

A Novel Two-stage Framework for Extracting Opinionated Sentences from News Articles

Pujari Rajkumar¹, Swara Desai², Niloy Ganguly¹ and Pawan Goyal¹

¹Dept. of Computer Science and Engineering,
Indian Institute of Technology Kharagpur, India – 721302

²Yahoo! India

¹rajkumarsaikorian@gmail.com, {niloy,pawang}@cse.iitkgp.ernet.in
²swara@yahoo-inc.com

Abstract

This paper presents a novel two-stage framework to extract opinionated sentences from a given news article. In the first stage, Naïve Bayes classifier by utilizing the local features assigns a score to each sentence - the score signifies the probability of the sentence to be opinionated. In the second stage, we use this prior within the HITS (Hyperlink-Induced Topic Search) schema to exploit the global structure of the article and relation between the sentences. In the HITS schema, the opinionated sentences are treated as Hubs and the facts around these opinions are treated as the Authorities. The algorithm is implemented and evaluated against a set of manually marked data. We show that using HITS significantly improves the precision over the baseline Naïve Bayes classifier. We also argue that the proposed method actually discovers the underlying structure of the article, thus extracting various opinions, grouped with supporting facts as well as other supporting opinions from the article.

1 Introduction

With the advertising based revenues becoming the main source of revenue, finding novel ways to increase focussed user engagement has become an important research topic. A typical problem faced by web publishing houses like Yahoo!, is understanding the nature of the comments posted by readers of 10⁵ articles posted at any moment on its website. A lot of users engage in discussions in the comments section of the articles. Each user has a different perspective and thus comments in that genre - this many a times, results in a situation where the discussions in the comment section wander far away from the articles topic. In order to assist users to discuss relevant points in the comments section, a possible methodology can be to generate questions from the article's content that seek user's opinions about various opinions conveyed in the article (Rokhlenko and Szpektor, 2013). It would also direct the users into thinking about a spectrum of various points that the article covers and encourage users to share their unique, personal,

daily-life experience in events relevant to the article. This would thus provide a broader view point for readers as well as perspective questions can be created thus catering to users with rich user generated content, this in turn can increase user engagement on the article pages. Generating such questions manually for huge volume of articles is very difficult. However, if one could identify the main opinionated sentences within the article, it will be much easier for an editor to generate certain questions around these. Otherwise, the sentences themselves may also serve as the points for discussion by the users.

Hence, in this paper we discuss a two-stage algorithm which picks opinionated sentences from the articles. The algorithm assumes an underlying structure for an article, that is, each opinionated sentence is supported by a few factual statements that justify the opinion. We use the HITS schema to exploit this underlying structure and pick opinionated sentences from the article.

The main contributions of this papers are as follows. First, we present a novel two-stage framework for extracting opinionated sentences from a news article. Secondly, we propose a new evaluation metric that takes into account the fact that since the amount of polarity (and thus, the number of opinionated sentences) within documents can vary a lot and thus, we should stress on the ratio of opinionated sentences in the top sentences, relative to the ratio of opinionated sentences in the article. Finally, discussions on how the proposed algorithm captures the underlying structure of the opinions and surrounding facts in a news article reveal that the algorithm does much more than just extracting opinionated sentences.

This paper has been organised as follows. Section 2 discusses related work in this field. In section 3, we discuss our two-stage model in further details. Section 4 discusses the experimental framework and the results. Further discussions on the underlying assumption behind using HITS along with error analysis are carried out in Section 5. Conclusions and future work are detailed in Section 6.

2 Related Work

Opinion mining has drawn a lot of attention in recent years. Research works have focused on mining

opinions from various information sources such as blogs (Conrad and Schilder, 2007; Harb et al., 2008), product reviews (Hu and Liu, 2004; Qadir, 2009; Dave et al., 2003), news articles (Kim and Hovy, 2006; Hu and Liu, 2006) etc. Various aspects in opinion mining have been explored over the years (Ku et al., 2006). One important dimension is to identify the opinion holders as well as opinion targets. (Lu, 2010) used dependency parser to identify the opinion holders and targets in Chinese news text. (Choi et al., 2005) use Conditional Random Fields to identify the sources of opinions from the sentences. (Kobayashi et al., 2005) propose a learning based anaphora resolution technique to extract the opinion tuple $\langle Subject, Attribute, Value \rangle$. Opinion summarization has been another important aspect (Kim et al., 2013).

A lot of research work has been done for opinion mining from product reviews where most of the text is opinion-rich. Opinion mining from news articles, however, poses its own challenges because in contrast with the product reviews, not all parts of news articles present opinions (Balahur et al., 2013) and thus finding opinionated sentences itself remains a major obstacle. Our work mainly focus on classifying a sentence in a news article as opinionated or factual. There have been works on sentiment classification (Wiebe and Riloff, 2005) but the task of finding opinionated sentences is different from finding sentiments, because sentiments mainly convey the emotions and not the opinions. There has been research on finding opinionated sentences from various information sources. Some of these works utilize a dictionary-based (Fei et al., 2012) or regular pattern based (Brun, 2012) approach to identify aspects in the sentences. (Kim and Hovy, 2006) utilize the presence of a single strong valence words as well as the total valence score of all words in a sentence to identify opinion-bearing sentences. (Zhai et al., 2011) work on finding ‘evaluative’ sentences in online discussions. They exploit the inter-relationship of aspects, evaluation words and emotion words to reinforce each other.

Thus, while ours is not the first attempt at opinion extraction from news articles, to the best of our knowledge, none of the previous works has exploited the global structure of a news article to classify a sentence as opinionated/factual. Though summarization algorithms (Erkan and Radev, 2004; Goyal et al., 2013) utilize the similarity between sentences in an article to find the important sentences, our formulation is different in that we conceptualize two different kinds of nodes in a document, as opposed to the summarization algorithms, which treat all the sentences equally.

In the next section, we describe the proposed two-stage algorithm in detail.

3 Our Approach

Figure 1 gives a flowchart of the proposed two-stage method for extracting opinionated sentences from news articles. First, each news article is pre-processed to get the dependency parse as well as the TF-IDF vector corresponding to each of the sentences present in the article. Then, various features are extracted from these sentences which are used as input to the Naïve Bayes classifier, as will be described in Section 3.1. The Naïve Bayes classifier, which corresponds to the first-stage of our method, assigns a probability score to each sentence as being an opinionated sentence. In the second stage, the entire article is viewed as a complete and directed graph with edges from every sentence to all other sentences, each edge having a weight suitably computed. Iterative HITS algorithm is applied to the sentence graph, with opinionated sentences conceptualized as hubs and factual sentences conceptualized as authorities. The two stages of our approach are detailed below.

3.1 Naïve Bayes Classifier

The Naïve Bayes classifier assigns the probability for each sentence being opinionated. The classifier is trained on 70 News articles from politics domain, sentences of which were marked by a group of annotators as being opinionated or factual. Each sentence was marked by two annotators. The inter-annotator agreement using Cohen’s kappa coefficient was found to be 0.71.

The features utilized for the classifier are detailed in Table 1. These features were adapted from those reported in (Qadir, 2009; Yu and Hatzivassiloglou, 2003). A list of positive and negative polar words, further expanded using wordnet synsets was taken from (Kim and Hovy, 2005). Stanford dependency parser (De Marneffe et al., 2006) was utilized to compute the dependencies for each sentence within the news article.

After the features are extracted from the sentences, we used the Weka implementation of Naïve Bayes to train the classifier¹.

Table 1: Features List for the Naïve Bayes Classifier

1.	Count of positive polar words
2.	Count of negative polar words
3.	Polarity of the root verb of the sentence
4.	Presence of aComp, xComp and advMod dependencies in the sentence

3.2 HITS

The Naïve Bayes classifier as discussed in Section 3.1 utilizes only the local features within a sentence. Thus, the probability that a sentence is opinionated remains

¹<http://www.cs.waikato.ac.nz/ml/weka/>

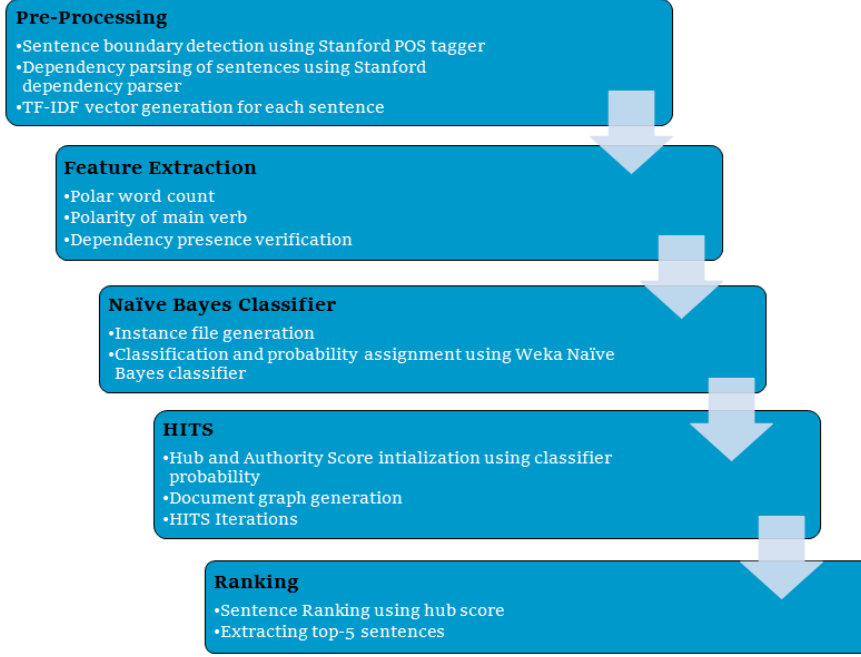


Figure 1: Flow Chart of Various Stages in Our Approach

independent of its context as well as the document structure. The main motivation behind formulating this problem in HITS schema is to utilize the hidden link structures among sentences. HITS stands for ‘Hyperlink-Induced Topic Search’; Originally, this algorithm was developed to rank Web-pages, with a particular insight that some of the webpages (**Hubs**) served as catalog of information, that could lead users directly to the other pages, which actually contained the information (**Authorities**).

The intuition behind applying HITS for the task of opinion extraction came from the following assumption about underlying structure of an article. A news article pertains to a specific theme and with that theme in mind, the author presents certain opinions. These opinions are justified with the facts present in the article itself. We conceptualize the opinionated sentences as **Hubs** and the associated facts for an opinionated sentence as **Authorities** for this **Hub**.

To describe the formulation of HITS parameters, let us give the notations. Let us denote a document D using a set of sentences $\{S_1, S_2, \dots, S_i, \dots, S_n\}$, where n corresponds to the number of sentences in the document D . We construct the sentence graph where nodes in the graph correspond to the sentences in the document. Let H_i and A_i denote the hub and authority scores for sentence S_i . In HITS, the edges always flow from a Hub to an Authority. In the original HITS algorithm, each edge is given the same weight. However, it has been reported that using weights in HITS update improves the performance significantly (Li et al., 2002). In our formulation, since each node has a non-zero probability of acting

as a hub as well as an authority, we have outgoing as well as incoming edges for every node. Therefore, the weights are assigned, keeping in mind the proximity between sentences as well as the probability (of being opinionated/factual) assigned by the classifier. The following criteria were used for deciding the weight function.

- An edge in the HITS graph goes from a hub (source node) to an authority (target node). So, the edge weight from a source node to a target node should be higher if the source node has a high hub score.
- A fact corresponding to an opinionated sentence should be discussing the same topic. So, the edge weight should be higher if the sentences are more similar.
- It is more probable that the facts around an opinion appear closer to that opinionated sentence in the article. So, the edge weight from a source to target node decreases as the distance between the two sentences increases.

Let W be the weight matrix such that W_{ij} denotes the weight for the edge from the sentence S_i to the sentence S_j . Based on the criteria outlined above, we formulate that the weight W_{ij} should be such that

$$\begin{aligned}
 W_{ij} &\propto H_i \\
 W_{ij} &\propto Sim_{ij} \\
 W_{ij} &\propto \frac{1}{dist_{ij}}
 \end{aligned}$$

where we use *cosine similarity* between the sentence vectors to compute Sim_{ij} . $dist_{ij}$ is simply the number

of sentences separating the source and target node. Various combinations of these factors were tried and will be discussed in section 4. While factors like sentence similarity and distance are symmetric, having the weight function depend on the hub score makes it asymmetric, consistent with the basic idea of HITS. Thus, an edge from the sentence S_i to S_j is given a high weight if S_i has a high probability score of being opinionated (i.e., acting as hub) as obtained the classifier.

Now, for applying the HITS algorithm iteratively, the Hubs and Authorities scores for each sentence are initialized using the probability scores assigned by the classifier. That is, if $P_i(\textit{Opinion})$ denotes the probability that S_i is an opinionated sentence as per the Naïve Bayes Classifier, $H_i(0)$ is initialized to $P_i(\textit{Opinion})$ and $A_i(0)$ is initialized to $1 - P_i(\textit{Opinion})$. The iterative HITS is then applied as follows:

$$H_i(k) = \sum_j W_{ij} A_j(k-1) \quad (1)$$

$$A_i(k) = \sum_j W_{ji} H_j(k-1) \quad (2)$$

where $H_i(k)$ denote the hub score for the i^{th} sentence during the k^{th} iteration of HITS. The iteration is stopped once the mean squared error between the Hub and Authority values at two different iterations is less than a threshold ϵ . After the HITS iteration is over, five sentences having the highest Hub scores are returned by the system.

4 Experimental Framework and Results

The experiment was conducted with 90 news articles in politics domain from Yahoo! website. The sentences in the articles were marked as opinionated or factual by a group of annotators. In the training set, 1393 out of 3142 sentences were found to be opinionated. In the test set, 347 out of 830 sentences were marked as opinionated. Out of these 90 articles, 70 articles were used for training the Naïve Bayes classifier as well as for tuning various parameters. The rest 20 articles were used for testing. The evaluation was done in an Information Retrieval setting. That is, the system returns the sentences in a decreasing order of their score (or probability in the case of Naïve Bayes) as being opinionated. We then utilize the human judgements (provided by the annotators) to compute precision at various points. Let $op(\cdot)$ be a binary function for a given rank such that $op(r) = 1$ if the sentence returned as rank r is opinionated as per the human judgements.

A $P@k$ precision is calculated as follows:

$$P@k = \frac{\sum_{r=1}^k op(r)}{k} \quad (3)$$

While the precision at various points indicates how reliable the results returned by the system are, it does not take into account the fact that some of the

documents are opinion-rich and some are not. For the opinion-rich documents, a high $P@k$ value might be similar to picking sentences randomly, whereas for the documents with a very few opinions, even a lower $P@k$ value might be useful. We, therefore, devise another evaluation metric $M@k$ that indicates the ratio of opinionated sentences at any point, normalized with respect to the ratio of opinionated sentences in the article.

Correspondingly, an $M@k$ value is calculated as

$$M@k = \frac{P@k}{Ratio_{op}} \quad (4)$$

where $Ratio_{op}$ denotes the fraction of opinionated sentences in the whole article. Thus

$$Ratio_{op} = \frac{\text{Number of opinionated sentences}}{\text{Number of sentences}} \quad (5)$$

The parameters that we needed to fix for the HITS algorithm were the weight function W_{ij} and the threshold ϵ at which we stop the iteration. We varied ϵ from 0.0001 to 0.1 multiplying it by 10 in each step. The results were not sensitive to the value of ϵ and we used $\epsilon = 0.01$. For fixing the weight function, we tried out various combinations using the criteria outlined in Section 3.2. Various weight functions and the corresponding $P@5$ and $M@5$ scores are shown in Table 2. Firstly, we varied k in Sim_{ij}^k and found that the square of the similarity function gives better results. Then, keeping it constant, we varied l in H_i^l and found the best results for $l = 3$. Then, keeping both of these constants, we varied α in $(\alpha + \frac{1}{d})$. We found the best results for $\alpha = 1.0$. With this α , we tried to vary l again but it only reduced the final score. Therefore, we fixed the weight function to be

$$W_{ij} = H_i^3(0) Sim_{ij}^2 (1 + \frac{1}{dist_{ij}}) \quad (6)$$

Note that $H_i(0)$ in Equation 6 corresponds to the probability assigned by the classifier that the sentence S_i is opinionated.

We use the classifier results as the baseline for the comparisons. The second-stage HITS algorithm is then applied and we compare the performance with respect to the classifier. Table 3 shows the comparison results for various precision scores for the classifier and the HITS algorithm. In practical situation, an editor requires quick identification of 3-5 opinionated sentences from the article, which she can then use to formulate questions. We thus report $P@k$ and $M@k$ values for $k = 3$ and $k = 5$.

From the results shown in Table 3, it is clear that applying the second-stage HITS over the Naïve Bayes Classifier improves the performance by a large degree, both in term of $P@k$ and $M@k$. For instance, the first-stage NB Classifier gives a $P@5$ of 0.52 and $P@3$ of 0.53. Using the classifier outputs during the second-stage HITS algorithm improves the

Table 2: Average $P@5$ and $M@5$ scores: Performance comparison between various functions for W_{ij}

Function	$P@5$	$M@5$
Sim_{ij}	0.48	0.94
Sim_{ij}^2	0.57	1.16
Sim_{ij}^3	0.53	1.11
$Sim_{ij}^2 H_i$	0.6	1.22
$Sim_{ij}^2 H_i^2$	0.61	1.27
$Sim_{ij}^2 H_i^3$	0.61	1.27
$Sim_{ij}^2 H_i^4$	0.58	1.21
$Sim_{ij}^2 H_i^3 \frac{1}{d}$	0.56	1.20
$Sim_{ij}^2 H_i^3 (0.2 + \frac{1}{d})$	0.60	1.25
$Sim_{ij}^2 H_i^3 (0.4 + \frac{1}{d})$	0.61	1.27
$Sim_{ij}^2 H_i^3 (0.6 + \frac{1}{d})$	0.62	1.31
$Sim_{ij}^2 H_i^3 (0.8 + \frac{1}{d})$	0.62	1.31
$Sim_{ij}^2 H_i^3 (1 + \frac{1}{d})$	0.63	1.33
$Sim_{ij}^2 H_i^3 (1.2 + \frac{1}{d})$	0.61	1.28
$Sim_{ij}^2 H_i^2 (1 + \frac{1}{d})$	0.6	1.23

Table 3: Average $P@5$, $M@5$, $P@3$ and $M@3$ scores: Performance comparison between the NB classifier and HITS

System	$P@5$	$M@5$	$P@3$	$M@3$
NB Classifier	0.52	1.13	0.53	1.17
HITS	0.63	1.33	0.72	1.53
Imp. (%)	+21.2	+17.7	+35.8	+30.8

performance by 21.2% to 0.63 in the case of $P@5$. For $P@3$, the improvements were much more significant and a 35.8% improvement was obtained over the NB classifier. $M@5$ and $M@3$ scores also improve by 17.7% and 30.8% respectively.

Strikingly, while the classifier gave nearly the same scores for $P@k$ and $M@k$ for $k = 3$ and $k = 5$, HITS gave much better results for $k = 3$ than $k = 5$. Specially, the $P@3$ and $M@3$ scores obtained by HITS were very encouraging, indicating that the proposed approach helps in pushing the opinionated sentences to the top. This clearly shows the advantage of using the global structure of the document in contrast with the features extracted from the sentence itself, ignoring the context.

Figures 2 and 3 show the $P@5$, $M@5$, $P@3$ and $M@3$ scores for individual documents as numbered from 1 to 20 on the X-axis. The articles are sorted as per the ratio of $P@5$ (and $M@5$) obtained using the HITS and NB classifier. Y-axis shows the corresponding scores. Two different lines are used to represent the results as returned by the classifier and the HITS algorithm. A dashed line denotes the scores obtained by HITS while a continuous line denotes the scores obtained by the NB classifier. A detailed analysis of these figures can help us draw the following conclusions:

- For 40% of the articles (numbered 13 to 20) HITS improves over the baseline NB classifier. For

40% of the articles (numbered 5 to 12) the results provided by HITS were the same as that of the baseline. For 20% of the articles (numbered 1 to 4) HITS gives a performance lower than that of the baseline. Thus, for 80% of the documents, the second-stage performs at least as good as the first stage. This indicates that the second-stage HITS is quite robust.

- $M@5$ results are much more robust for the HITS, with 75% of the documents having an $M@5$ score > 1 . An $M@k$ score > 1 indicates that the ratio of opinionated sentences in top k sentences, picked up by the algorithm, is higher than the overall ratio in the article.
- For 45% of the articles, (numbered 6, 9 – 11 and 15 – 20), HITS was able to achieve a $P@3 = 1.0$. Thus, for these 9 articles, the top 3 sentences picked up by the algorithm were all marked as opinionated.

The graphs also indicate a high correlation between the results obtained by the NB classifier and HITS. We used Pearson’s correlation to find the correlation strength. For the $P@5$ values, the correlation was found to be 0.6021 and for the $M@5$ values, the correlation was obtained as 0.5954.

In the next section, we will first attempt to further analyze the basic assumption behind using HITS, by looking at some actual Hub-Authority structures, captured by the algorithm. We will also take some cases of failure and perform error analysis.

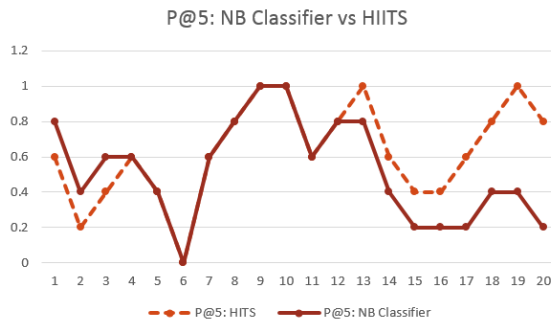
5 Discussion

First point that we wanted to verify was, whether HITS is really capturing the underlying structure of the document. That is, are the sentences identified as authorities for a given hub really correspond to the facts supporting the particular opinion, expressed by the hub sentence.

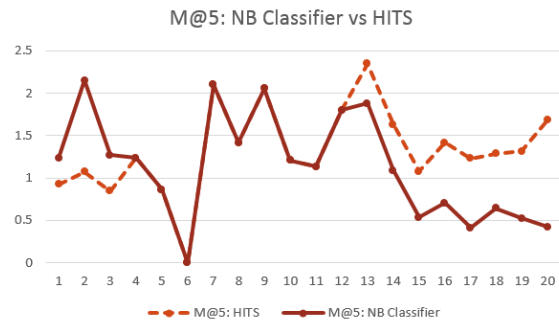
Figure 4 gives two examples of the Hub-Authority structure, as captured by the HITS algorithm, for two different articles. For each of these examples, we show the sentence identified as Hub in the center along with the top four sentences, identified as Authorities for that hub. We also give the annotations as to whether the sentences were marked as ‘opinionated’ or ‘factual’ by the annotators.

In both of these examples, the hubs were actually marked as ‘opinionated’ by the annotators. Additionally, we find that all the four sentences, identified as authorities to the hub, are very relevant to the opinion expressed by the hub. In the first example, top 3 authority sentences are marked as ‘factual’ by the annotator. Although the fourth sentence is marked as ‘opinionated’, it can be seen that this sentence presents a supporting opinion for the hub sentence.

While studying the second example, we found that while the first authority does not present an important fact, the fourth authority surely does. Both of these

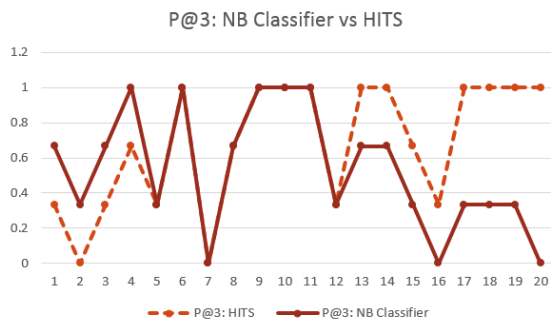


(a) Comparison of P@5 values

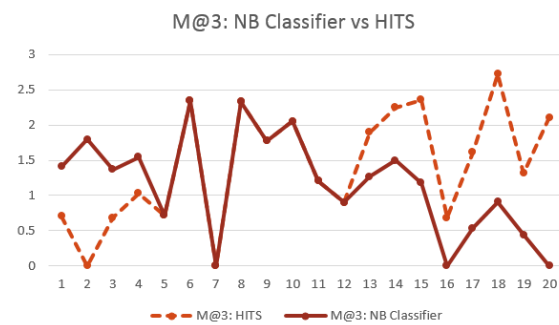


(b) Comparison of M@5 values

Figure 2: Comparison Results for 20 Test articles between the Classifier and HIITS: P@5 and M@5

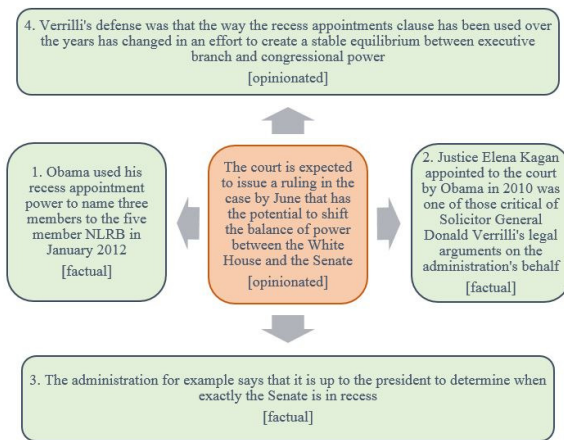


(a) Comparison of P@3 values

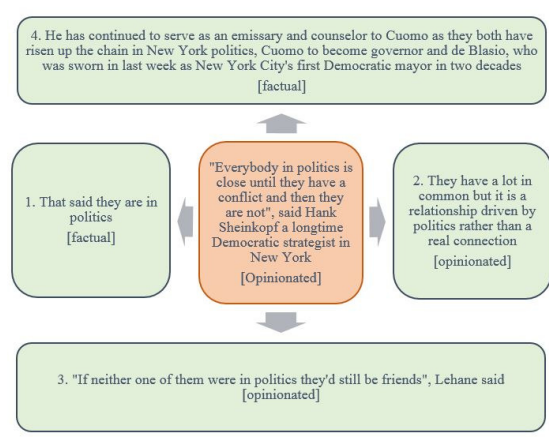


(b) Comparison of M@3 values

Figure 3: Comparison Results for 20 Test articles between the Classifier and HIITS: P@3 and M@3



(a) Hub-Authority Structure: Example 1



(b) Hub-Authority Structure: Example 2

Figure 4: Example from two different test articles capturing the Hub-Authority Structure

were marked as ‘factual’ by the annotators. In this particular example, although the second and third authority sentences were annotated as ‘opinionated’, these can be seen as supporting the opinion expressed by the hub sentence. This example also gives us an interesting idea to improve diversification in the final results. That is, once an opinionated sentence is identified by the algorithm, the hub score of all its authorities can be reduced proportional to the edge

weight. This will reduce the chances of the supporting opinions being returned by the system, at a later stage as a main opinion.

We then attempted to test our tool on a recently published article, “**What’s Wrong with a Meritocracy Rug?**”². The tool could pick up a very

²<http://news.yahoo.com/whats-wrong-meritocracy-rug-070000354.html>

important opinion in the article, “*Most people tend to think that the most qualified person is someone who looks just like them, only younger.*”, which was ranked 2nd by the system. The supporting facts and opinions for this sentence, as discovered by the algorithm were also quite relevant. For instance, the top two authorities corresponding to this sentence hub were:

1. *And that appreciation, we learned painfully, can easily be tinged with all kinds of gendered elements without the person who is making the decisions even realizing it.*
2. *And many of the traits we value, and how we value them, also end up being laden with gender overtones.*

5.1 Error Analysis

We then tried to analyze certain cases of failures. Firstly, we wanted to understand why HITS was not performing as good as the classifier for 3 articles (Figures 2 and 3). The analysis revealed that the supporting sentences for the opinionated sentences, extracted by the classifier, were not very similar on the textual level. Thus a low cosine similarity score resulted in having lower edge weights, thereby getting a lower hub score after applying HITS. For one of the articles, the sentence picked up by HITS was wrongly annotated as a factual sentence.

Then, we looked at one case of failure due to the error introduced by the classifier prior probabilities. For instance, the sentence, “*The civil war between establishment and tea party Republicans **intensified** this week when House Speaker John Boehner slammed outside **conservative** groups for **ridiculous** pushback against the bipartisan budget agreement which cleared his chamber Thursday.*” was classified as an opinionated sentence, whereas this is a factual sentence. Looking closely, we found that the sentence contains three polar words (marked in bold), as well as an *advMod* dependency between the pair (slammed,when). Thus the sentence got a high initial prior by the classifier. As a result, the outgoing edges from this node got a higher H_i^3 factor. Some of the authorities identified for this sentence were:

- *For Democrats, the tea party is the gift that keeps on giving.*
- *Tea party sympathetic organizations, Boehner later said, “are pushing our members in places where they don’t want to be”.*

which had words, similar to the original sentence, thus having a higher Sim_{ij} factor as well. We found that these sentences were also very close within the article. Thus, a high hub prior along with a high outgoing weight gave rise to this sentence having a high hub score after the HITS iterations.

5.2 Online Interface

To facilitate easy usage and understanding of the system by others, a web interface has been built for

the system³. The webpage caters for users to either input a new article in form of text to get top opinionated sentences or view the output analysis of the system over manually marked test data consisting of 20 articles.

The words in green color are positive polar words, red indicates negative polar words. Words marked in violet are the root verbs of the sentences. The colored graph shows top ranked opinionated sentences in yellow box along with top supporting factual sentences for that particular opinionated sentence in purple boxes. Snapshots from the online interface are provided in Figures 5 and 6.

6 Conclusions and Future Work

In this paper, we presented a novel two-stage framework for extracting the opinionated sentences in the news articles. The problem of identifying top opinionated sentences from news articles is very challenging, especially because the opinions are not as explicit in a news article as in a discussion forum. It was also evident from the inter-annotator agreement and the kappa coefficient was found to be 0.71.

The experiments conducted over 90 News articles (70 for training and 20 for testing) clearly indicate that the proposed two-stage method almost always improves the performance of the baseline classifier-based approach. Specifically, the improvements are much higher for $P@3$ and $M@3$ scores (35.8% and 30.8% over the NB classifier). An $M@3$ score of 1.5 and $P@3$ score of 0.72 indicates that the proposed method was able to push the opinionated sentences to the top. On an average, 2 out of top 3 sentences returned by the system were actually opinionated. This is very much desired in a practical scenario, where an editor requires quick identification of 3-5 opinionated sentences, which she can then use to formulate questions.

The examples discussed in Section 5 bring out another important aspect of the proposed algorithm. In addition to the main objective of extracting the opinionated sentences within the article, the proposed method actually discovers the underlying structure of the article and would certainly be useful to present various opinions, grouped with supporting facts as well as supporting opinions in the article.

While the initial results are encouraging, there is scope for improvement. We saw that the results obtained via HITS were highly correlated with the Naïve Bayes classifier results, which were used in assigning a weight to the document graph. One direction for the future work would be to experiment with other features to improve the precision of the classifier. Additionally, in the current evaluation, we are not evaluating the degree of diversity of the opinions returned by the system. The Hub-Authority

³available at <http://cse.iitkgp.ac.in/resgrp/cnerg/temp2/final.php>

Opinion Sentence Finder

Please insert text from any article to see the most opinionated sentences in that article

Article:

Pick Opinionated Sentences

Marked Data

Article 1

Show Output Analysis

Top Opinionated Sentences picked by the classifier are:

[Factual] The president **fielded** questions a few hours after the government announced the economy grew at a **solid** 4.1 percent annual rate from July through September the **fastest** pace since late 2011 and significantly higher than previously believed

[Opinionated] If you're **measuring** this by polls my polls have gone up and down a lot over the course of my career he said and then repeated that the economy was finally showing **significant progress**

[Opinionated] The rollout of his health care website bombed and high visibility parts of his agenda **have** yet to make it through Congress including a call for gun safety legislation in the wake of the shooting at a Newtown Conn. elementary school a year ago and a **sweeping** overhaul of immigration laws

[Opinionated] But it's also **fair** to say we're not **condemned** to endless gridlock he said

[Factual] A presidential advisory panel this week **recommended sweeping** changes to government surveillance including limiting the bulk collection of Americans' phone records by stripping the NSA of its ability to store the data in its own facilities

Figure 5: Screenshot from the Web Interface

Top opinionated sentences picked up after applying HITS algorithm:

1.

b. It's probably too early to declare an outbreak of bipartisanship	[Opinionated] But it's also fair to say we're not condemned to endless gridlock he said	c. Obama did not mention it but the stock market is also at or near record levels
	d. It's a responsibility of Congress he said although he added that he was willing to discuss other issues separately	

2.

	a. Yet he suggested that given widespread criticism he may alter the power of the National Security Agency to collect information on Americans	
b. I have confidence that the NSA is not engaged in domestic surveillance or snooping around he said	[Opinionated] Yet he added we may have to refine this further to give people more confidence	c. As for health care Obama said that despite the problems including the rollout of the website more than 2 million people have signed up or more since enrollment began
	d. It's a responsibility of Congress he said although he added that he was willing to discuss other issues separately	

Figure 6: Hub-Authority Structure as output on the Web Interface

structure of the second example gives us an interesting idea to improve diversification and we would like to implement that in future.

In the future, we would also like to apply this work to track an event over time, based on the opinionated sentences present in the articles. When an event occurs, articles start out with more factual sentences. Over time, opinions start surfacing on the event, and as the event matures, opinions predominate the facts in the articles. For example, a set of articles on a plane crash would start out as factual, and would offer expert opinions over time. This work can be used to plot the maturity of the media coverage by keeping track of facts v/s opinions on any event, and this can be used by organizations to provide a timeline for the event. We would also like to experiment with this model on

a different media like microblogs.

References

- Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.
- Caroline Brun. 2012. Learning opinionated patterns for contextual opinion detection. In *COLING (Posters)*, pages 165–174.
- Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction

- patterns. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 355–362. Association for Computational Linguistics.
- Jack G Conrad and Frank Schilder. 2007. Opinion mining in legal blogs. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 231–236. ACM.
- Kushal Dave, Steve Lawrence, and David M Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528. ACM.
- Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454.
- Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res.(JAIR)*, 22(1):457–479.
- Geli Fei, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2012. A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In *Proceedings of COLING 2012 (Posters)*.
- Pawan Goyal, Laxmidhar Behera, and Thomas Martin McGinnity. 2013. A context-based word indexing model for document summarization. *Knowledge and Data Engineering, IEEE Transactions on*, 25(8):1693–1705.
- Ali Harb, Michel Plantié, Gerard Dray, Mathieu Roche, François Troussel, and Pascal Poncelet. 2008. Web opinion mining: How to extract opinions from blogs? In *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*, pages 211–217. ACM.
- Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *Proceedings of Nineteenth National Conference on Artificial Intelligence (AAAI)*.
- Minqing Hu and Bing Liu. 2006. Opinion extraction and summarization on the web. In *AAAI*, volume 7, pages 1621–1624.
- Soo-Min Kim and Eduard Hovy. 2005. Automatic detection of opinion bearing words and sentences. In *Proceedings of IJCNLP*, volume 5.
- Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8. Association for Computational Linguistics.
- Hyun Duk Kim, Malu Castellanos, Meichun Hsu, ChengXiang Zhai, Umeshwar Dayal, and Riddhiman Ghosh. 2013. Compact explanatory opinion summarization. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1697–1702. ACM.
- Nozomi Kobayashi, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2005. Opinion extraction using a learning-based anaphora resolution technique. In *The Second International Joint Conference on Natural Language Processing (IJCNLP), Companion Volume to the Proceeding of Conference including Posters/Demos and Tutorial Abstracts*.
- Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, volume 100107.
- Longzhuang Li, Yi Shang, and Wei Zhang. 2002. Improvement of hits-based algorithms on web documents. In *Proceedings of the 11th international conference on World Wide Web*, pages 527–535. ACM.
- Bin Lu. 2010. Identifying opinion holders and targets with dependency parser in chinese news texts. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*.
- Ashequl Qadir. 2009. Detecting opinion sentences specific to product features in customer reviews using typed dependency relations. In *Proceedings of the Workshop on Events in Emerging Text Types, eETTs '09*, pages 38–43.
- Oleg Rokhlenko and Idan Szpektor. 2013. Generating synthetic comparable questions for news articles. In *ACL*, pages 742–751.
- Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Computational Linguistics and Intelligent Text Processing*, pages 486–497. Springer.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP '03*, pages 129–136.
- Zhongwu Zhai, Bing Liu, Lei Zhang, Hua Xu, and Peifa Jia. 2011. Identifying evaluative sentences in online discussions. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*.