Predicting Subjectivity Orientation of Online Forum Threads

Prakhar Biyani[1], Cornelia Caragea[2], and Prasenjit Mitra[1]

[1] The Pennsylvania State University, US
[2] University of North Texas, US
{pxb5080,pmitra}@ist.psu.edu, ccaragea@unt.edu

Abstract. Online forums contain huge amounts of valuable information in the form of discussions between forum users. The topics of discussions can be subjective seeking opinions of other users on some issue or non-subjective seeking factual answer to specific questions. Internet users search these forums for different types of information such as opinions, evaluations, speculations, facts, etc. Hence, knowing subjectivity orientation of forum threads would improve information search in online forums. In this paper, we study methods to analyze subjectivity of online forum threads. We build binary classifiers on textual features extracted from thread content to classify threads as subjective or non-subjective. We demonstrate the effectiveness of our methods on two popular online forums.

1 Introduction

Online forums contain huge amounts of discussions between Internet users on various domain-specific problems such as Mac OS products, cameras, operating systems, music, traveling, health, as well as 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. Topics of discussion in online forum threads can be *subjective* or *non-subjective*. Subjective topics seek personal opinions or viewpoints, whereas non-subjective topics seek factual information.

Different users have different needs. Some search the web for subjective information like discussions on a certain topic to educate themselves about multiple points of view related to the topic, people's emotions, etc. Others pose queries that are objective and have short factual answers. Specifically, a user may want to learn what other people think about some problem, e.g., "*which is the best camera for beginners?*" or they may want un-opinionated information such as facts or verifiable information, e.g., "*what do the numbers on camera lenses mean?*". We call the former question as *subjective* and the latter as *non-subjective*.

Subjective information needs are more likely to be satisfied by forum threads discussing subjective topics and non-subjective information needs are more likely to be satisfied by forum threads discussing non-subjective topics. Let us consider this example. A user has two information needs related to Canon 7D camera that

A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 109–120, 2013.

he conveys to some camera forum's search engine by issuing the following queries: 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 user's intent is different across the two queries. In the first query, the user seeks opinions of different camera users on the resolution of the Canon 7D camera, i.e., how different users feel about the resolution, what are their experiences (good, bad, excellent, etc.) with Canon 7D as far as its resolution is concerned; hence, the query is subjective. In the second query, the user does not seek opinions but an answer to a specific question, which in this case, is the value of the resolution and therefore the query is non-subjective. Hence, prior knowledge of the subjectivity of threads would help in satisfying users' information needs more effectively by taking into account the user's intent in addition to the keywords in the query. In order to answer such queries effectively, forum search engines need to identify subjective threads in online forums and differentiate them from threads providing non-subjective information. Threads can be filtered by matching their subjectivity orientation with that of the query or they can be ranked by combining scores of lexical relevance and subjectivity match with the query.

Here, we address the first part of this vision; we show how to identify the subjectivity of threads in an online forum with high accuracy using simple word features. Recent works on online forum thread retrieval have taken into account the distinctive properties of online threads such as conversational structure [1], and hyperlinking patterns and non-textual metadata [2] to improve their retrieval. Previous works on subjectivity analysis in social media have mainly focused on online review sites for opinion mining and sentiment analysis [3,4,5] and on improving question-answering in community QA [6,7,8,9]. In contrast, our focus is on analyzing subjectivity in online forums using content based features.

We propose a simple and effective classification method using textual features obtained from online forum threads to identify subjective threads of discussion. We model the task as a binary classification of threads in one of the two classes: subjective and non-subjective. We say a thread is subjective if its topic of discussion is subjective and non-subjective if its topic is non-subjective. We used combinations of words and their parts-of-speech tags as features. The features were generated from the text in: (i) the title of a thread, (ii) the title and initial post of a thread and (iii) the entire thread. We performed experiments on two popular online forums (Dpreview and Trip Advisor–New York forums). We used ensemble techniques to improve learning of classifiers on unbalanced datasets and also explored the effects of feature selection to improve the performance of our classifiers. Our experiments show that our classifiers using textual features produce highly accurate results with respect to F1-measure.

Our contributions are as follows. We show that simple features generated from n-grams and parts-of-speech tags work *effectively* for identifying subjective and non-subjective discussion threads in online forums. We believe that online forum search engines can improve their ranking functions by taking into account the subjectivity match between users' queries and threads.

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 [3,10,4,5]. An integral part of sentiment analysis is to separate opinionated (generally subjective) sentences from un-opinionated (non-subjective) sentences [10] 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 [4]. 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 [3,5]. In contrast, our work aims at predicting subjectivity orientation of forum threads for use in improving retrieval. In sentiment analysis, only subjective sentences are of interest because sentiments are generally expressed in subjective languages whereas in our case, 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.

Other recent works have used subjectivity analysis to improve question-answering in social media [6,7,8,9,11] and multi-document summarization [12,13]. For example, Stoyanov *et al.*, [8] 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.,* [11] 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.* [6] used classification to identify subjectivity orientation of questions in community QA. Gurevych *et al.* [7] 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.*

In our previous work [14], we performed thread level subjectivity classification using thread-specific non-lexical features. In contrast, in this work, we use ensembles of classifiers built on balanced samples using lexical features.

Next, we state our problem and describe various features used in the subjectivity classification task.

3 Problem Statement and Approach

An online forum thread starts with a topic of discussion posted by the (thread) starter in the title and initial post of the thread. The topic can either be subjective or non-subjective. Following the definitions of subjective and objective sentences given by Bruce et. al.[15], we say that a thread's topic is *subjective* if the thread starter seeks private states of minds of other people such as opinions, evaluations, speculations, etc. and *non-subjective* if the thread starter seeks factual and/or verifiable information. We call a thread subjective if its topic of discussion is subjective and non-subjective if it discusses a non-subjective topic. We assume that subjective threads have discussions, mainly, in subjective languages whereas non-subjective threads discuss, mainly, in factual languages. We note that there may be cases where this assumption does not hold good, however, analysis of such exceptional cases is not the focus of this paper and is left for future work.

Problem Statement: Given an online forum thread T, classify it into one of the two classes: subjective (denoted by +1) or non-subjective (denoted by -1).

In this work, we assume that a thread discusses a single topic which is specified by the thread starter in the title and the initial post. Analyzing subjectivity of threads with multiple topics is a separate research problem that is out of scope of this work.

3.1 Feature Generation

Intuitively, in online forums, threads discussing subjective topics would contain more subjective sentences compared to threads discussing non-subjective topics. This difference usually results in different vocabulary and grammatical structures of these two types of sentences [16]. To capture this intuition, we used words, parts-of-speech tags and their combinations as the features for classification. These features have been shown to perform well in other subjectivity analysis tasks [17,18,19]. 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 elements (title, initial post, reply posts) of a thread:

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

Table 1 describes feature generation on a sentence containing three words W_i, W_{i+1} and W_{i+2} and 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 we empirically found it to be more effective than *tf-idf* and *binary*) as the weighting scheme 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`

Table 1. Feature generation for sentence W_i W_{i+1} W_{i+2}

Feature type	Generated feature
BoW	W_i, W_{i+1}, W_{i+2}
BoW+POS	$W_i, POS_i, W_{i+1}, POS_{i+1}, W_{i+2},\ POS_{i+2}$
BoW+Bi	$W_i, W_{i+1}, W_{i+1}, W_iW_{i+1},\ W_{i+1}W_{i+2}$
BoW+Bi+POS	$W_i, POS_i, W_{i+1}, POS_{i+1}, W_{i+2}, POS_{i+2}, W_iW_{i+1}, W_iPOS_{i+1},$ $POS_iW_{i+1}, W_{i+1}W_{i+2}, W_{i+1}POS_{i+2}, POS_{i+1}W_{i+2}$

3.2 Model Training

We used a Naive Bayes classifier [20] for classification as it performs well on word features. We experimented with Support Vector Machines and Logistic Classifiers with *tf*, *tf-idf*, and *binary* as the feature encoding schemes, and found that the Naive Bayes classifier gave the best results. The Naive Bayes classifier outputs the following two probabilities for a test thread T: $P(+1|T)$, i.e., the probability of thread T belonging to the subjective class and $P(-1|T)$, i.e., the probability of thread T belonging to the non-subjective class, where $P(+1|T) + P(-1|T) = 1$.

Our datasets are highly unbalanced (as described in Section 4) with a majority of the threads belonging to the subjective class. In this setting, even a classifier labeling all the instances as subjective would give reasonably high overall accuracy while performing poorly on the minority class (the non-subjective class). To address this problem, one way is to create a balanced dataset by undersampling from the majority class an equal number of instances to the minority class size and then train a classifier on that dataset. Such a classifier is highly dependent on the small sample.

To address this problem, we used an *ensemble* of classifiers approach [21]. We created multiple balanced samples by taking all the threads of the minority class and sampling (multiple times) an equal number of threads from the majority class. We trained a classifier on each balanced sample. However, our test sets retain the "natural" distribution of the data, which is unbalanced. On the test set, we combined the predictions of all the classifiers for each instance. More precisely, we created n balanced datasets $D_1, \cdots, D_n$ and trained n classifiers $C_1, \cdots, C_n$ such that C_i is trained on D_i. For a test instance T, the final prediction of the ensemble is computed by averaging the prediction of all the classifiers. That is: $P_{ens}(+1|T) = \frac{1}{n}\sum_{i=1}^{n} P_{C_i}(+1|T)$, where $P_{C_i}(+1|T)$ is the probability estimate given by classifier C_i of thread T belonging to the subjective class. $P_{ens}(-1|T) = 1 - P_{ens}(+1|T)$. For classification, we used a threshold of 0.5 on the ensemble's prediction.

4 Datasets

To evaluate our approach, we used threads from the two popular online forums: Digital Photography Review (denoted by **dpreview**) and Trip Advisor–New

Table 2. Sample queries used for data collection from dpreview forum

Subjective Queries	Non-subjective Queries
nikon DSLR vs. sony DSLR	what is flash focal length
which camera should I buy for all round photography?	what does a wide angle lens do
carl zeiss better than canon	what is exposure compensation

York (denoted by **trip-advisor**), described below. The choice for these forums is that we wanted to evaluate our models across the two popular genres of online forums, namely, technical and non-technical online forums, **dpreview** is a technical forum whereas **trip-advisor** is a non-technical forum.

1. **dpreview** is an online forum with discussions related to digital cameras and digital photography[2]. We manually framed 39 queries, mix of subjective and non-subjective, on topics related to digital cameras (see Table 2 for several examples) and ran them on the Google search engine. We limited the search space of Google to the website *http://forums.dpreview.com/forums*, ensuring the results are discussion threads from the dpreview forum only. For each query, the top 200 returned threads were crawled and processed to identify structural elements (such as title, posts, authors, etc). Note that, in some cases, less than 200 threads were retrieved by the search engine.
2. **trip-advisor** is an online forum having travel related discussions mainly for New York city[3]. We used a publicly available dataset[4] [2] that had 83072 threads from which we randomly selected 700 threads for our experiments. The processing of threads for identifying thread elements (i.e., title, posts, authors, etc) is the same as for dpreview.

Data Annotation. Threads in our datasets were annotated by two human annotators. The annotators were asked to annotate a thread as subjective if its topic of discussion is subjective and 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[5].

The annotations for each dataset were conducted in three stages. First, the annotators were asked to annotate a sample of 20 threads (for which we already had annotations) from the dataset using the instruction set. Second, separate discussions were held between the first author and each annotator. Each annotator was asked to provide arguments (for the annotations) and, in case of inconsistencies, they were educated through discussions to attain a common understanding of subjectivity. Third, they were given the full dataset for annotation.

[2] `http://forums.dpreview.com/forums/`

[3] `http://www.tripadvisor.com/ShowForum-g60763-i5-New_York_City_New_York.html`

[4] `http://www.cse.psu.edu/ sub194/datasets/ForumData.tar.gz`

[5] `blindreview.com`

Table 3. Distribution of threads in the two classes

	Dpreview	Trip–Advisor New York
No. of subjective threads	3320	412
No. of non-subjective threads	536	197

The overall percentage agreement between the annotators was 90% on the **dpreview** dataset and 87% on **trip-advisor** dataset. For our experiments, *we used only the data on which the annotators agreed.* Table 3 shows the number of threads in the two classes. There are much more subjective than non-subjective threads in the two forums, which confirms that online forum users tend to discuss subjective topics. This observation is consistent with previous works on subjectivity analysis of other online social media such as community question answering sites. For example, Li *et al.* [6] found that 66% of the questions asked in Yahoo! Answers were subjective.

5 Experiments and Results

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

5.1 Experimental Setting

We used k-fold cross validation to evaluate our classification models. k-fold cross validation is a popular method for performance evaluation of classifiers when the data do not have dependencies. Since the method randomly partitions the data into training and test set, if there are dependent data points in the training and test, the prediction of the classifier will be biased. In our case, there were dependencies in the **dpreview** dataset. Threads corresponding to a query discussed similar topics and, hence, would contain similar words and would have similar subjectivity orientations. Their presence in both training and test sets would make the sets dependent. In such a setting, a classifier's performance may be overestimated because of the dependence bias. To address this problem, we used *leave-one-out* cross validation at the query level. Threads corresponding to a query were held-out and the classifier was trained on the remaining threads. Testing was done on the held-out set. This holding out was done for each query and the average of the classifiers' performance over all queries was computed. For the **trip-advisor** dataset, since there were not any inbuilt dependencies, we used *k-fold* cross validation with $k = 5$. We used the Weka data mining toolkit [22] 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 elements (title (t), initial post (I), reply posts (R)) of the threads. First, we trained a basic model where we used only the text of the titles (denoted by t) for classification; that is our baseline. Then, we incorporated the text of initial posts (denoted by t+I) and finally, we used the textual content of the entire thread (denoted by t+I+R) for classification. For each dataset, we performed experiments using: (i) a single classifier trained on a balanced sample, (ii) a single classifier trained on the entire unbalanced dataset, and (iii) an ensemble of n classifiers, with each classifier in the ensemble being trained on a balanced sample of the data. For the ensemble, we empirically determined the value of n, that is, we conducted experiments with different values of n and used the value corresponding to the best results, $n = 20$ for **dpreview** and $n = 7$ for **trip-advisor**. Also, we investigated the effect of feature selection on the classification performance. We ranked the features using Information Gain [23] to get the most informative ones with respect to the class variable. We trained classifiers for various numbers of selected features, starting from 100 and ending at 2000, in steps of 100.

5.2 Results

Table 4 is divided into two halves. The upper half shows the results for **dpreview** and the lower half shows the results for **trip-advisor**. We used macro averaged F1-measure to report the classification performance of our models.

Effect of Different Features: For **dpreview**, the combination of unigrams, bigrams and part-of-speech tags (BoW+Bi+POS) extracted from title and the initial post gave the best F1-measure (0.884), using an ensemble of classifiers, whereas for **trip-advisor**, the same combination of unigrams, bigrams and part-of-speech tags (BoW+Bi+POS) this time extracted from title, the initial post, and the reply posts gave the best F1-measure (0.745), using again an ensemble of classifiers. However, for **trip-advisor**, the improvement in performance by incorporating parts-of-speech tags over BoW+Bi is not statistically significant.

Effect of Different Structural Units: In Table 4, we see that incorporating text from the first post (t+I) improves the classification performance over the baseline (t) for the two datasets. This observation suggests that initial posts along with titles convey more information than titles alone about the subjectivity orientation of online threads, which is intuitive as titles contain only a few keywords about the topic whereas initial posts contain full details about the topic. Incorporation of text from the reply posts has different effects for the two datasets. For **dpreview**, the classification performance remains almost the same as compared to t+I setting. However, for **trip-advisor**, there is a high improvement in performance. In principle, this observation says that for the dpreview forum the subjectivity orientation of threads is mainly determined by their titles and initial posts combined, and the reply posts do not convey any significant additional information about the subjectivity orientation. For the trip-advisor forum, the subjectivity orientation of threads is determined by the entire thread including its reply posts. We conjecture the reason of this difference to be the *more* informal nature of trip-advisor than dpreview as the former is a non technical forum and the latter is a technical forum. In trip-advisor threads, there is

Table 4. Classification performance (F1-measure) 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 that we used (explained in Table 1). [Sin] and [Ens] denote experiments with single balanced sample and with ensembling (i.e., using multiple balanced samples) respectively.

Dpreview dataset (leave-one-out cross validation)				
	BoW	BoW+POS	BoW+Bi	BoW+Bi+POS
t[Sin]	0.791	0.802	0.787	0.793
t+I[Sin]	**0.862**	**0.865**	**0.871**	**0.877**
t+I+R[Sin]	0.859	0.859	0.876	0.875
t [Ens]	0.807	0.811	0.807	0.801
t+I [Ens]	**0.865**	**0.865**	**0.877**	**0.884**
t+I+R [Ens]	0.867	0.863	0.876	0.878
Trip Advisor–New York dataset (5-fold cross validation)				
	BoW	BoW+POS	BoW+Bi	BoW+Bi+POS
t[Sin]	0.557	0.572	0.561	0.552
t+I[Sin]	0.606	0.618	0.642	0.666
t+I+R[Sin]	**0.701**	**0.702**	**0.729**	**0.738**
t [Ens]	0.565	0.564	0.568	0.566
t+I [Ens]	0.633	0.641	0.674	0.691
t+I+R [Ens]	**0.723**	**0.717**	**0.74**	**0.745**

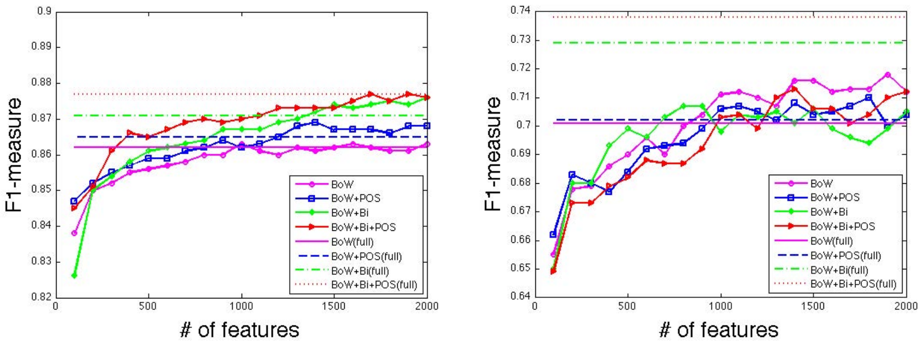
generally more topic drift, i.e., there are discussions that are not related to the topic specified by the titles and initial posts of the threads. Hence, the subjectivity orientation is no longer, mainly, determined by titles and initial posts of the threads. We plan to investigate this difference in more detail as part of future research on subjectivity analysis of online forums.

To verify that these differences (in results) are not due to the difference in sizes of the two datasets, we conducted additional experiments with the **dpreview** dataset. We experimented with a small fraction of dpreview, i.e., 0.35, obtained by under-sampling [24] from the entire dataset. Specifically, we first under-sampled from the minority class of dpreview a small subset that approximately matched the size of the minority class in **trip-advisor**; we then under-sampled from the majority class of **dpreview** to obtained a balanced subset (same number of instances from both classes). Hence, on **dpreview**, we trained classifiers on approximately the same sized balanced samples as in **trip-advisor**, where the size of balanced sample is 394 (197 subjective and 197 non-subjective). The under-sampling was performed only on the training set (the test set remained unbalanced). Table 5 provides results for this experiment.

Effect of ensembling: For both datasets, using an ensemble of classifiers, with each classifier trained on a balanced sample, improves the performance of a single classifier trained on a balanced sample. However, the improvement is generally small, especially for **dpreview** (see Table 4). This implies that the classifiers learn almost the same patterns from the different random samples of the majority class.

Table 5. The performance of classifiers (in terms of F1-measure) trained on smaller balanced samples of the **dpreview** dataset. The number of threads in the balanced sample is 376 (188 subjective and 188 non-subjective). As can be seen, performance of t+I is similar to that of t+I+R.

	BoW	BoW+POS	BoW+Bi	BoW+Bi+POS
t	0.772	0.777	0.764	0.764
t+I	0.863	0.863	0.869	0.87
t+I+R	0.876	0.878	0.859	0.857



(a) Dpreview **(b)** Trip Advisor–New York

Fig. 1. Classification performance of top 2000 features for the two datasets for settings t+I (for dpreview) and t+I+R (for Trip Advisor–New York). Straight lines represent performance corresponding to all the features for a particular kind of representation (Table 4).

Effect of Feature Selection: Figures 1(a) and 1(b) show the performance of *single* classifiers (not ensembling) as a function of the number of features, ranging from 100 to 2000 in steps of 100, for **dpreview** and **trip-advisor**, respectively. Due to space constraints, we only report the results for the two best performing experimental settings for the two datasets: t+I for **dpreview** and t+I+R for **trip-advisor**. We used all the feature representations described in Table 1. For **dpreview**, the performance of the BoW+Bi+POS-based classifier using all the features ($\approx 100,000$ features) is matched by that of the BoW+Bi+POS-based classifier using only the top 1700 selected features (F1-measure = 0.877). On the other hand, for **trip-advisor**, the BoW-based classifier using feature selection (with the number of features ranging between 100 and 2000) achieves the highest performance (F1-measure =0.718) using 1900 features, which is worse than that of BoW+Bi+POS-based classifier using all the features (F1-measure =0.738). However, in every case (for the two datasets) the number of features corresponding to the best performance is much smaller compared to the total number of features.

Table 6. True positive rates (for minority class) of classifiers trained on unbalanced and balanced data for the two datasets for BoW features

	Dpreview		**Trip Advisor–New York**	
	Unbalanced	Balanced	Unbalanced	Balanced
t	0.53	0.752	0.305	0.635
t+I	0.56	0.73	0.467	0.66
t+I+R	0.558	0.618	0.426	0.545

Unbalanced Dataset vs. Balanced Dataset: Table 6 compares true positive rates (for the minority class) of single classifiers trained on balanced and unbalanced (entire) data for the two datasets. As expected, classifiers built on unbalanced data performed worse on the minority class when compared to those trained on balanced datasets. We show the results only for BoW features for the three experimental settings, (t), (t+I), (t+I+R), but the same behavior was observed for other types of features.

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 methods showed that features generated from n-grams and parts-of-speech tags of the textual content of forum threads give promising results. In the future, we plan to use the subjectivity analysis to improve search in online forums.

Acknowledgement. This material is based upon work supported by the National Science Foundation under Grant No. 0845487.

References

1. Seo, J., Croft, W., Smith, D.: Online community search using thread structure. In: CIKM, pp. 1907–1910. ACM (2009)
2. Bhatia, S., Mitra, P.: Adopting inference networks for online thread retrieval. In: Proceedings of the 24th AAAI, pp. 1300–1305 (2010)
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: SIGKDD, pp. 168–177. ACM (2004)
4. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing (2010) ISBN: 978–1420085921
5. Ly, D., Sugiyama, K., Lin, Z., Kan, M.: Product review summarization from a deeper perspective. In: JCDL, pp. 311–314 (2011)
6. Li, B., Liu, Y., Ram, A., Garcia, E., Agichtein, E.: Exploring question subjectivity prediction in community qa. In: SIGIR, pp. 735–736. ACM (2008)

7. Gurevych, I., Bernhard, D., Ignatova, K., Toprak, C.: Educational question answering based on social media content. In: Proceedings of the 14th International Conference on Artificial Intelligence in Education. Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, pp. 133–140 (2009)
8. Stoyanov, V., Cardie, C., Wiebe, J.: Multi-perspective question answering using the opqa corpus. In: EMNLP, pp. 923–930. Association for Computational Linguistics (2005)
9. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: EMNLP, pp. 129–136. Association for Computational Linguistics (2003)
10. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2, 1–135 (2008)
11. Somasundaran, S., Wilson, T., Wiebe, J., Stoyanov, V.: Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In: ICWSM (2007)
12. Seki, Y., Eguchi, K., Kando, N., Aono, M.: Multi-document summarization with subjectivity analysis at duc 2005. In: DUC. Citeseer (2005)
13. Carenini, G., Ng, R., Pauls, A.: Multi-document summarization of evaluative text. In: EACL, pp. 305–312 (2006)
14. Biyani, P., Bhatia, S., Caragea, C., Mitra, P.: Thread specific features are helpful for identifying subjectivity orientation of online forum threads. In: COLING, pp. 295–310. ACL, Mumbai (2012)
15. Bruce, R., Wiebe, J.: Recognizing subjectivity: a case study in manual tagging. Natural Language Engineering 5, 187–205 (1999)
16. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: EMNLP, pp. 105–112. ACL (2003)
17. Li, B., Liu, Y., Agichtein, E.: Cocqa: co-training over questions and answers with an application to predicting question subjectivity orientation. In: EMNLP, pp. 937–946. ACL, Stroudsburg (2008)
18. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: EMNLP, pp. 129–136. ACL, Stroudsburg (2003)
19. Aikawa, N., Sakai, T., Yamana, H.: Community qa question classification: Is the asker looking for subjective answers or not? IPSJ Online Transactions 4, 160–168 (2011)
20. Mitchell, T.: Machine learning. McGraw-Hill, Burr Ridge (1997)
21. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000)
22. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The weka data mining software: an update. ACM SIGKDD Explorations Newsletter 11, 10–18 (2009)
23. Yang, Y., Pedersen, J.: Feature selection in statistical learning of text categorization. In: Proc. ICML, pp. 412–420 (1997)
24. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: One-sided selection. In: ICML, pp. 179–186. Morgan Kaufmann (1997)