in a text. It then weights each sentence in the text (in our case a micropost) by computing the average probability of the words in the sentence. In each iteration it picks the highest weighted document and from it the highest weighted word. It uses an update function which penalises words which have already been picked.

Hybrid TFIDF (`TFIDF`) It is similar to `SB`, however rather than computing the initial word probabilities based on word frequencies it weights terms based on TFIDF. In this case the document frequency is computed as the number of times a word appears in a micropost from the collection $\mathcal{C}$. Following the same procedure as `SB` it returns the top x weighted terms.

Maximal Marginal Relevance (`MMR`) This is a relevance based ranking algorithm (Carbonell and Goldstein, 1998), which avoids redundancy in the documents used for generating a summary. It measures the degree of dissimilarity between the documents considered and previously selected ones already in the ranked list.

Text Rank (`TR`) This is a graph-based summariser method (Mihalcea and Tarau, 2004) where each word is a vertex. The relevance of a vertex (term) to the graph is computed based on global information recursively drawn from the whole graph. It uses the PageRank algorithm (Brin and Page, 1998) to recursively change the weight of the vertices. The final score of a word is therefore not only dependent on the terms immediately connected to it but also on how these terms connect to others. To assign the weight of an edge between two terms, TextRank computes word co-occurrence in windows of N words (in our case $N = 10$). Once a final score is calculated for each vertex of the graph, TextRank sorts the terms in a reverse order and provided the top T vertices in the ranking. Each of these algorithms produces a label of a desired length x for a given topic k.

3 Experimental Setup

3.1 Dataset

Our Twitter Corpus (`TW`) was collected between November 2010 and January 2011. `TW` comprises over 1 million tweets. We used the OpenCalais' document categorisation service[1] to generate categorical sets. In particular, we considered four different categories which contain many real-world events, namely: War and Conflict (`War`), Disaster and Accident (`DisAc`), Education (`Edu`) and Law and Crime (`LawCri`). The final `TW` dataset after removing retweets and short microposts (less than 5 words after removing stopwords) contains 7000 tweets in each category.

We preprocessed `TW` by first removing: punctuation, numbers, non-alphabet characters, stop words, user mentions, and URL links. We then performed Porter stemming (Porter, 1980) in order to reduce the vocabulary size. Finally to address the issue of data sparseness in the `TW` dataset, we removed words with a frequency lower than 5.

3.2 Generating the Gold Standard

Evaluation of automatic topic labelling often relied on human assessment which requires heavy manual effort (Lau et al., 2011; Hulpus et al., 2013). However performing human evaluations of Social Media test sets comprising thousands of inputs become a difficult task. This is due to both the corpus size, the diversity of event-related topics and the limited availability of domain experts. To alleviate this issue here, we followed the distribution similarity approach, which has been widely applied in the automatic generation of gold standards (`GSs`) for summary evaluations (Donaway et al., 2000; Lin et al., 2006; Louis and Nenkova, 2009; Louis and Nenkova, 2013). This approach compares two corpora, one for which no `GS` labels exist, against a reference corpus for which a `GS` exists. In our case these corpora correspond to the `TW` and a Newswire dataset (`NW`). Since previous

---

[1]OpenCalais service, http://www.opencalais.com

research has shown that headlines are good indicators of the main focus of a text, both in structure and content, and that they can act as a human produced abstract (Nenkova, 2005), we used headlines as the GS labels of `NW`.

The News Corpus (`NW`) was collected during the same period of time as the `TW` corpus. `NW` consists of a collection of news articles crawled from traditional news media (BBC, CNN, and New York Times) comprising over 77,000 articles which include supplemental metadata (e.g. headline, author, publishing date). We also used the OpenCalais' document categorisation service to automatically label news articles and considered the same four topical categories, (`War`, `DisAc`, `Edu` and `LawCri`). The same preprocessing steps were performed on `NW`.

Therefore, following a similarity alignment approach we performed the steps oulined in Algorithm 1 for generating the `GS` topic labels of a topic in `TW`.

Algorithm 1 `GS` for Topic Labels
***

Input: LDA topics for `TW`, and the LDA topics for `NW` for category c.

Output: Gold standard topic label for each of the LDA topics for `TW`.

1: **for** each topic $i \in \{1, 2, ..., 100\}$ from `TW` **do**
2:     **for** each topic $j \in \{1, 2..., 100\}$ from `NW` **do**
3:         Compute the Cosine similarity between word distributions of topic t_i and topic t_j.
4:     **end for**
5:     Select topic j which has the highest similarity to i and whose similarity measure is greater than a threshold (in this case 0.7)
6: **end for**
7: **for** each of the extracted topic pairs $(t_i - t_j)$ **do**
8:     Collect relevant news articles $\mathcal{C}_{NW}^{j}$ of topic t_j from the `NW` set.
9:     Extract the headlines of news articles from $\mathcal{C}_{NW}^{j}$ and select the top x most frequent words as the gold standard label for topic t_i in the `TW` set
10: **end for**

These steps can be outlined as follows:1) We ran LDA on `TW` and `NW` separately for each category with the number of topics set to 100; 2) We then aligned the Twitter topics and Newswire topics by the similarity measurement of word distributions of these topics (Ercan and Cicekli, 2008; Haghighi and Vanderwende, 2009; Wang et al., 2009; Delort and Alfonseca, 2012); 3) Finally to generate the GS label for each aligned topic pair $(t_i - t_j)$, we extracted the headlines of the news articles relevant to t_j and selected the top x most frequent words (after stop word removal and stemming). The generated label was used as the gold

standard label for the corresponding Twitter topic t_i in the topic pair.

4 Experimental Results

We compared the results of the summarisation techniques with the top terms (`TT`) of a topic as our baseline. These `TT` set corresponds to the top x terms ranked based on the probability of the word given the topic ($p(w|k)$) from the topic model. We evaluated these summarisation approaches with the ROUGE-1 method (Lin, 2004), a widely used summarisation evaluation metric that correlates well with human evaluation (Liu and Liu, 2008). This method measures the overlap of words between the generated summary and a reference, in our case the `GS` generated from the `NW` dataset.

The evaluation was performed at $x = \{1, .., 10\}$. Figure 1 presents the ROUGE-1 performance of the summarisation approaches as the length x of the generated topic label increases. We can see in all four categories that the `SB` and `TFIDF` approaches provide a better summarisation coverage as the length of the topic label increases. In particular, in both the `Education` and `Law & Crime` categories, both `SB` and `TFIDF` outperforms `TT` and `TR` by a large margin. The obtained ROUGE-1 performance is within the same range of performance previously reported on Social Media summarisation (Inouye and Kalita, 2011; Nichols et al., 2012; Ren et al., 2013).

Table 1 presents average results for ROUGE-1 in the four categories. Particularly the `SB` and `TFIDF` summarisation techniques consistently outperform the `TT` baseline across all four categories. `SB` gives the best results in three categories except `War`.

	ROUGE-1				
	TT	SB	TFIDF	MMR	TR
War	0.162	**0.184**	**0.192**	**0.154**	**0.141**
DisAc	0.134	**0.194**	**0.160**	0.132	0.124
Edu	0.106	**0.240**	**0.187**	**0.104**	0.023
LawCri	0.035	**0.159**	**0.149**	0.034	**0.115**

Table 1: Average ROUGE-1 for topic labels at $x = \{1..10\}$, generated from the `TW` dataset.

The generated labels with summarisation at $x = 5$ are presented in Table 2, where `GS` represents the label generated from the Newswire headlines.

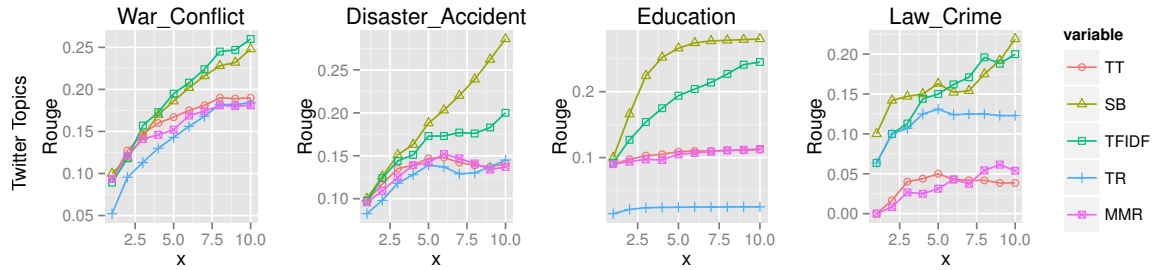Different summarisation techniques reveal words which do not appear in the top terms but

Figure 1: Performance in ROUGE for Twitter-derived topic labels, where x is the number of terms in the generated label

which are relevant to the information clustered by the topic. In this way, the labels generated for topics belonging to different categories generally extend the information provided by the top terms. For example in Table 2, the `DisAc` headline is characteristic of the New Zealand's Pike River's coal mine blast accident, which is an event occurred in November 2010.

Although the top 5 terms set from the LDA topic extracted from `TW` (listed under `TT`) does capture relevant information related to the event, it does not provide information regarding the blast. In this sense the topic label generated by `SB` more accurately describes this event.

We can also notice that the `GS` labels generated from Newswire media presented in Table 2 appear on their own, to be good labels for the `TW` topics. However as we described in the introduction we want to avoid relaying on external sources for the derivation of topic labels.

This experiment shows that frequency based summarisation techniques outperform graph-based and relevance based summarisation techniques for generating topic labels that improve upon the top-terms baseline, without relying on external sources. This is an attractive property for automatically generating topic labels for tweets where their event-related content might not have a counter part on existing external sources.

5   Conclusions and Future Work

In this paper we proposed a novel alternative to topic labelling which do not rely on external data sources. To the best of out knowledge no existing work has been formally studied for automatic labelling through summarisation. This experiment shows that existing summarisation techniques can be exploited to provide a better label of a topic, extending in this way a topic's information by pro-

	War	DisAc
GS	protest brief polic afghanistan attack world leader bomb obama pakistan	mine zealand rescu miner coal fire blast kill man dis-ast
TT	polic offic milit recent mosqu	mine coal pike river zealand
SB	terror war polic arrest offic	mine coal explos river pike
TFIDF	polic war arrest offic terror	mine coal pike safeti zealand
MMR	recent milit arrest attack target	trap zealand coal mine ex-plos
TR	war world peac terror hope	mine zealand plan fire fda

	Edu	LawCri
GS	school protest student fee choic motherlod tuition teacher anger polic	man charg murder arrest polic brief woman attack inquiri found
TT	student univers protest oc-cupi plan	man law child deal jail
SB	student univers school protest educ	man arrest law kill judg
TFIDF	student univers protest plan colleg	man arrest law judg kill
MMR	nation colleg protest stu-dent occupi	found kid wife student jail
TR	student tuition fee group hit	man law child deal jail

Table 2: Labelling examples for topics generated from the `TW` Dataset. GS represents the gold-standard generated from the relevant Newswire dataset. All terms are Porter stemmed as described in subsection 3.1

viding a richer context than top-terms. These results show that there is room to further improve upon existing summarisation techniques to cater for generating candidate labels.

References

Nikolaos Aletras and Mark Stevenson. 2013. Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 13–22, Potsdam, Germany, March. Association for Computational Linguistics.

David Meir Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. In *J. Mach. Learn. Res. 3*, pages 993–1022.

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

Jaime Carbonell and Jade Goldstein. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 335–336, New York, NY, USA. ACM.

Jean-Yves Delort and Enrique Alfonseca. 2012. Dualsum: A topic-model based approach for update summarization. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 214–223, Stroudsburg, PA, USA. Association for Computational Linguistics.

Qiming Diao, Jing Jiang, Feida Zhu, and Ee-Peng Lim. 2012. Finding bursty topics from microblogs. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 536–544, Jeju Island, Korea, July. Association for Computational Linguistics.

Robert L. Donaway, Kevin W. Drummey, and Laura A. Mather. 2000. A comparison of rankings produced by summarization evaluation measures. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization*, NAACL-ANLP-AutoSum '00, pages 69–78, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gonenc Ercan and Ilyas Cicekli. 2008. Lexical cohesion based topic modeling for summarization. In *Proceedings of the 9th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing'08, pages 582–592, Berlin, Heidelberg. Springer-Verlag.

Thomas L. Griffiths and Mark Steyvers. 2004. Finding scientific topics. *PNAS*, 101(suppl. 1):5228–5235.

Aria Haghighi and Lucy Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 362–370, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yulan He. 2012. Incorporating sentiment prior knowledge for weakly supervised sentiment analysis. *ACM Transactions on Asian Language Information Processing*, 11(2):4:1–4:19, June.

Ioana Hulpus, Conor Hayes, Marcel Karnstedt, and Derek Greene. 2013. Unsupervised graph-based topic labelling using dbpedia. In *Proceedings of the sixth ACM international conference on Web search and data mining*, WSDM '13, pages 465–474, New York, NY, USA. ACM.

David Inouye and Jugal K. Kalita. 2011. Comparing twitter summarization algorithms for multiple post summaries. In *SocialCom/PASSAT*, pages 298–306. IEEE.

Jey Han Lau, David Newman, Karimi Sarvnaz, and Timothy Baldwin. 2010. Best Topic Word Selection for Topic Labelling. *CoLing*.

Jey Han Lau, Karl Grieser, David Newman, and Timothy Baldwin. 2011. Automatic labelling of topic models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1536–1545, Stroudsburg, PA, USA. Association for Computational Linguistics.

Chin-Yew Lin, Guihong Cao, Jianfeng Gao, and Jian-Yun Nie. 2006. An information-theoretic approach to automatic evaluation of summaries. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL '06, pages 463–470, Stroudsburg, PA, USA. Association for Computational Linguistics.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Stan Szpakowicz Marie-Francine Moens, editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, Barcelona, Spain, July. Association for Computational Linguistics.

Feifan Liu and Yang Liu. 2008. Correlation between rouge and human evaluation of extractive meeting summaries. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, HLT-Short '08, pages 201–204, Stroudsburg, PA, USA. Association for Computational Linguistics.

Annie Louis and Ani Nenkova. 2009. Automatically evaluating content selection in summarization without human models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 306–314, Stroudsburg, PA, USA. Association for Computational Linguistics.

Annie Louis and Ani Nenkova. 2013. Automatically assessing machine summary content without a gold

standard. *Computational Linguistics*, 39(2):267–300.

Davide Magatti, Silvia Calegari, Davide Ciucci, and Fabio Stella. 2009. Automatic labeling of topics. In *Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications*, ISDA '09, pages 1227–1232, Washington, DC, USA. IEEE Computer Society.

Qiaozhu Mei, Xuehua Shen, and ChengXiang Zhai. 2007. Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 490–499, New York, NY, USA. ACM.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Texts. In *Conference on Empirical Methods in Natural Language Processing*, EMNLP '04, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 262–272, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ani Nenkova and Lucy Vanderwende. 2005. The impact of frequency on summarization. *Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101*.

Ani Nenkova. 2005. Automatic text summarization of newswire: Lessons learned from the document understanding conference. In *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 3*, AAAI'05, pages 1436–1441. AAAI Press.

David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 100–108, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. 2012. Summarizing sporting events using twitter. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, pages 189–198, New York, NY, USA. ACM.

Martin Porter. 1980. An algorithm for suffix stripping. *Program*, 14(3):130–137.

Zhaochun Ren, Shangsong Liang, Edgar Meij, and Maarten de Rijke. 2013. Personalized time-aware tweets summarization. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 513–522, New York, NY, USA. ACM.

Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. 2013. A participant-based approach for event summarization using twitter streams. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '13, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dingding Wang, Shenghuo Zhu, Tao Li, and Yihong Gong. 2009. Multi-document summarization using sentence-based topic models. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 297–300, Stroudsburg, PA, USA. Association for Computational Linguistics.

Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan, and Xiaoming Li. 2012. Identifying event-related bursts via social media activities. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1466–1477, Jeju Island, Korea, July. Association for Computational Linguistics.