Twitter Topic Modeling by Tweet Aggregation

Asbjørn Ottesen Steinskog **Jonas Foyn Therkelsen** **Björn Gambäck**
Department of Computer Science
Norwegian University of Science and Technology
NO–7491 Trondheim, Norway
asbjorn@steinskog.me jonas.foyn@gmail.com gamback@ntnu.no

Abstract

Conventional topic modeling schemes, such as Latent Dirichlet Allocation, are known to perform inadequately when applied to tweets, due to the sparsity of short documents. To alleviate these disadvantages, we apply several pooling techniques, aggregating similar tweets into individual documents, and specifically study the aggregation of tweets sharing authors or hashtags. The results show that aggregating similar tweets into individual documents significantly increases topic coherence.

1 Introduction

Due to the tremendous amount of data broadcasted on microblog sites like Twitter, extracting information from microblogs has turned out to be useful for establishing the public opinion on different issues. O'Connor et al. (2010) found a correlation between word frequencies in Twitter and public opinion surveys in politics. Analyzing tweets (Twitter messages) over a timespan can give great insights into what happened during that time, as people tend to tweet about what concerns them and their surroundings. Many influential people post messages on Twitter, and investigating the relation between the underlying topics of different authors' messages could yield interesting results about people's interests. One could for example compare the topics different politicians tend to talk about to obtain a greater understanding of their similarities and differences. Twitter has an abundance of messages, and the enormous amount of tweets posted every second makes Twitter suitable for such tasks. However, detecting topics in tweets can be a challenging task due to their informal type of language and since tweets usually are more incoherent than traditional documents. The community has also spawned user-generated metatags, like hashtags and mentions, that have analytical value for opinion mining.

The paper describes a system aimed at discovering trending topics and events in a corpus of tweets, as well as exploring the topics of different Twitter users and how they relate to each other. Utilizing Twitter metadata mitigates the disadvantages tweets typically have when using standard topic modeling methods; user information as well as hashtag co-occurrences can give a lot of insight into what topics are currently trending.

The rest of the text is outlined as follows: Section 2 describes the topic modeling task and some previous work in the field, while Section 3 outlines our topic modeling strategies, and Section 4 details a set of experiments using these. Section 5 then discusses and sums up the results, before pointing to some directions for future research.

Proceedings of the 21st Nordic Conference of Computational Linguistics, pages 77–86,
Gothenburg, Sweden, 23-24 May 2017.

2 Topic modeling

Topic models are statistical methods used to represent latent topics in document collections. These probabilistic models usually present topics as multinomial distributions over words, assuming that each document in a collection can be described as a mixture of topics. The language used in tweets is often informal, containing grammatically creative text, slang, emoticons and abbreviations, making it more difficult to extract topics from tweets than from more formal text.

The 2015 International Workshop on Semantic Evaluation (SemEval) presented a task on Topic-Based Message Polarity Classification, similar to the topic of this paper. The most successful systems used text preprocessing and standard methods: Boag et al. (2015) took a supervised learning approach using linear SVM (Support Vector Machines), heavily focused on feature engineering, to reach the best performance of all. Plotnikova et al. (2015) came in second utilizing another supervised method, Maximum Entropy, with lexicon and emoticon scores and trigrams, while essentially ignoring topics, which is interesting given the task. Zhang et al. (2015) differed from the other techniques by focusing on word embedding features, as well as the traditional textual features, but argued that to only extend the model with the word embeddings did not necessarily significantly improve results.

Although the informal language and sparse text make it difficult to retrieve the underlying topics in tweets, Weng et al. (2010) previously found that **Latent Dirichlet Allocation** (LDA) produced decent results on tweets. LDA (Blei et al., 2003) is an unsupervised probabilistic model which generates mixtures of latent topics from a collection of documents, where each mixture of topics produces words from the collection's vocabulary with certain probabilities. A distribution over topics is first sampled from a Dirichlet distribution, and a topic is chosen based on this distribution. Each document is modeled as a distribution over topics, with topics represented as distributions over words (Blei, 2012).

Koltsova and Koltcov (2013) used LDA mainly on topics regarding Russian presidential elections, but also on recreational and other topics, with a dataset of all posts by 2,000 LiveJournal bloggers. Despite the broad categories, LDA showed its robustness by correctly identifying 30–40% of the topics. Sotiropoulos et al. (2014) obtained similar results on targeted sentiment towards topics related to two US telecommunication firms, while Waila et al. (2013) identified socio-political events and entities during the Arab Spring, to find global sentiment towards these.

The **Author-topic model** (Rosen-Zvi et al., 2004) is an LDA extension taking information about an author into account: for each word in a document d, an author from the document's set of authors is chosen at random. A topic t is then chosen from a distribution over topics specific to the author, and the word is generated from that topic. The model gives information about the diversity of the topics covered by an author, and makes it possible to calculate the distance between the topics covered by different authors, to see how similar they are in their themes and topics.

Topic modeling algorithms have gained increased attention in modeling tweets. However, tweets pose some difficulties because of their sparseness, as the short documents might not contain sufficient data to establish satisfactory term co-occurrences. Therefore, pooling techniques (which involve aggregating related tweets into individual documents) might improve the results produced by standard topic model methods. Pooling techniques include, among others, aggregation of tweets that share hashtags and aggregation of tweets that share author. Hong and Davison (2010) compare the LDA topic model with an Author-topic model for tweets, finding that the topics learned from these methods differ from each other. By aggregating tweets written by an author into one individual document, they mitigate the disadvantages caused by the sparse nature of tweets. Moreover, Quan et al. (2015) present a solution for topic modeling for sparse documents, finding that automatic text aggregation during topic modeling is able to produce more interpretable topics from short texts than standard topic models.

3 Extracting topic models

Topic models can be extracted in several ways, in addition to the LDA-based methods and SemEval methods outlined above. Specifically, here three sources of information are singled out for this purpose: topic model scores, topic clustering, and hashtags.

3.1 Topic model scoring

The unsupervised nature of topic discovery makes the assessment of topic models challenging. Quantitative metrics do not necessarily provide accurate reflections of a human's perception of a topic model, and hence a variety of evaluation metrics have been proposed.

The **UMass coherence metric** (Mimno et al., 2011) measures *topic coherence*: $C = \sum_{m=2}^{M} \sum_{l=1}^{m-1} log \frac{D(w_m,w_l)+1}{D(w_l)}$ with $(w_1,...,w_M)$ being the M most probable words in the topic, $D(w)$ the number of documents that contain word w, and $D(w_m,w_l)$ the number of documents that contain both words w_m and w_l. The metric utilizes word co-occurrence statistics gathered from the corpus, which ideally already should be accounted for in the topic model. Mimno et al. (2011) achieved reasonable results when comparing the scores obtained by this measure with human scoring on a corpus of 300,000 health journal abstracts.

However, statistical methods cannot model a human's perception of the coherence in a topic model perfectly, so **human judgement** is commonly used to evaluate topic models. Chang et al. (2009) propose two tasks where humans can evaluate topic models: *Word intrusion* lets humans measure the coherence of the topics in a model by evaluating the latent space in the topics. The human subject is presented with six words, and the task is to find the *intruder*, which is the one word that does not belong with the others. The idea is that the subject should easily identify that word when the set of words minus the intruder makes sense together. In *topic intrusion*, subjects are shown a document's title along with the first few words of the document and four topics. Three of those are the highest probability topics assigned to the document, while the *intruder topic* is chosen randomly.

In addition to these methods, we introduce a way to evaluate author-topic models, specifically with the Twitter domain in mind. A topic mixture for each author is obtained from the model. The human subjects should know the authors in advance, and have a fair understanding of which topics the authors are generally interested in. The subjects are then presented a list of authors, along with topic distributions for each author (represented by the 10 most probable topics, with each topic given by its 10 most probable words). The task of the subject is to deduce which topic distribution belongs to which author. The idea is that coherent topics would make it easy to recognize authors from a topic mixture, as author interests would be reflected by topic probabilities.

3.2 Clustering tweets

An important task related to topic modeling is determining the number of clusters, k, to use for the model. There is usually not a single correct optimal value: too few clusters will produce topics that are overly broad and too many clusters will result in overlapping or too similar topics. The Elbow method can be used to estimate the optimal number of clusters, by running k-means clustering on the dataset for different values of k, and then calculating the sum of squared error (SSE = $\sum_{i=1}^{n}[y_i - f(x_i)]^2$) for each value.

For text datasets, Can and Ozkarahan (1990) propose that the number of clusters can be expressed by the formula $\frac{mn}{t}$, where m is the number of documents, n the number of terms, and t the number of non-zero entries in the document by term matrix. Greene et al. (2014) introduce a term-centric stability analysis strategy, assuming that a topic model with an optimal number of clusters is more robust to deviations in the dataset. However, they validated the method on news articles, that are much longer and usually more coherent than tweets. Greene et al. (2014) released a Python implementation[1] of the stability analysis approach, which we used to predict the optimal number of clusters for a Twitter dataset.

To estimate the number of topics in a tweet corpus, stability analysis was applied to 10,000 tweets posted on January 27, 2016, using a $[2,10]$ k range for the number of topics. An initial topic model was generated from the whole dataset. Proceedingly, τ random subsets of the dataset were generated, with one topic model per k value for each subset $S_1,...,S_\tau$. The stability score for a k value is generated by computing the mean agreement score between a reference set S_0 and a sample ranking set S_i for k: $\sum_{i=1}^{\tau} agree(S_0, S_i)$ (Greene et al., 2014). The number of terms to consider, t, also affects the agreement. A t value of 20 indicates that the top 10 terms for each topic were used. The stability scores were overall low, e.g., $t = 20$ ranging from 0.43 at $k = 2$ to 0.31 at $k = 10$. The low scores are likely caused by the sparse and noisy data in tweets, as this method was originally used for longer, more coherent documents. The method's estimation of number of topics is therefore not a good indication of the number of underlying topics in Twitter corpora.

[1] https://github.com/derekgreene/topic-stability

Good results have been achieved by using a higher number of topics for tweets than what is needed for larger, more coherent corpora, since short messages and the diverse themes in tweets require more topics. Hong and Davison (2010) use a range of topics for tweets from 20 to 150, obtaining the best results for $t = 50$, although the chance of duplicate or overlapping topics increase with the amount of topics.

3.3 Hashtags

Tweets have user-generated metatags that can aid the topic and sentiment analysis. A hashtag is a term or an abbreviation preceded by the hash mark (#). Being labels that users tag their messages with, they serve as indicators of which underlying topics are contained in a tweet. Moreover, hashtags can help discover emerging events and breaking news, by looking at new or uncommon hashtags that suddenly rise in attention. We here present the usage of hashtag co-occurrences to divulge the hidden thematic structure in a corpus, using a collection of 3 million tweets retrieved during the Super Bowl game, February 7th, 2016.

Hashtag co-occurrences show how hashtags appear together in tweets. Since different hashtags appearing in the same tweet usually share the underlying topics, a hashtag co-occurence graph might give interesting information regarding the topics central to the tweet. Looking at which hashtags co-occur with the hashtag *#SuperBowl* gives us more information about other important themes and topics related to Super Bowl. Table 1 shows the 10 most popular hashtags from the Super Bowl corpus, about half of them being re-

Hashtag	Freq.
SB50	10,234
KCA	3,624
SuperBowl	2,985
FollowMeCameronDallas	1,899
Broncos	1,290
EsuranceSweepstakes	1,079
followmecaniff	995
FollowMeCarterReynolds	938
KeepPounding	794
SuperBowl50	783

Table 1: Hashtags during Super Bowl 2016

lated to the Super Bowl. Interestingly, (Denver) *Broncos*, the winning team of the match, was the 5th most mentioned hashtag, while the losing team (Carolina) *Panthers* only was the 17th most popular hashtag that day.

Some hashtags are related to a topic without it being apparent, since it requires further knowledge to understand how they are related: "Keep Pounding" is a quote by the late Carolina Panthers player and coach Sam Mills.

Hashtag co-occurrences help reveal such related hashtags, and hashtag-graph based topic models have been found to enhance the semantic relations displayed by standard topic model techniques (Wang et al., 2014). Figure 1 displays the 20 most co-occurring hashtags in a co-occurrence network for the Super Bowl corpus. Three clusters of hashtags emerge, with Super Bowl related hashtags forming the largest. Related topics and terms become more apparent when displayed in a co-occurrence graph like this, with *KeepPounding* and *SB50* being the 8th most co-occurring hashtag pair, and the artists *Beyonce* and *Coldplay* appearing in the Super Bowl cluster since they performed during the halftime show. The graph also indicates *EsuranceSweepstakes* to be related to Super Bowl, and indeed the company Esurance run an ad during the match, encouraging people to tweet using the hashtag *EsuranceSpeestakes*.

Another cluster consists of the three hash-

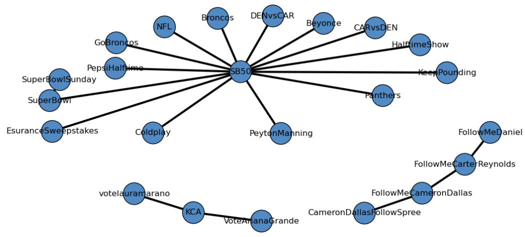

Figure 1: Hashtag co-occurrence network

tags *votelauramarano*, *KCA* and *VoteArianaGrande*. KCA is an abbreviation of *Kid's Choice Awards*, an annual award show where people can vote for their favorite television, movie and music acts, by tweeting the hashtag *KCA* with a specific nominee-specific voting hashtag (e.g., *VoteArianaGrande*).

4 Topic Modeling Experiments

Extending the studies of the previous section, a set of experiments related to topic modeling were conducted, comparing a standard LDA topic model to a hashtag-aggregated model, and comparing two author-topic models.

4.1 Hashtag-aggregated topic model

A pooling technique that involves aggregating tweets sharing hashtags was applied, the assumption being that tweets that share hashtags also share underlying topics. The main goal of this method is the same as for the Author-topic model and other pooling techniques; alleviating the disadvantages of short documents by aggregating documents that are likely to share latent topics. Some restrictions were introduced: only *single-hashtag* tweets were used, and only hashtags that appeared in at least 20 of the documents in the corpus.

Table 2 shows a sample of the resulting topics. They appear more coherent than the topics generated on tweets as individual documents, even though many of the less probable words in each topic might seem somewhat random. It is, however, easier to get an un-

Topic #7	Topic #21	Topic #24	Topic #34
revivaltour	new	purposetourboston	trump
selenagomez	soundcloud	justinbieber	hillary
whumun	news	boston	bernie
wtf	video	one	realdonaldtrump
getting	favorite	best	will
boyfriend	sounds	tonight	clinton
bitch	health	yet	greysanatomy
mad	blessed	shows	vote
resulted	efc	bitcoin	president
blend	mealamovie	redsox	people

Table 2: Four topics after hashtag aggregation

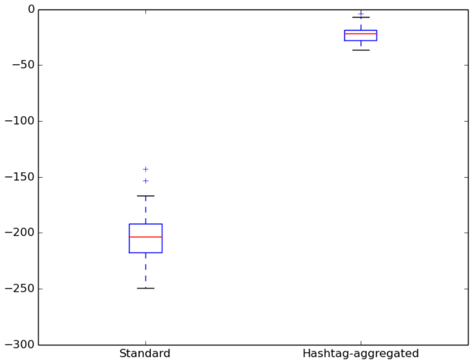

Figure 2: Coherence score of LDA topic model vs. hashtag-aggregated topic model

derstanding of the underlying topics conveyed in the tweets, and aggregating tweets sharing hashtags can produce more coherence than a topic model generated by single tweets as documents. The UMass coherence scores for the topics in this topic model are also much higher than for standard LDA, as shown in Figure 2.

4.2 Author-topic model experiments

Tweets from six popular Twitter users were obtained through the Twitter API, selecting users known for tweeting about different topics, so that the results would be distinguishable. Barack Obama would be expected to tweet mainly about topics related to politics, while the astrophysicist Neil deGrasse Tyson would assumingly tweet about science-related topics. The themes communicated by Obama and Donald Trump ought to be similar, both being politicians, while the inventor Elon Musk ought to show similarities with Tyson. Tweets from two pop artists, Justin Bieber and Taylor Swift, were also included and expected not to share many topics with the other users. To obtain tweets reflecting the author's interests and views, all retweets and quote tweets were discarded, as well as tweets containing media or URLs. Two approaches to author topic-modeling were compared, based on Rosen-Zvi et al. (2004) and on Hong and Davison (2010), respectively.

Ten topics were generated from the Rosen-Zvi et al. (2004) author-topic model, each topic being represented by the 10 most probable words. The resulting topics are reasonably coherent, and can be seen in Table 3. The quality of an author-topic model can be measured in its ability to accurately portray the user's interests. A person that has knowledge of which themes and topics a user usually talks about, should be able to recognize the user by their topic distribution. Eight persons were shown topic distributions such as the one in Figure 3 without knowing which user it belonged to, and asked to identify the users based on the topic distributions.

All participants managed to figure out which topic distribution belonged to which author for all author's, except the distributions of Taylor Swift and Justin Bieber, which were very similar, both having Topic 4 and 5 as most probable. The remaining author's had easily recognizable topic distributions, which was confirmed by the experiment.

The author-topic model proposed by Hong and Davison (2010) performs standard LDA on aggregated user profiles. To conduct the experiments, the model was thus first trained on a corpus where each document contains ag-

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
will	just	earth	night	tonight	just	people	great	president	don
new	like	moon	today	love	one	much	thank	obama	fyi
now	will	day	get	thank	know	time	trump2016	america	time
get	now	sun	new	thanks	orbit	won	will	sotu	tesla
america	good	world	happy	ts1989	two	tonight	cruz	actonclimate	first
make	think	ask	back	taylurking	might	way	hillary	economy	rocket
poll	also	universe	time	show	long	bad	makeamericagreatagain	work	science
many	around	full	see	got	planet	show	big	change	space
trump	live	space	one	crowd	star	morning	cnn	americans	launch
don	landing	year	good	tomorrow	instead	really	said	jobs	model

Table 3: The ten topics generated for the Rosen-Zvi et al. (2004) author-topic model

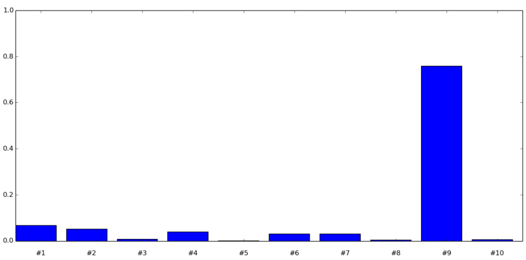

Figure 3: Topic distribution for Obama

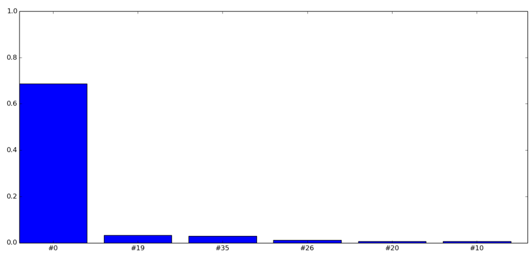

Figure 4: Obama's aggregated topics

gregated tweets for each user. Furthermore, new tweets for each author (which were not part of the training data) were downloaded, and the topic distribution inferred for each of the new tweets. Finally, the topic distribution for each user was calculated as the average topic distribution over all new tweets written by that user. Since a low number of topics produces too many topics containing the same popular words, 50 topics were used instead of the 10 in the previous experiments.

An example of the resulting topic mixtures for the authors can be seen in Figure 4, with the most probable topics for each of the authors tabulated in Table 4. As opposed to the previous topic mixtures, these topic mixtures generally had one topic that was much more probable than the remaining topics. Therefore, the diversity in the language by each author might not be captured as well by this model. On the other hand, the most probable topic for each author generally describes the author with a high precision. It is therefore easier to distinguish Justin Bieber from Taylor Swift than it was in the previous experiment.

5 Discussion and conclusion

A topic modeling system for modeling tweet corpora was created, utilizing pooling techniques to improve the coherence and interpretability of standard topic models. The results indicate that techniques such as author aggregating and hashtag aggregation generate more coherent topics.

Various methods for estimating the optimal number of topics for tweets were tested. The Elbow method almost exclusively suggested a k value between 3 and 5, no matter how large or diverse the corpus was. The stability score (Greene et al., 2014) also produced rather poor estimates for a number of topics when applied to a tweet corpus. The spar-

#0 (Obama)	#20 (Musk)	#26 (Tyson)	#35 (Trump)	#43 (Bieber)	#19 (Swift)
president	tesla	earth	will	thanks	tonight
obama	will	moon	great	love	ts1989
america	rocket	just	thank	whatdoyoumean	taylurking
sotu	just	day	trump2016	mean	just
actonclimate	model	one	just	purpose	love
time	launch	time	cruz	thank	thank
work	good	sun	hillary	lol	crowd
economy	dragon	people	new	good	night
americans	falcon	space	people	great	now
change	now	will	makeamericagreatagain	see	show

Table 4: The most probable topic for the six authors inferred from aggregated topic distribution

sity of tweets is likely the cause of this; the documents do not contain enough words to produce sufficient term co-occurrences. Hong and Davison (2010) found that 50 topics produced the optimal results for their author-topic model, although the optimal number of topics dependents on the diversity of the corpora. Thus *k* values of 10 and 50 were used in the experiments, with 50 for large corpora where a diverse collection of documents was expected.

Hashtag co-occurrences were used to divulge latent networks of topics and events in a collection of tweets. A hashtag aggregating technique was shown to mitigate the negative impacts sparse texts have on the coherence of a topic model. Hashtag aggregation technique is especially interesting, as it utilizes a specific metadata tag that is not present in standard documents. A hashtag aggregated topic model produced a much better coherence than the standard LDA variation for the same corpus; this is also consistent with recent research on topic models for microblogs. Two Author-topic models were used in our experiments, one using the Rosen-Zvi et al. (2004) topic model and the aggregated author-topic model proposed by Hong and Davison (2010), both seeming to produce interpretable topics. It is worth noting that there is no standardized methods for evaluating topic models, as most quantitative ways try to estimate human judgement. Moreover, there is no precise definition of a *gold standard* for topic models, which makes the task of comparing and ranking topic models difficult. A combination of a computational method and human judgement was therefore used in the evaluations.

One way to extend the topic modeling system would be to apply online analysis by implementing automatic updates of a topic model, continuously extended by new tweets retrieved from the Twitter Streaming API. This would help in detect emerging events, in a fashion similar to Lau et al. (2012). Moreover, Dynamic Topic Models could be considered to provide better temporal modeling of Twitter data. A limitation to topic modeling in general is the difficulties in evaluating the accuracy of the models. Computational methods try to simulate human judgement, which poses difficulties, as human judgement is not clearly defined. Further research could help provide better methods for evaluating topic models. In this paper, we aggregated tweets sharing authors and hashtags. Further work should look into other pooling schemes, and see how they compare to author and hashtag aggregation. One example would be to aggregate conversations on Twitter into individual documents. Tweets contain a lot of metadata that can aid the aggregation process.

References

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet Allocation. In *the Journal of Machine Learning Research*, volume 3, pages 993–1022, MIT, Massachusetts, USA. JMLR. org.

David M. Blei. 2012. Probabilistic Topic Models. In *Communications of Association for Computer Machinery*, volume 55, New York, NY, USA, April. ACM.

William Boag, Peter Potash, and Anna Rumshisky. 2015. TwitterHawk: A Feature Bucket Based Approach to Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 640–646, Denver, Colorado, June. Association for Computational Linguistics.

Fazli Can and Esen A. Ozkarahan. 1990. Concepts and Effectiveness of the Cover-coefficient-based Clustering Methodology for Text Databases. In *ACM Transitional Database Systems*, volume 15, pages 483–517, New York, NY, USA, December. Association for Computer Machinery.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, pages 288–296, Vancouver, British Columbia.

Derek Greene, Derek O'Callaghan, editor="Calders Toon Cunningham, Pádraig", Floriana Esposito, Eyke Hüllermeier, and Rosa Meo, 2014. *How Many Topics? Stability Analysis for Topic Models*, pages 498–513. Springer Berlin Heidelberg, Berlin, Heidelberg.

Liangjie Hong and Brian D. Davison. 2010. Empirical Study of Topic Modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 80–88, New York, NY, USA. ACM.

Olessia Koltsova and Sergei Koltcov. 2013. Mapping the public agenda with topic modeling: The case of the Russian LiveJournal. In *Policy & Internet*, volume 5, pages 207–227, Russia.

Jey Han Lau, Nigel Collier, and Timothy Baldwin. 2012. On-line Trend Analysis with Topic Models:\# Twitter Trends Detection Topic Model Online. In *Proceedings of COLING 2012: Technical Papers, pages 1519–1534*, pages 1519–1534, Mumbai, India.

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.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *International Conference on Web and Social Media*, volume 11, pages 1–2, Washington DC, USA.

Nataliia Plotnikova, Micha Kohl, Kevin Volkert, Andreas Lerner, Natalie Dykes, Heiko Emer, and Stefan Evert. 2015. KLUEless: Polarity Classification and Association. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Erlangen, Germany. Friedrich-Alexander-Universitat Erlangen-Nurnberg.

Xiaojun Quan, Chunyu Kit, Yong Ge, and Sinno Jialin Pan. 2015. Short and sparse text topic modeling via self-aggregation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 2270–2276. AAAI Press.

Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The Author-topic Model for Authors and Documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 487–494, Arlington, Virginia, United States. AUAI Press.

Dionisios N Sotiropoulos, Chris D Kounavis, Panos Kourouthanassis, and George M Giaglis. 2014. What drives social sentiment? An entropic measure-based clustering approach towards identifying factors that influence social sentiment polarity. In *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference*, pages 361–373, Chania Crete, Greece. IEEE.

Pranav Waila, VK Singh, and Manish K Singh. 2013. Blog text analysis using topic modeling, named entity recognition and sentiment classifier combine. In *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, pages 1166–1171, Mysore, India. IEEE.

Y. Wang, J. Liu, J. Qu, Y. Huang, J. Chen, and X. Feng. 2014. Hashtag Graph Based Topic Model for Tweet Mining. In *2014 IEEE International Conference on Data Mining*, pages 1025–1030, Shenzhen, China, Dec.

Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. TwitterRank: Finding Topic-sensitive Influential Twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 261–270, New York, NY, USA. ACM.

Zhihua Zhang, Guoshun Wu, and Man Lan. 2015. East China Normal University, ECNU: Multi-level Sentiment Analysis on Twitter Using Traditional Linguistic Features and Word Embedding Features. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Shanghai, China. East China Normal University Shanghai.