Online Debate Summarization using Topic Directed Sentiment Analysis

Sarvesh Ranade, Jayant Gupta, Vasudeva Varma, Radhika Mamidi
International Institute of Information Technology, Hyderabad
{sarvesh.ranade | jayant.gupta}@research.iiit.ac.in
{vv | radhika.mamidi}@iiit.ac.in

ABSTRACT

Social networking sites provide users a virtual community interaction platform to share their thoughts, life experiences and opinions. Online debate forum is one such platform where people can take a stance and argue in support or opposition of debate topics. An important feature of such forums is that, they are dynamic and increase rapidly. In such situations, effective opinion summarization approaches are needed so that readers need not go through the entire debate. This paper aims to summarize online debates by extracting highly topic relevant and sentiment rich sentences. The proposed approach takes into account topic relevant, document relevant and sentiment based features to capture topic opinionated sentences. ROUGE scores are used to evaluate our system. Our system significantly outperforms several baseline systems and show 5.2% (ROUGE-1), 7.3% (ROUGE-2) and 5.5% (ROUGE-L) improvement over the state-of-the-art opinion summarization system. The results verify that topic directed sentiment features are most important to generate effective debate summaries.

1. INTRODUCTION

With the exponential growth in the use of World Wide Web, online users express themselves on continuously emerging social networking sites. These sites provide users a wide variety of choices. Micro-blogging sites like twitter allows them to express their opinion (140 characters) on trending topics. Users can express their views and share their experiences on popular web blogs like wordpress, blogspot, etc. Facebook allows community oriented interaction, restricted within one's friend circle. E-commerce users can provide product specific reviews on online shopping sites.

Amongst social networking platforms, online debate sites ('convinceme.net', '4forums.net', 'onlinedebate.net') have become popular in recent times. These online debate forums provide users an option to express their opinion about their favorite debate topics [28]. From the research point of view, they provide a rich collection of differing opinions on various topics. In ideological two-sided debates, users support their stance by cleverly stating arguments supporting their stance or opposing other stance [27]. Conversation sentences between two users are very common in this multi-party conversation. People rebut to another user's post and express their viewpoint on other's opinion [2]. Because of dynamic nature of debates and large number of posts (194 per debate) they contain, it is essential to generate effective summaries for them so that readers need not go through the entire debate.

To help this cause, we need to summarize online debates such that, after reading the summaries, user gets a good idea about the information debate presents and the opinions users express. This paper aims to summarize online debates collected from a popular debate site called 'convinceme.net' with the intention of capturing good topic information as well as opinion rich sentences from the debate.

Compared to generic summarization, opinion summarization is a relatively novel area. Unlike traditional methods, two factors, the sentiment degree and the correlated events, play a major role in opinion summarization. Previous methods [29, 15] have effectively used these factors over news, blogs and conversation domain. However, online debates is a domain which is yet to be explored. An important aspect of online debates is that most of its sentences are sentiment rich and topic relevant. Thus, topic directed sentiment analysis is an important feature to create effective summaries. We have successfully used this feature and results validate the effectiveness of our approach.

In our method, we analyzed factors governing important sentences in debate summaries. We observed 3 important factors: informative sentences, sentiment rich sentences and sentences which describes topic related entities. Thus, topic related and sentiment carrying features are used in the proposed approach. We have also used positional features as they have been effective in generic summarization approaches. Document relevant features such as tf-idf scores are used to capture content relevant sentences.

Our system, *DEBSumm* generates extractive summaries using the aforementioned features. ROUGE [18] scores have been used to evaluate the system summaries. Final results show that sentiment words that are relevant to topic are the most important feature to create effective summaries. The results also show that our system achieve better results than previous state-of-the-art and several baseline systems.

The rest of the paper is organized as follows: Section 2 describes related work; Section 3 gives a detailed description of our approach. Section 4 describes experimental setup for experiments; Section 5 discuss results of different experiments performed; Section 6 concludes the paper with final comments and future work.

2. RELATED WORK

An extensive work has been done in the field of extractive text summarization. Earliest work by Luhn [19] used frequency based features to score sentences. Later work [9, 16] added features such as topic signature, cue words, data annotation, etc. These features were used to score sentences and top sentences were selected for summary. In MEAD [25], clusters were created and sentences were scored using sentence and inter sentence features. Redundancy removal has been a big issue in summarization for which a benchmark approach was proposed by Carbonnell et al. [7]. This approach used MMR to create a balance between information novelty and importance to create non redundant summaries.Graph based approaches [26, 20] represent text as a graph where a text entity (sentence, phrases) represent vertices connected by similarity based measures. Salton et al. [26] use cosine similarity to connect vertices, then a greedy graph traversal technique is applied in chronological order to form summary.

In context of opinion summarization, identification of opinion containing sentences is important. Sentence relevance is further decided by their sentiment scores, topic relevance and other lexical and positional features. Earlier works mainly focused on reviews [24, 13, 22] which used lexical features (unigram, bigram and trigram), part-of-speech tags and dependency relations.

Ku et al. [15] performed opinion summarization in news and blog domain. They propose opinion extraction at word, sentence and document level. For each new word, distribution of its characters (Chinese) as positive and negative polarity in the seed vocabulary (created manually) is used to determine sentiment of the word. These scores are compounded to compute sentence scores and then document scores. Presence of negation operators decided the sentiment tendency at sentence level which further propagated to document level.

Wang et al. [29] performed opinion summarization on conversations. They used linear combination of features from different aspects including topic relevance, subjectivity and sentence importance to score sentences. They also proposed a graph based method, which incorporates topic and sentiment information, as well as additional information about sentence to sentence relations extracted based on dialogue structures.

Mining of opinion goes hand in hand with analyzing sentiments. In this perspective, a detailed study of past work, present trends and futures needs has been done by Cambria et al. [5]. Significant work has been done in social media on target directed sentiment analysis [1, 23, 21]. Agarwal et al. [1] used syntactic features as target dependent features to differentiate sentiment words' effect on different targets in a tweet. O'Hare et al. [23] employed word-based approach to calculate sentiments directed towards companies and their stocks from financial blogs.

Cambria et al. [6] applied semantic multi-dimensional scaling on a knowledge base of affective common-sense knowledge for text classification, emotion recognition, and patient opinion mining. Mukherjee et al. [21] applied clustering to extract feature specific opinions and calculated overall feature sentiment using subjectivity lexicon.

Opinion summarization in the specific domain of online debates is a novel field. This domain differs from chatting and conversation because it is more formal and focuses on specific topics. It may be possible that the argument contains various different factual knowledge but they are usually related to one or the other topic. Similarly, it is different from news and blogs because it is comparatively more rich in sentiment. Therefore, opinion mining in debates is an interesting and challenging task.

3. APPROACH

Extractive summaries are generated by ranking the Dialogue Acts (DAs)[1] from the original documents. We calculate their importance according to linear combination of scores using several features. Equation 1 is used to assign score to each DA s.

$$\begin{aligned} score(s) =& \lambda_{topicRel} topicRel(s, topics) + \lambda_{docRel} docRel(s, D) \\ &+ \lambda_{sentiRel} sentiRel(s) + \lambda_{conRel} conRel(s, D) \end{aligned} \tag{1}$$

Most highly ranked DAs are chosen until summary length constraint is satisfied. Table 1 lists the set of features used in this equation. We describe each of these features in the subsequent subsections.

Feature Category	Feature Names
Topic Relevance	Topic Directed Sentiment Score Topic Co-occurrence
Document Relevance	tf-idf Sentiment Score
Sentiment Relevance	Number of Sentiment Words Sentiment Strength
Context Relevance	Sentence position Sentence length

Table 1: Argument Structure Examples

3.1 Topic Relevant Features

Debate posts present users' opinion towards debate topics. Thus, sentences which provide information or express opinion about debate topics are most important in the context of debate summarization. We use topic directed sentiment scores and topic co-occurrence measure to capture topic relevance of the DAs.

3.1.1 Topic Directed Sentiment Score

Topic related sentiment carrying DAs are very important in the context of online debates. They represent the sentiments directed by DA toward debate topics and thus, a key feature in the task of debate summarization. In the proposed approach, the sentiment score directed towards debate topics is calculated using dependency parse of the DAs and sentiment lexicon *SentiWordNet [4]*.

Pronoun referencing is resolved using Stanford co-reference resolution system [17]. Then using Stanford dependency parse [8], DAs are represented in tree format where each node represents a DA word storing its sentiment score and the edges represent dependency relations. Each DA word is looked in SentiWordNet and the sentiment score calculated using Algorithm 1 is stored in the word's tree node.

[1]Dialogue Act is smallest unit of debate.

Algorithm 1 Word Sentiment Score

$S \leftarrow Senses\,of\,word\,W$
$wordScore \leftarrow 0$
for all $s \in S$ **do**
  $s_{score} = s_{posScore} - s_{negscore}$
  $wordScore = wordScore + s_{score}$
end for
$wordScore = \frac{wordScore}{|S|}$

SentiWordNet is a lexical corpus used for opinion mining. It stores positive and negative sentiment scores for every sense of the word present in WordNet [10]. For words missing from SentiWordNet, average of sentiment scores of its synset member words is stored in the word's tree node, otherwise zero sentiment score is stored. If words are modified by negation words like {'never','not','nonetheless', etc.}, their sentiment scores are negated.

In noun phrases 'great warrior', 'cruel person', etc. first word being the adjective of the latter, influences its sentiment score. Thus, based on the semantic significance of the dependency relation each edge holds, sentiment score of parent nodes are updated with that of child nodes using Algorithm 2. In DAs like "Batman killed a bad guy.", sentiment score of word 'Batman' is affected by action 'kill'. Thus, for verb-predicate relations like {'nsubj','dobj','cobj','iobj', etc.}, predicate sentiment scores are updated with that of verb scores using Algorithm 2.

Algorithm 2 Update Word Sentiment Score

$node \leftarrow Word's\,Tree\,Node$
$childs \leftarrow Word's\,child\,nodes$
for all $c \in childs$ **do**
  $updateScore(c)$
  $node_{score} \leftarrow sign(node_{score}) * (|node_{score}| + (c_{score}))$
end for

Tree structure and recursive nature of Algorithm 2 ensures that sentiment scores of child nodes are updated before updating their parents' sentiment scores. Table 2 lists the semantically significant dependency relations used to update parent node scores.

Modification Type	**Dependency Relations**
Noun Modifying	nn, amod, appos, abbrev, infmod, poss, rcmod, rel, prep
Verb Modifying	advmod, acomp, advcl, ccomp, prt, purpcl, xcomp, parataxis, prep

Table 2: List of Dependency Relations

Extended Targets (ET): Extended targets are the entities closely related to debate topics. For example, 'Joker', 'Clarke Kent' are related to 'Batman' and 'Darth Vader', 'Yoda' to 'Star Wars'. To extract the extended targets, we capture named entities (NE) from Wikipedia page of the debate topic using Stanford Named Entity Recognizer [11] and sort them based on their frequency. Out of *top-k* ($k = 20$) NEs, some can belong to both of the debate topics. For example, 'DC Comics' is common between 'Superman' and 'Batman'. We remove these NEs from individual lists and the remaining NEs are treated as extended targets (*extendedTargets*) of the debate topics.

Now that we have a list of extended targets for debate topics and a sentiment score for each DA word, topic directed sentiment scores are calculated for each debate topic using Equation 2.

$$Topic\ Score_{DA} = \sum_{\substack{w \in DA \\ w \in ET(Topic)}} (Score(w)) \tag{2}$$

We refer to these scores as *AScore* and *BScore* representing scores directed towards topics A and B in debate between these two topics respectively.

Absolute value of both topic directed sentiment scores are added reprensenting DA's topic directed sentiment score. These scores are normalized with the sum of topic directed sentiment score of all the DAs.

3.1.2 Topic Co-occurrence Measure

Topic co-occurrence measure captures DAs containing high sentiment words which highly co-occur with debate topic. Extended targets previously described represent debate topic entities. Topic co-occurrence measure is computed using HAL from the Equation 3, capturing co-occurrence measure of DA words and their sentiment strengths. Sentiment score is calculated using Algorithm 1.

$$Co{-}occur_{DA} = \sum_{w \in DA} \left(\sum_{t \in ET} (HAL(w|t)) * sentiScore(w) \right) \tag{3}$$

Topic-occurrence measure is normalized with the sum of co-occurrence scores of all the DAs. We sum up topic directed sentiment scores and topic co-occurrence measure giving us the topic relevance feature score for DAs.

3.2 Document Relevance Features

Tf-idf and sentiment score of the words are used to compute document relevance of the DAs using Equation 4.

$$tf - idf_{DA} = \sum_{w\,\in\,DA} (tf - idf(w)\ *\ sentiScore(w)) \tag{4}$$

Tf-idf score reflects how important a word is to a document in a collection or corpus. Sentiment score carrying words' sentiment strength reflects subjective importance of the word in the context of opinion DAs. Thus, this feature captures the DAs containing highly frequent sentiment rich words. Document relevance score of the whole debate DAs is used to normalize individual scores.

3.3 Sentiment Relevance Features

This dimension captures the presence of sentiment carrying words and their strength in the DAs.

1. *sentiCount* is the count of sentiment carrying words in the DAs. *sentiCount* is normalized with total number of sentiment words present in the debate.

2. Sentiment score of each DA word is calculated using Algorithm 1 and Equation 5 is used to compute DAs' sentiment strength. Sentiment score for each DA is normalized with overall debate's sentiment score.

$$sentiScore_{DA} = \sum_{w \in DA} sentiScore(w) \tag{5}$$

Sentiment score and number of sentiment words in DAs are added which represents the sentiment relevance feature score of DAs.

3.4 Document Context Features

3.4.1 Sentence Position

Sentence position plays important role in predicting the presence of DAs in summary. In debates, initial and ending DAs of the debate posts are more important than the middle ones. So, we have used Equation 6 to compute sentences' position based score which gives higher values for initial and ending sentence than the middle ones. This score is normalized by dividing it with number of DAs in debate posts[2].

$$posScore_{DA} = \frac{\mid \frac{N}{2} - DA_{position} \mid}{N}, N = Total\ DAs\ in\ Post \quad (6)$$

3.4.2 Sentence Length

As the longer sentences tend to contain more information, we have used sentence length as document context feature. It also avoids short sentences (smaller than 5 words) which are less likely to contribute to summary because of incompleteness or less information. Sentence length is the number of words in the DAs. We have normalized the sentence length with the number of words in the whole debate.

We sum sentence position and sentence length scores to compute document context feature score of DAs.

Note that all the values have been normalized over all DAs in the debate so that the different feature scores are comparable.

4. EXPERIMENTAL SETUP

In this study, we extracted 10 online debate discussions from *www.convinceme.net*. These discussions are freely available on aforementioned site and Table 3 shows the statistics of the dataset used. Each of these discussions focus upon different topics allowing us to produce results over various domains.

Number of users	Number of posts	Number of DA
1168	1945	23681

Table 3: Statistics of the dataset

For evaluation, extractive gold set summaries were created by 2 language editors. They were asked to create 500, 1000, 1500, 2000 word summaries. Inter-editor agreement was calculated to be 71.7%[3]. The editors were asked to select the sentences on the following order:

1. Sentiment rich which contains highly topic-relevant information.
2. Sentiment rich with relevant information (low noise).
3. Less subjective content but rich in information.
4. Highly subjective sentence with no relevant information and factual statements should be selected with care. The reason being that they add noise without taking any particular stand.

All the evaluation scores are computed using ROUGE [18] which stands for Recall Oriented Understudy of Gisting Evaluation. It has been widely used by DUC to evaluate system summaries. ROUGE measures summary quality by counting overlapping units such as the n-gram, word-sequences and word-pairs between system summaries and human summaries. Three automatic evaluation methods ROUGE-1, ROUGE-2 and ROUGE-L were chosen to calculate scores. They compute unigram recall, bigram recall and longest common subsequence respectively.

We have conducted the following experiments :

1. Comparison of $DEBSumm$ summaries with proven baseline and state-of-the-art summarization systems explained in Section 5.
2. Effect of variable summary size on $DEBSumm$ and state-of-the-art systems.

5. RESULTS AND DISCUSSION

Grid search was used to compute best parameter values for Equation 1. Following values gave the best results as indicated by ROUGE results: $\lambda_{topicRel} = 0.3, \lambda_{docRel} = 0.1, \lambda_{sentiRel} = 0.5, \lambda_{conRel} = 0.1$.[4]

Scores show that better summaries are obtained when sentiment rich sentences are selected. Furthermore, sentiments which are directed towards the topic words are also given higher weightage. Other measures like sentence position and length give a better fine tuning to summaries as they help to differentiate between similar sentences. Low weightage to document relevance score is understandable because it is a redundant feature to identify sentiment rich document words.

We compared our system ($DEBSumm$) to the following systems:

1. **Max-length [12]:** Longest sentences were selected from all the users. In case, summary is short of length second-longest sentences are selected. This step is iterated until summary reaches required length. This is a proven strong baseline for conversation summarization.
2. **Lead [30]:** Top sentences from each user were selected where each sentence has to be greater than 4 words. In case, summary is short of length, next sentence is selected. This step is iterated until summary reaches required length.
3. **pHAL [14]:** Sentence (S) score was calculated by combining the pHAL scores of each of sentence words. pHAL score of each word is calculated as follows,

$$pHAL(w) = \sum_{w' \epsilon ET} \frac{HAL(w'|w)}{n(w) * K}$$

$$Score(S) = \sum_{w_i \epsilon S} (P(w_i) \times pHAL(w_i))$$

[2]Post represents a user argument and consists of multiple DAs

[3]Number of common sentences were averaged over the complete set of debates.

[4]All the further experiments were conducted using these values.

For summary creation, top scored sentences were selected from sorted list of sentences.

4. **tf-Idf [3]:** Sentences were scored by combining the tf-idf measures of their words[5]. For summary creation, top most sentences were selected from sorted list of sentences.

5. **OpinionSumm [29]:**[6] This is a sentence scoring approach where sentence are scored based on their document similarity, topic relevance, sentiment and length. We have used the same parameter values experimentally calculated in their work. This is a state-of-the-art opinion summarization system.

In the field of generic summarization, system 2 and 4 are proven strong baseline and system 3 is a state-of-the-art system.

Table 4: ROUGE Scores (Average F-measure) of System Summaries (1000 words)

System	ROUGE-1	ROUGE-2	ROUGE-L
Max-Length	0.49892	0.18453	0.48343
Lead	0.49068	0.14759	0.47839
pHAL	0.48985	0.16468	0.46955
tf-idf	0.49922	0.17585	0.48035
OpinionSumm	0.51631	0.20364	0.49849
DEBSumm	**0.56833**	**0.27044**	**0.55326**

Table 4 shows ROUGE scores (Average F-measure) of different systems. The summary size is taken to be 1000 words. Note that, each of the systems 1, 2, 3 and 4, is one of the lower weighted components of the function used to compute our ($DEBSumm$) scores. On the other hand, OpinionSumm represents the higher weighted sentiment component of $DEBSumm$. The results show that $DEBSumm$ comprehensively outperforms the state-of-the-art systems. They also show an improvement of 5.2% (ROUGE-1), 7.3% (ROUGE-2) and 5.5% (ROUGE-L) over OpinionSumm. The above results show that sentiment, topic directed or independent of it, is very important factor to compute effective summaries.

Evaluating systems over variable summary size allows us to judge systems over wide range of summary length. Shorter summaries require higher precision and longer summaries require high recall. As the summary size increases, number of sentences which add novel relevant information decreases. Thus, rate of change in scores is not significant. However, in our graph (Figure 1) we find that there is a slight decrease in scores of OpinionSumm and $DEBSumm$ from 500 to 1000 words. We believe the reason of such a behavior to inclusion of new noisy data as compared to relevant data. This suggests that more relevance should be given to structural and document features over features representing sentiments. Overall, Figure 1 shows that $DEBSumm$ consistently outperforms other systems over different summary sizes.

[5]Each user discussion is considered as a single document while calculating tf-idf values

[6]Note that OpinionSumm is the name given to this system to refer it, throughout, this paper only.

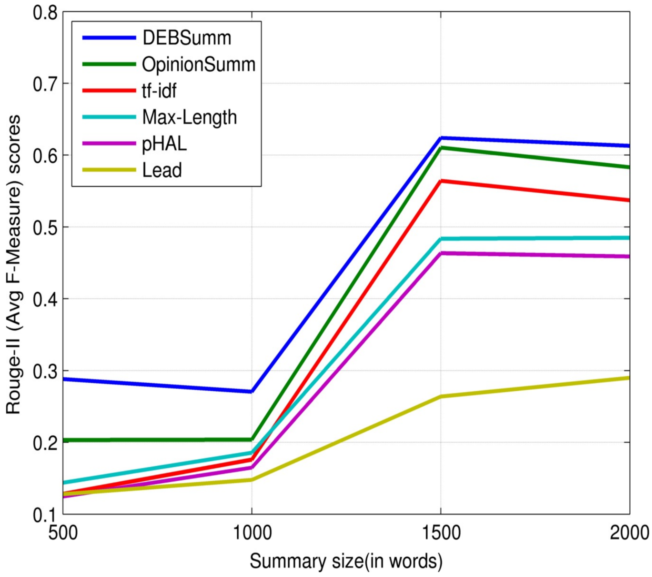

Figure 1: ROUGE-2 (Average F-measure) scores v/s Summary Size (in words)

6. CONCLUSION AND FUTURE WORK

Online debates provides topic related information as well users' opinion regarding the debate topics. Because of large amount of differing opinions, it is necessary to summarize these debates. This paper focuses on summarizing the online debates on the basis of topic directed sentiment and topic related information rich features. To our knowledge, this is the first work in the area of debate summarization. Our sentence ranking based approach ranks debate sentences based on the features related to topic information, sentiment and document statistics. Topic directed sentiment analysis is used to capture sentiment directed towards debate topics whereas topic based co-occurrence measure is used to describe debate topics and sentence closeness. Features commonly used in general text summarization approaches like tf-idf, sentence length and position are also used to capture document statistics based rich sentences. We have compared our system's performance with generic and opinion based state-of-the-art systems. The results show that our system beats all these systems comprehensively.

In this approach, we are averaging sentiment scores of all senses of a word, because of poor state of word sense disambiguation in current scenario, which will not work in all cases. Some words carry different sentiment in different domains for example, 'refined' word is good for 'oil products' whereas bad in the domain of 'agriculture products'. Therefore, next we will be using word sense disambiguation and domain specific sentiment analysis in our system. We will also include debate structure features. These features can leverage DAs occurring along with a high scoring DA. They can also identify related DAs spanned across different users and help in identifying relevant DAs more effectively.

Creating users' profile by capturing their intentions, support by other users and rebuttal arguments can prove a crucial factor in terms of determining the users' expertise. Therefore, we plan to investigate the role of opinion holder in the task of debate summarization.

7. REFERENCES

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data. pages 30–38. LSM, 2011.

[2] P. Anand, M. Walker, R. Abbott, J. Tree, R. Bowmani, and M. Minor. Cats rule and dogs drool!: Classifying stance in online debate. In WASSA 2011, pages 1–9, 2011.

[3] C. Aone, M. E. Okurowski, and J. Gorlinsky. Trainable, scalable summarization using robust nlp and machine learning. In COLING, 1998, pages 62–66.

[4] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. LREC, Valletta, Malta, 2010. ELRA, 2010.

[5] E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems 28(2)*, pages 15–21, 2013.

[6] E. Cambria, Y. Song, H. Wang, and N. Howard. Semantic multi-dimensional scaling for open-domain sentiment analysis. *IEEE Intelligent Systems*, 2013.

[7] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In SIGIR, 1998, pages 335–336.

[8] M. De Marneffe, B. MacCartney, and C. Manning. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454, 2006.

[9] H. Edmundson. New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2):264–285, 1969.

[10] C. Fellbaum. *WordNet*. Springer, 2010.

[11] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In ACL, 2005, pages 363–370.

[12] D. Gillick, K. Riedhammer, B. Favre, and D. Hakkani-Tur. A global optimization framework for meeting summarization. In ICASSP, 2009, pages 4769–4772.

[13] M. Hu and B. Liu. Mining and summarizing customer reviews. In SIGKDD, 2004, pages 168–177.

[14] J. Jagadeesh, P. Pingali, and V. Varma. A relevance-based language modeling approach to duc 2005. In HLT-EMNLP, Vancouver, Canada, 2005.

[15] L.-W. Ku, Y.-T. Liang, and H.-H. Chen. Opinion extraction, summarization and tracking in news and blog corpora. AAAI, 2006, volume 2001.

[16] J. Kupiec, J. Pedersen, and F. Chen. A trainable document summarizer. In SIGIR, 1995, pages 68–73.

[17] H. Lee, Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. CONLL : Shared Task, ACL, 2011, pages 28–34.

[18] C. Lin. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, 2004.

[19] H. Luhn. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, 1958.

[20] R. Mihalcea. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In ACL, 2004, page 20.

[21] S. Mukherjee and P. Bhattacharyya. Feature specific sentiment analysis for product reviews. CICLING, 2012, pages 475–487.

[22] V. Ng, S. Dasgupta, and S. Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In COLING-ACL, 2006, pages 611–618.

[23] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. F. Smeaton. Topic-dependent sentiment analysis of financial blogs. In CIKM, 2009, pages 9–16.

[24] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. ACL, 2004, page 271.

[25] D. Radev, H. Jing, M. Styś, and D. Tam. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938, 2004.

[26] G. Salton, J. Allan, C. Buckley, and A. Singhal. Automatic analysis, theme generation, and summarization of machine-readable texts. In *Information retrieval and hypertext*, pages 51–73. Springer, 1996.

[27] S. Somasundaran and J. Wiebe. Recognizing stances in online debates. In ACL-IJCNLP of AFNLP, 2009, pages 226–234.

[28] S. Somasundaran and J. Wiebe. Recognizing stances in ideological on-line debates. In NAACL-HLT, 2010, pages 116–124.

[29] D. Wang and Y. Liu. A pilot study of opinion summarization in conversations. In ACL, 2011.

[30] M. Wasson. Using leading text for news summaries: Evaluation results and implications for commercial summarization applications. In COLING-ACL, 1998, pages 1364–1368.