Measuring Topic Coherence through Optimal Word Buckets

Nitin Ramrakhiyani[1], **Sachin Pawar**[1,2], **Swapnil Hingmire**[1,3], **and Girish K. Palshikar**[1]

[1]TCS Research, Tata Consultancy Services, Pune
[2]Indian Institute of Technology Bombay, Mumbai
[3]Indian Institute of Technology Madras, Chennai
{nitin.ramrakhiyani,sachin7.p,swapnil.hingmire,gk.palshikar}@tcs.com

Abstract

Measuring topic quality is essential for scoring the learned topics and their subsequent use in Information Retrieval and Text classification. To measure quality of Latent Dirichlet Allocation (LDA) based topics learned from text, we propose a novel approach based on grouping of topic words into buckets (TBuckets). A single large bucket signifies a single coherent theme, in turn indicating high topic coherence. TBuckets uses word embeddings of topic words and employs singular value decomposition (SVD) and Integer Linear Programming based optimization to create coherent word buckets. TBuckets outperforms the state-of-the-art techniques when evaluated using 3 publicly available datasets and on another one proposed in this paper.

1 Introduction

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) based topic modelling uses statistical relations between words like word co-occurrence while inferring topics and not semantic relations. Hence, topics inferred by LDA may not correlate well with human judgements even though they better optimize perplexity on held-out documents (Chang et al., 2009). Given the growing importance of topic models like LDA in text mining techniques and applications (Hingmire et al., 2013; Wang et al., 2009; Lin and He, 2009; Pawar et al., 2016), it is crucial to ensure that the inferred topics are of as high quality as possible. As shown in (Aletras et al., 2017), computing topic coherence is also important for developing better topic representation methods for use in Information Retrieval. An attractive feature of the probabilistic topic models is that the inferred topics can be interpreted by humans, each topic being just a bag of probabilistically selected "prominent" words in that topic's distribution. This has opened up a research area which explores use of human expertise or automated techniques to measure the quality of topics and improve the topic modelling techniques by incorporating these measures. As an example, consider two topics inferred from a document collection (topics are represented by their 10 most probable words):

{loan, foreclosure, mortgage, home, property, lender, housing, bank, homeowner, claim}

{horse, sullivan, business, secretariat, owner, get, truck, back, old, merchant}

The first topic is easily interpretable by humans whereas the second topic is incoherent and less understandable. One could evaluate a single topic or an entire set of topics ("topic model") for quality. Several approaches have been proposed in the literature for measuring the quality of a single topic or of an entire topic model (see Section 2).

In this paper, we aim at measuring the quality of a single topic and propose a novel approach - TBuckets, which groups a topic's words into thematic groups (which we call *buckets*). The intuition is that if a single large bucket is obtained from a topic, the topic carries a single coherent theme. TBuckets combines Singular Value Decomposition (SVD) and Integer Linear Programming (ILP) to achieve an optimal word bucket distribution. We evaluate our technique by correlating its estimated coherence scores with human annotated scores and compare with state-of-the-art results reported in Röder et al. (2015) and Nikolenko (2016). The TBuckets approach not only outperforms the state-of-the-art but also is parameter free. This makes TBuckets directly applicable to topics of a topic model without any searching in a parameter space.

Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 437–442,
Valencia, Spain, April 3-7, 2017.

2 Related Work

Several authors hypothesize that *coherence* of the N most probable words of a topic capture its semantic interpretability. Newman et al. (2010) used the set of N most probable words of a topic and computed its coherence (C_{UCI}) based on *pointwise mutual information* (PMI) between all possible word pairs of N words. In (Aletras and Stevenson, 2013) the authors propose a variant of C_{UCI} by using normalized PMI (NPMI) computed based on distributional similarity between the words of the topic. Each word of a topic is represented by a context vector based on a window context in Wikipedia and coherence is computed as average of cosine similarities between the topic's centroid vector and each word. Mimno et al. (2011) proposes (C_{UMASS}) that uses *log conditional probability* (LCP) instead of PMI and uses the same corpus on which topics are inferred to estimate LCP.

Röder et al. (2015) propose a unifying framework that represents a coherence measure as a composition of parts, that can be freely combined to form a configuration space of coherence definitions. These parts can be grouped into four dimensions: 1) ways a word set can be divided into smaller pieces, 2) word pair agreement measures like PMI or NPMI, 3) ways to estimate word probabilities and 4) methods to aggregate scalar values. This framework spans over a large number of configuration space of coherence measures and it becomes tedious to find an appropriate coherence measure for a set of topics.

Nikolenko (2016), one of the state-of-the-art, also uses distributional properties of words and proposes coherence measures based on word embeddings. Topic quality is defined as average distance between topic words, and four distance functions - cosine, L1, L2 and co-ordinate are proposed. The paper reports strong results on datasets in Russian. Fang et al. (2016) also uses cosine similarity between word embeddings to compute coherence scores for twitter topics. Two other major approaches are based on topic word probability distributions (Alsumait et al., 2009) and coverage and specificities of WordNet hierarchies for topic words (Musat et al., 2011).

3 TBuckets: Creating buckets of topic words

The idea of viewing a topic as a set of coherent word buckets is based on how we humans observe a topic and decide its coherence. A human would observe the topic words one by one and put them in some form of coherent groups (or *buckets*, as we call them). Starting with a fresh bucket for the first word, every new word is put in an already created bucket if the word is semantically similar or semantically associated with the words in the bucket; otherwise the word is put in a new bucket. On completion of this exercise, all topic words would be distributed in various buckets. A distribution with a single large bucket and few small buckets would signify better coherence. However, a distribution with multiple medium sized buckets would indicate lower coherence.

For a coherent topic like {`storm`, `weather`, `wind`, `temperature`, `rain`, `snow`, `air`, `high`, `cold`, `northern`}, which deals with weather and associated factors, the above procedure leads to the following bucket distribution:
Bucket-1: {`storm`, `weather`, `wind`, `temperature`, `rain`, `snow`, `air`, `cold`};
Bucket-2: {`high`};
Bucket-3: {`northern`}

But for a non-coherent topic like {`karzai`, `afghan`, `miner`, `official`, `mine`, `assange`, `government`, `kabul`, `afghanistan`, `wikileaks`} the same procedure leads to the following bucket distribution:
Bucket-1: {`karzai`, `afghan`, `kabul`, `afghanistan`};
Bucket-2: {`miner`, `mine`};
Bucket-3: {`official`, `government`};
Bucket-4: {`assange`, `wikileaks`}

It is evident from above examples that the final distribution of topic words into buckets, reflects the coherence of a topic closely. Based on this idea, we devise the TBuckets approach which enables us to perform this bucketing automatically and generate a coherence score for a topic. It only requires word embeddings of topic words, which are not difficult to obtain as embeddings of a large set of words, trained on various corpora, are now available publicly (Mikolov et al., 2013; Pennington et al., 2014; Levy and Goldberg, 2014)

The idea of clustering arises intuitively when we think of forming related groups among a set of items (words here). However, an important limitation of clustering is that the resulting clusters are sensitive to choice of parameters like linkage configuration, threshold on maximum distance, number of clusters, etc. Furthermore, cluster cen-

troids computed using average of word embeddings might not represent the underlying themes among the words. To really find the underlying themes, it is important to focus on interactions among the features of topic words. The values on dimensions of a word's embeddings can be regarded as the word's abstract features. Considering a matrix capturing interactions among the features of topic words, we hypothesize that the principal eigenvector of this matrix should capture the central theme of the topic. Further, we say that a topic is coherent if most of its words are aligned to this central theme. Additionally other eigenvectors would capture other themes, if any.

To capture this notion, we propose use of Singular Value Decomposition (SVD) and Integer Linear Programming (ILP) for obtaining optimal word theme alignments. We begin by constructing a $n \times d$ rectangular matrix A comprising d dimensional word embeddings of n words of a topic. We then apply SVD on A to obtain a product USV^T where columns of the V matrix are eigenvectors of the feature-feature interaction matrix A^TA. These d dimensional eigenvectors represent the underlying themes we are interested in. The eigenvector corresponding to the largest singular value is the principal eigenvector[1], representing the central theme. Now to determine an initial assignment of words with the eigenvectors, we use the first n eigenvectors in V as bucket identifiers to assign words to. The assignment is n$\ddot{a}$ive - the word goes to the bucket represented by the word's most similar eigenvector. We use cosine similarity to measure similarity between the word's embedding and an eigenvector. We define the principal bucket as the one corresponding to the principal eigenvector.

We believe that this n$\ddot{a}$ive assignment is strict and may lead to formation of multiple distinct but related themes. This may lead to splitting of the central theme across multiple buckets and hence words that should align with the central theme may get aligned to other (related) themes. Hence, to improve the n$\ddot{a}$ive assignment we propose an ILP based optimization and attain an optimal word theme alignment. The details of the optimization formulation are presented in Table 1. We consider the following example topic from the NYT dataset to understand the ILP formulation: {`baby, birth, pregnant, woman, pregnancy, bat, allergy, mother, born, american`}. The human assigned coherence score is 2.15 on a scale of 1 to 3, which is considerable but not too high. The topic's bucket distribution obtained using SVD is:
Bucket-1: {`baby, birth, pregnant, woman, pregnancy, mother`};
Bucket-2: {`allergy`};
Bucket-3: {`american`};
Bucket-4: {`bat`};
Bucket-5: {`born`}

[1] without loss of generality we assume the principal eigenvector to be the first eigenvector

Parameters:
n: No. of eigenvectors/No. of words in a topic
E: Matrix of dimensions $n \times n$, where E_{ij} represents similarity of the j^{th} word with the i^{th} eigenvector
W: Matrix of dimensions $n \times n$, where W_{ij} represents similarity of the i^{th} word with the j^{th} word
L: Matrix of dimensions $(n-1) \times n$, where $L_{ij} = 1$ if $E_{(i+1)j} > E_{1j}$ else 0
Variable:
X: Matrix of dimensions $n \times n$, where $X_{ij} = 1$ only when j^{th} word is assigned to the bucket associated with i^{th} eigenvector
Objective:
Maximize $\sum_{i=1}^{n}\sum_{j=1}^{n} E_{ij} \cdot X_{ij} - \sum_{i=2}^{n}\sum_{j=1}^{n} E_{1j} \cdot X_{ij}$
Constraints:
C_1: $\forall_j$ s.t. $1 \le j \le n$ $\sum_{i=1}^{n} X_{ij} = 1$
C_2: Single constraint $\sum_{j=1}^{n} X_{1j} \ge 1$
C_3: $\forall_{i,j,k}$ s.t. $2 \le i \le n, 1 \le j,k \le n, j \ne k$ $E_{ij} \cdot X_{ij} >= W_{jk} \cdot (X_{1k} - X_{1j} - \sum_{m=2,m\ne i}^{n} X_{mj})$
C_4: $\forall_j$ s.t. $1 \le j \le n$ $X_{1j} \cdot (\sum_{i=1}^{n-1} L_{ij}) <= 1$
C_5: Single constraint $2 \cdot \sum_{j=1}^{n}(X_{1j} \cdot (\sum_{i=1}^{n-1} L_{ij})) <= \sum_{j=1}^{n} X_{1j}$

Table 1: Integer Linear Program (ILP) formulation

3.1 Objective

The objective function consists of two terms. The first term $\sum_{i=1}^{n}\sum_{j=1}^{n} E_{ij} \cdot X_{ij}$ maximizes the similarity between any word with the eigenvector to which it is assigned. Optimizing only this term is equivalent to obtaining the SVD based assignments, as each word gets assigned to the bucket corresponding to its closest eigenvector. The second term $-\sum_{i=2}^{n}\sum_{j=1}^{n} E_{1j} \cdot X_{ij}$ minimizes the penalty for the words which are *not* assigned to the principal eigenvector. The penalty is equal to their similarity with the principal eigenvector. The penalty term favours word assignments to the principal eigenvector by pushing to it some words which are not "too dissimilar" to its theme. The constraints described in the next subsection, bal-

ance addition and restriction of word assignments to the principal eigenvector ensuring a coherent principal bucket.

3.2 Constraints

The first two constraints ensure sanity of the assignments. Constraint C_1 ensures that any word is assigned to one and only one eigenvector and constraint C_2 makes sure that at least one word is assigned to the principal eigenvector.

Constraint C_3 makes sure that any word j which is assigned to a non-principal eigenvector i has more similarity to the eigenvector i than its similarity with any word k assigned to the principal eigenvector. When the j^{th} word itself is assigned to the principal eigenvector then the LHS is always zero and the RHS is either zero or negative; hence satisfying the constraint trivially. When the j^{th} word is assigned to a non-principal eigenvector i, then $E_{ij} \cdot X_{ij}$ represents its similarity with the i^{th} eigenvector. As both the terms X_{1j} and $\sum_{m=2,m\neq i}^{n} X_{mj}$ would be zero, the RHS will reduce to $W_{jk} \cdot X_{1k}$ which is similarity of the j^{th} word with the k^{th} word when the k^{th} word is assigned to the principal eigenvector.

It can be observed that the penalty term and constraint C_3, both favour assignments to the principal eigenvector. If the ILP formulation is restricted to only the three constraints C_1, C_2 and C_3, the example topic results in the following bucket distribution:

Bucket-1: {baby, birth, pregnant, woman, pregnancy, mother, born, american};
Bucket-2: {allergy};
Bucket-3: {bat}

The constraint C_4 ensures that for any word which is assigned to the principal eigenvector, it is either the word's most similar eigenvector or second most similar eigenvector. This constraint ensures that words highly dissimilar to the principal eigenvector do not get forced to the principal bucket. For any word j, the sum $\sum_{i=1}^{n-1} L_{ij}$ represents the number of eigenvectors which are more similar to it than the principal eigenvector. Hence, for each word assigned to the principal eigenvector, the LHS simply counts the number of other more similar eigenvectors and the constraint restricts this count to 1. Therefore, constraint C_4 ensures that there can be only two types of words in the principal bucket: i) words for which the principal eigenvector is the most similar and ii) words for which the principal eigenvector is the second most similar.

It is important to further improve the set of words that get attached to the principal eigenvector. Maintaining that words of type (i) are always in majority would imply adding lesser words which have the principal eigenvector as their second most similar eigenvector. Constraint C_5 ensures that words of type (i) are always in majority.

It can be observed that as against the principal-eigenvector-favouring nature of the penalty term and constraint C_3, both constraints C_4 and C_5 inhibit addition of dissimilar terms and ensure thematic coherence in the principal bucket. The complete ILP formulation for the example topic results in the following bucket distribution. It is evident that constraints C_4 and C_5 evict the term american, ensuring a coherent principal bucket.

Bucket-1: {baby, birth, pregnant, woman, pregnancy, mother, born};
Bucket-2: {american};
Bucket-3: {allergy};
Bucket-4: {bat}

The constraints in the ILP formulation can also be viewed as a set of flexible settings, and depending on the desired representation of the learned topics, the constraints can be loosened or tightened leading to an optimal bucket distribution.

The coherence score of the topic is defined as the size of the principal bucket after optimization.

4 Experimental Analysis

4.1 Datasets

We evaluate TBuckets on 4 datasets - 20 NewsGroups (20NG), New York Times (NYT), Genomics and ACL. Each dataset consists of a set of 100 topics where each topic is represented by its 10 most probable words. Each topic is associated with a real number between 1 and 3 indicating human judgement of its coherence. Detailed description of 20NG, NYT and Genomics datasets is provided in Röder et. al (2015).

We inferred the 100 topics for the ACL dataset[2] on the ACL Anthology Reference Corpus (Bird, 2008). We obtained the gold coherence scores for these topics from three annotators by following the methodology described in Röder et. al (2015).

[2]topics and coherence scores are available at https://www.cse.iitb.ac.in/~sachinpawar/TopicQuality/dataset.html

	Setting	NYT	20NG	Genomics	ACL	Mean
(Röder et al., 2015)	CV	0.803	0.859	**0.773**	0.160	0.649
	CP	0.757	0.825	0.721	0.215	0.629
	CA	0.747	0.739	0.53	0.167	0.546
	NPMI	0.806	0.78	0.594	0.228	0.602
	UCI	0.783	0.696	0.478	0.190	0.537
	UMASS	0.543	0.562	0.442	0.078	0.406
(Nikolenko, 2016)	Cosine	0.75	0.766	0.648	0.248	0.603
	L1	0.431	0.492	0.369	0.017	0.327
	L2	0.448	0.535	0.38	0.021	0.346
	Co-ord	0.447	0.536	0.388	0.131	0.376
Clustering		0.745	0.856	0.709	**0.293**	0.651
SVD		0.758	0.867	0.698	0.227	0.638
Tbuckets		**0.819**	**0.87**	0.729	0.272	**0.673**

Table 2: Pearson Correlation based performance

For all our experiments, we use the 300 dimensional pre-trained word embeddings provided by the GloVe framework (Pennington et al., 2014).

4.2 Evaluation

We use the same evaluation scheme used in (Röder et al., 2015). Each technique generates coherence scores for all the topics in a dataset. Pearson's r correlation co-efficient is computed between the coherence scores based on human judgement and the coherence scores automatically generated by the technique. Higher the correlation with human scores, better is the performance of the technique at measuring coherence.

Table 2 shows the Pearson's r values obtained from the state-of-the-art (Röder et al. (2015) and Nikolenko (2016)) and baselines (Clustering and Only SVD) compared with TBuckets. We consider scores on NYT, 20NG and Genomics as reported in (Röder et al., 2015) and obtain scores on the ACL dataset using the web demo provided by the authors at `http://palmetto.aksw.org/palmetto-webapp/`

As observed in Table 2, TBuckets outperforms (Röder et al., 2015) on 3 out of 4 and (Nikolenko, 2016) on all 4 datasets. It also outperforms all the baselines considering average performance across all datasets. This is significant considering the fact that TBuckets is parameter less whereas the state-of-the-art technique (Röder et al., 2015) requires considerable tuning of multiple parameters. This also is a sound validation of the TBuckets idea for measuring topic coherence.

Effect of word polysemy: The TBuckets approach relies on word embeddings for capturing the semantic relations among topic words. An important limitation of word embeddings is that a single representation of a word is learned irrespective of its senses. Hence it is observed that infrequent or domain-specific senses of polysemous words are not represented sufficiently. Coherent topics containing such polysemous words can still be judged coherent by humans as they can easily consider the appropriate sense of these words looking at the context of other topic words. TBuckets however, is unable to consider infrequent or domain-specific senses of such words, resulting into multiple unnecessary buckets and lower coherence. For a coherent topic from the ACL dataset: {`derivation, probabilistic, pcfg, collins, subtree, production, child, charniak, parser, treebank`}, TBuckets produces three non-principal buckets for the words `child`, `production` and `collins`. A similar example from 20NG is {`game, team, player, baseball, win, fan, run, season, hit, play`}, where TBuckets creates a separate bucket for the word `fan` due to its infrequent sense of "sports fan".

5 Conclusion and Future Work

We proposed a novel approach TBuckets to measure quality of Latent Dirichlet Allocation (LDA) based topics, based on grouping of topic words into buckets. TBuckets uses singular value decomposition (SVD) to discover important themes in topic words and ILP based optimization to find optimal word-bucket assignments. We evaluated TBuckets on LDA topics of 4 datasets, by correlating the estimated coherence scores with human annotated scores and demonstrated the best average performance across datasets. Moreover, as compared to the state-of-the-art techniques which need to tune multiple parameters, TBuckets requires no parameter tuning.

In future, we plan to devise better ways to compute word similarities which would be more suitable for specific domains like Genomics. One possible way is to train word embeddings on a domain specific corpus and use the learned embeddings. Also we intend to study the impact of using coherent topics for text classification and other NLP applications. We would also like to explore a new topic generation process which incorporates semantic relations between words, in addition to their statistical co-occurrence, leading to generation of semantically coherent topics.

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.

Nikolaos Aletras, Timothy Baldwin, Jey Han Lau, and Mark Stevenson. 2017. Evaluating topic representations for exploring document collections. *Journal of the Association for Information Science and Technology*, 68(1):154–167.

Loulwah Alsumait, Daniel Barbar, James Gentle, and Carlotta Domeniconi. 2009. Topic Significance Ranking of LDA Generative Models. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part I*, pages 67–82. Springer-Verlag.

Steven Bird. 2008. Defining a Core Body of Knowledge for the Introductory Computational Linguistics Curriculum. In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, pages 27–35, Columbus, Ohio, June. Association for Computational Linguistics.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, Jan.

Jonathan Chang, Jordan L. Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems 22*, pages 288–296.

Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel. 2016. Using Word Embedding to Evaluate the Coherence of Topics from Twitter Data. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1057–1060. ACM.

Swapnil Hingmire, Sandeep Chougule, Girish K. Palshikar, and Sutanu Chakraborti. 2013. Document Classification by Topic Labeling. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 877–880. ACM.

Omer Levy and Yoav Goldberg. 2014. Dependency-Based Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308, Baltimore, Maryland, June. Association for Computational Linguistics.

Chenghua Lin and Yulan He. 2009. Joint Sentiment/Topic Model for Sentiment Analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 375–384. ACM.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781.

David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing Semantic Coherence in Topic Models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 262–272, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Claudiu Cristian Musat, Julien Velcin, Stefan Trausan-Matu, and Marian-Andrei Rizoiu. 2011. Improving Topic Evaluation Using Conceptual Knowledge. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 1866–1871. 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*, pages 100–108, Los Angeles, California, June. Association for Computational Linguistics.

Sergey I. Nikolenko. 2016. Topic Quality Metrics Based on Distributed Word Representations. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1029–1032. ACM.

Sachin Pawar, Nitin Ramrakhiyani, Swapnil Hingmire, and Girish Palshikar. 2016. Topics and Label Propagation: Best of Both Worlds for Weakly Supervised Text Classification. In *Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2016), LNCS 9624*. Springer.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October. Association for Computational Linguistics.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 399–408. ACM.

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*, pages 297–300, Suntec, Singapore, August. Association for Computational Linguistics.