I want what I need! Analyzing Subjectivity of Online Forum Threads

Prakhar Biyani
College of Information Sciences and Technology, The Pennsylvania State University
prakharbiyani@gmail.com

Cornelia Caragea
College of Information Sciences and Technology, The Pennsylvania State University
ccaragea@ist.psu.edu

Amit Kumar Singh
IBM Research, Bangalore
amitkumarsingh@in.ibm.com

Prasenjit Mitra
College of Information Sciences and Technology, The Pennsylvania State University
pmitra@ist.psu.edu

ABSTRACT

Online forums have become a popular source of information due to the unique nature of information they contain. Internet users use these forums to get opinions of other people on issues and to find factual answers to specific questions. Topics discussed in online forum threads can be subjective seeking personal opinions or non-subjective seeking factual information. Hence, knowing subjectivity orientation of threads would help forum search engines to satisfy user's information needs more effectively by matching the subjectivities of user's query and topics discussed in the threads in addition to lexical match between the two. We study methods to analyze the subjectivity of online forum threads. Experimental results on a popular online forum demonstrate the effectiveness of our methods.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services – Web-based services

General Terms

Human factors, Algorithms, Experimentation

Keywords

Online forums, Subjectivity.

1. INTRODUCTION

Online forums contain huge amounts of discussions between Internet users on domain-specific problems such as camera, operating systems, notebooks, traveling, health, etc., as well as discussions related to daily life experiences. Such information is difficult to find in other online sources (e.g., product manuals, Wikipedia, etc.) and, hence, these forums are increasingly becoming popular among Internet users. Since these forums contain massive volume of information, finding relevant information often becomes challenging for users.

Internet users search online forums, generally, for two types of information. Some search the forums for subjective information like discussions on a certain topic to educate themselves about multiple points of view related to the topic, people's opinions, emotions, evaluations, etc. Others pose queries that are objective and have short factual answers. Previous works on online forum search have focused on improving lexical match between searcher's query keywords and thread content [13, 1, 2]. However, these works do not take into account a searcher's intent, i.e., the *type of* information a searcher wants. We explain this point with the following example. Consider the two queries issued by a searcher to some camera forum search engine: 1. "How is the resolution of canon 7D?", and 2. "What is the resolution of canon 7D?". Both queries are about the resolution of canon 7D (and may look similar at first sight), but the searcher's intent is different across the two queries. In the first query, he seeks *opinions* of different camera users on the resolution of the Canon 7D camera, i.e., how different users *feel* about the resolution, what is their experience (good, bad, excellent, etc.) with Canon 7D as far as the resolution is concerned; hence the query is *subjective* in nature. In the second query, the searcher does not seek opinions, but a *factual* answer to a specific question, which, in this case, is the value of the resolution, and therefore, the query is *objective* in nature. Hence, queries having similar keywords may differ in their intents. Search algorithms based only on keyword search would perform badly for these types of queries. In order to answer these types of queries effectively, forum search engines need to first identify the type of information a searcher wants, as well as the type of information a document unit (usually a thread) contains, and then match the two in addition to their lexical match.

The current work addresses a part of this problem. We propose methods to identify the type of information a forum thread contains. To the best of our knowledge, this problem has not been addressed by any previous work on subjectivity analysis and on online forums. We identify two types of threads in an online forum: *subjective* and *non-subjective*. Subjective threads discuss subjective topics that seek personal opinions, viewpoints, evaluations, and other private states of people, whereas non-subjective threads discuss non-subjective topics that seek factual information. Figure 1 shows a subjective thread from an online forum, Trip–Advisor New York. Figure 2 shows a non-subjective thread from the same forum. In the former, the topic of discussion is *whether to tip or not after bad service?*, which seeks opinions, whereas the latter seeks factual information about *bands/artists playing in December in Madison Square Gardens*.

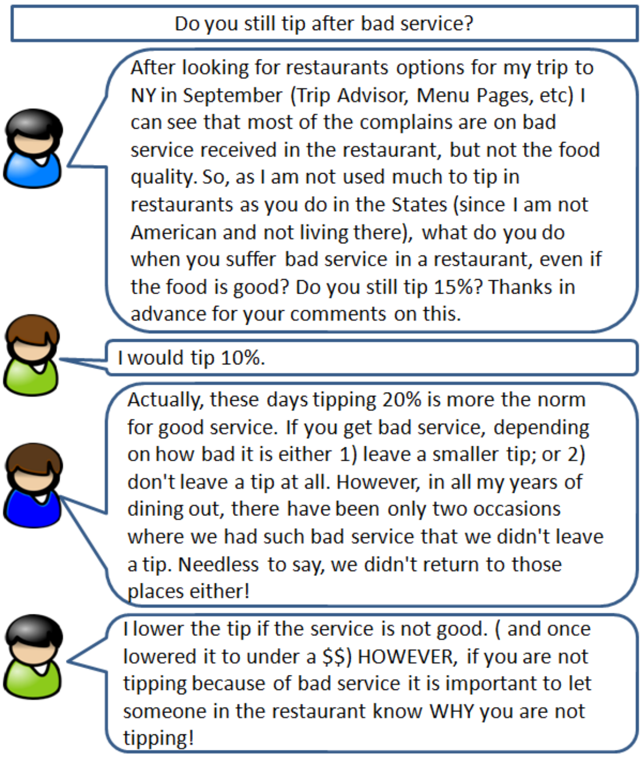

Figure 1: An example thread with subjective topic.

Figure 2: An example thread with non-subjective topic.

We propose a simple and effective classification method using textual features obtained from online forum threads to identify the two types of threads. We model the task as a binary classification of threads in one of the two classes: ***Subjective*** and ***Non-subjective***. We use combinations of words and their parts-of-speech tags as features. The features are generated from different structural units of a thread such as title, initial post, reply posts and their combinations. We performed experiments on a popular online forum, Trip Advisor–New York.

Our contribution. Our work is the first to perform subjectivity analysis of online forum threads automatically. We believe that retrieval models can improve ranking functions by incorporating subjectivity match between users' queries and threads. We show that simple features generated from n-grams and parts-of-speech tags work reasonably well for identifying subjective and non-subjective discussion threads in online forums.

The rest of the paper is organized as follows: The next section overviews the related work in the field of subjectivity analysis. Section 3 describes the features used for classification. Section 4 gives details of data collection and annotation. We describe our experimental setting and present the results in Section 5. Section 6 concludes the paper and discusses the future work.

2. RELATED WORK

Subjectivity analysis has received a lot of attention in the recent literature. For example, subjectivity analysis of sentences has been widely researched in the field of Sentiment Analysis [6, 11, 8, 9]. An integral part of sentiment analysis is to separate opinionated (generally subjective) sentences from un-opinionated (non-subjective) sentences [11] by classifying sentences as subjective or non-subjective and then sentiments in the opinionated sentences are classified as positive or negative. Finally, a summary of sentiments is generated [8]. Previous works in this field have mainly focused on online product reviews sites where the aim is to summarize product reviews given by the users [6, 9]. In contrast, our work aims at predicting subjectivity orientation of online forum threads for use in improving their retrieval. In sentiment analysis, only subjective sentences are of interest because sentiments are generally expressed in subjective language whereas in our application, a user's query governs the interest, i.e., threads having similar subjectivity orientation (subjective or non-subjective) as that of a user's query are of interest.

In the domain of online forums, there have been two recent works that are close to our work. Hassan *et al.,* [5] performed sentence-level attitude classification in online discussions to model user interaction that may be helpful in facilitating collaborations. Zhai *et al.,* [18] classified sentences in online discussions as evaluative or non-evaluative for getting relevant opinion sentences. In contrast, our work does thread-level subjectivity classification as we are more interested in the overall topic of discussion of the thread and not in the fine level details in individual sentences.

Other recent works have used subjectivity analysis to improve question-answering in social media [7, 3, 15, 17, 14]. For example, Stoyanov *et al.,* [15] identify opinions and facts in questions and answers to make multi-perspective question-answering more effective. They showed that answers to opinion questions have different properties than answers to factual questions, e.g., opinion answers were approximately twice as long as fact answers. They used these differences to filter factual answers for opinion questions thereby improving answer retrieval for opinion questions. Somasundaran *et al.,* [14] recognized two types of attitudes in opinion sentences: sentiment and arguing and used it to improve answering of attitude questions by matching the attitude type of the questions and answers in multi-perspective QA. Li *et.al.* [7] used classification to identify subjectivity orientation of questions in community QA. Gurevych *et.al.* [3] used an unsupervised lexicon based approach to classify questions as subjective or factoid (non-subjective). They manually extracted patterns of words that are indicative of subjectivity from annotated questions and scored test questions based on the number of patterns present in them. These works analyzed the subjectivity of questions and answers that are usually given by *single authors in community sites*. In contrast, we analyze the subjectivity of *online forum threads that contain replies from multiple authors.*

3. FEATURE GENERATION

We assume that threads discussing subjective topics would contain more subjective sentences in contrast to the threads discussing non-subjective topics. A subjective sentence exposes private states of mind such as emotions, opinions, evaluations, speculations, etc. of the author whereas a non-subjective (objective) sentence contains factual materials [16]. This difference usually results in different vocabulary and grammatical structure of these two types of sentences [12]. To capture these differences, we used words, parts-of-speech tags and their combinations as the features for classification. We used the *Lingua-en-tagger* package from CPAN[1] for part-of-speech tagging. The following features were extracted for a sentence in different structural units (title, initial post, reply posts, etc.) of a thread:

(a) **Bag of Words (BoW)**: all words of a sentence.

(b) **Unigrams + POS tags (BoW+POS)**: all words of a sentence and their parts-of-speech tags.

(c) **Unigrams + bigrams (BoW+Bi)**: all words and sequences of 2 consecutive words in a sentence.

(d) **Unigrams + bigrams + POS tags (BoW+Bi+POS)**: all words, their parts-of-speech tags and sequences of 2 consecutive words in a sentence.

Feature type	Generated feature
Bag of Words (BoW)	W_i, W_{i+1}, W_{i+2}
Unigrams + POS tags (BoW+POS)	$W_i, POS_i, W_{i+1}, POS_{i+1}, W_{i+2}, POS_{i+2}$
Unigrams + Bigrams (BoW+Bi)	$W_i, W_{i+1}, W_{i+1}, W_iW_{i+1}, W_{i+1}W_{i+2}$
Unigrams + Bigrams + POS tags (BoW+Bi+POS)	$W_i, POS_i, W_{i+1}, POS_{i+1}, W_{i+2}, POS_{i+2}, W_iW_{i+1}, W_iPOS_{i+1}, POS_iW_{i+1}, W_{i+1}W_{i+2}, W_{i+1}POS_{i+2}, POS_{i+1}W_{i+2}$

Table 1: Feature Generation for sentence $W_i\ W_{i+1}\ W_{i+2}$.

Table 1 describes feature generation on a sentence containing three words W_i, W_{i+1} and W_{i+2}. POS_i, POS_{i+1} and POS_{i+2} are the parts-of-speech tags for the words W_i, W_{i+1} and W_{i+2}, respectively. For feature representation, we used term frequency as the weighting scheme (we empirically found it to be more effective than *tf-idf* and *binary* representations), and used minimum document frequency for a term as 3 (we experimented with minimum document frequency 3, 5 and 10 and 3 gave the best results).

[1]http://search.cpan.org/dist/Lingua-EN-Tagger/Tagger.pm

4. DATA

To evaluate our approach, we used threads from a popular online travel forum: **Trip Advisor–New York** that contains travel related discussions mainly for New York city [2]. We used a publicly available dataset [3] [1] that contains 83072 threads from which we randomly selected 700 threads for our experiments.

We hired two human annotators for tagging the threads. The annotators were asked to tag a thread as subjective if its topic of discussion is subjective or non-subjective if the topic of discussion is non-subjective.The annotators were provided with a set of instructions for annotations. The set contained definitions of subjective and non-subjective topics with examples and guidelines for doing annotations. The instruction set and the tagged dataset can be downloaded from the authors' website. [4]

The overall percentage agreement between the annotators and Kappa value were 87% and 0.713. For our experiments, we used the data on which the annotators agreed. There were 412 subjective and 197 non-subjective threads which indicates that online forum users tend to discuss subjective topics more.

Total # threads	609
Total # posts	6591
Total # users	1206
Average thread length (in terms of # posts)	10.82
Average thread length (in terms of # words)	907
Average # users in a thread	1.98

Table 2: Statistics of the Dataset

5. EXPERIMENTS AND RESULTS

In this section, we describe our experimental setting and report the results.

5.1 Experimental Setting

We used a Multinomial Naive Bayes classifier [10] for classification because it performs well on word features (We also experimented with Support Vector Machines (SVM), Logistic Classifiers, Bagging and Boosting with tf, tf-idf and binary as the feature encoding schemes and found that Naive Bayes outperformed all the others except SVM where the performances were almost equal). We used 5-fold cross validation to evaluate the performance of our classifiers and Weka data mining toolkit [4] with default settings to conduct our experiments.

As described in Section 3, we conducted experiments with four kinds of features: (i) bag of words (BoW), (ii) unigrams and POS tags (BoW+POS), (iii) unigrams and bigrams (BoW+Bi), (iv) unigrams, bigrams and POS tags (Bow+Bi+POS) extracted from the textual content of different structural units (title, initial post, reply posts) of the threads. First, we built a basic model where we used only the text of the titles (denoted by t) for classification. Then, we used the text of initial posts and reply posts. We experimented with the following four settings: title (t), initial post (I), title and initial post (t+I), entire thread (t+I+R).

[2]http://www.tripadvisor.com/ShowForum-g60763-i5-New_York_City_New_York.html
[3]http://www.cse.psu.edu/ sub194/datasets/ForumData.tar.gz
[4]www.personal.psu.edu/pxb5080

	BoW			BoW+POS			BoW+Bi			BoW+Bi+POS		
	Pr.	Re.	F-1	Pr.	Re.	F-1	Pr.	Re.	F-1	Pr.	Re.	F-1
t	0.618	0.644	0.625	0.626	0.647	0.633	0.606	0.631	0.614	0.606	0.616	0.610
I	0.662	0.665	0.664	0.669	0.673	0.671	0.713	0.718	0.715^{α}	0.701	0.711	0.704^{α}
t+I	0.671	0.673	0.672	0.686	0.69	0.688^{α}	0.700	0.704	0.702^{α}	0.701	0.709	0.704^{α}
t+I+R	0.703	0.716	**0.706**$^{\alpha}$	0.701	0.713	**0.704**$^{\alpha}$	0.738	0.747	**0.723**$^{\alpha}$	0.733	0.741	**0.71**$^{\alpha}$

Table 3: Classification performance of different features extracted from different structural components of the forum threads. t, I and R are title, initial post and set of all reply posts of a thread respectively. BoW, BoW+POS, BoW+Bi and BoW+Bi+POS are the different kinds of features explained in table 1. Superscript α indicates statistical significance over title (t) with a significance level of 0.05 (paired one-sided t-test).

5.2 Results

Table 3 reports the results of the thread classification. We report macro averaged (the weighted average of a metric for subjective class and non-subjective class) precision, recall and F-1 measure. To test the statistical significance of our results, we used the cross-validated paired t test for the difference in two F-1 measures with significance level of 0.05. A naive baseline that classifies all the threads in the subjective (majority) class will have a macro-averaged precision, recall and F-1 measure of 0.457, 0.676 and 0.545 respectively. Our most basic title (t) model (Table 3) beats this baseline.

Effect of different features: We expect the BoW+Bi and BoW+Bi+POS features to perform better than BoW and BoW+POS features as they contain more information about the thread text. Contrary to our expectation, for the title (t) setting, BoW and BoW+POS perform better than the two more advanced features. We conjecture that this is due to the informal nature of the titles. In many forum threads, titles only contain keywords related to the topic of discussion and hence are not *well-formed* sentences with proper grammar. For other settings, BoW+Bi and BoW+Bi+POS perform better than BoW and BoW+POS as expected.

Effect of different structural units: We see that titles give fair estimate of thread's subjectivity and initial posts (I) provide a better estimate. We expect this as initial posts contain the entire problem of discussion whereas titles only contain some key words related to the problem. Incorporating text from initial post in title (t+I) improves the performance slightly over the initial post (I) model. Further, adding the text of reply posts (t+I+R) gives the best performance.

Many online forums provide search functionalities for threads based only on titles as titles contain important information about the discussion problem [1]. But for subjectivity analysis, it turns out that initial posts and reply posts are better indicators of a thread's subjectivity. Hence, in order to use subjectivity analysis of threads in improving online forum search, search engines would need to reconsider the importance to be given to titles, initial posts and reply posts. This is an interesting direction in online forum search and we plan to investigate it as a future work.

6. CONCLUSION AND FUTURE WORK

In this paper, we presented a supervised machine-learning approach to classifying online forum threads as subjective or non-subjective. Our experiments show that features generated from n-grams and parts-of-speech tags of the textual content of forum threads give promising results and using the text of initial post and reply posts significantly improve the classification performance over the title (t) model. In the future, we plan to investigate the use of potential lexical clues of subjectivity and other features to further improve the subjectivity classification.

7. ACKNOWLEDGEMENTS

We would like to thank Sujay Patel for his help in conducting the experiments.

8. REFERENCES

[1] S. Bhatia and P. Mitra. Adopting inference networks for online thread retrieval. In *AAAI 2010, Atlanta, Georgia, USA, July 11-15*, pages 1300–1305, 2010.

[2] H. Duan and C. Zhai. Exploiting thread structures to improve smoothing of language models for forum post retrieval. *Advances in Information Retrieval*, pages 350–361, 2011.

[3] I. Gurevych, D. Bernhard, K. Ignatova, and C. Toprak. Educational question answering based on social media content. In *Proc. of the 14th International Conf. on Artificial Intelligence in Education*, pages 133–140, 2009.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[5] A. Hassan, V. Qazvinian, and D. R. Radev. What's with the attitude? identifying sentences with attitude in online discussions. In *EMNLP 2010*, pages 1245–1255. ACL, 2010.

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

[7] B. Li, Y. Liu, A. Ram, E. Garcia, and E. Agichtein. Exploring question subjectivity prediction in community qa. In *SIGIR 2008*, pages 735–736. ACM, 2008.

[8] B. Liu. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing,*, pages 978–1420085921, 2010.

[9] D. Ly, K. Sugiyama, Z. Lin, and M. Kan. Product review summarization from a deeper perspective. In *JCDL 2011*, pages 311–314, 2011.

[10] A. Mccallum and K. Nigam. A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on 'Learning for Text Categorization'*, 1998.

[11] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[12] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. In *EMNLP*, pages 105–112. ACL, 2003.

[13] J. Seo, W. B. Croft, and D. A. Smith. Online community search using thread structure. In *CIKM 2009*, pages 1907–1910, New York, NY, USA, 2009. ACM.

[14] S. Somasundaran, T. Wilson, J. Wiebe, and V. Stoyanov. Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In *ICWSM 2007*, 2007.

[15] V. Stoyanov, C. Cardie, and J. Wiebe. Multi-perspective question answering using the opqa corpus. In *EMNLP 2005*, pages 923–930. ACL, 2005.

[16] J. Wiebe, R. Bruce, and T. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *ACL 1999*, pages 246–253. ACL, 1999.

[17] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP 2003*, pages 129–136. ACL, 2003.

[18] Z. Zhai, B. Liu, L. Zhang, H. Xu, and P. Jia. Identifying evaluative sentences in online discussions. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.