# Learning to Identify Subjective Sentences

**Girish K. Palshikar, Manoj Apte, Deepak Pandita**
TCS Research, Tata Consultancy Services
Pune 411013
{gk.palshikar, manoj.apte, deepak.p7}@tcs.com

**Vikram Singh**
M. Tech., Computer Science & Engineering
Indian Institute of Technology, Patna
vikramsinghiitp@gmail.com

## Abstract

Subjective sentences describe people's opinions, points-of-view, interpretations, comparisons, sentiments, judgments, appraisals or feelings toward entities, events and their properties. Identifying subjective sentences is the basis for opinion mining and sentiment analysis and is important in applications like political analysis, social media analytics and product review analytics. We use standard classifiers to build models of SUBJECTIVE vs. NON-SUBJECTIVE sentences and demonstrate that they outperform the approaches reported in the literature on several interesting datasets. We discuss two novel applications of this work: prevalence of subjective sentences in performance appraisal text and scientific papers. We demonstrate that scientific papers also contain a substantial fraction of subjective sentences. We compare the nature of subjective sentences in performance appraisals text and scientific papers, and observe *different* reasons why some sentences in these domains are subjective. We propose the need to further investigate the linguistic and semantic basis for subjective sentences across different domains.

## 1 Introduction

Documents contain a mixture of facts, opinions and other kinds of sentences, such as questions or instructions. Factual sentences include *objective expressions* about entities, events and their properties. Opinion sentences usually include *subjective expressions* that describe people's sentiments, judgments, appraisals or feelings toward entities, events and their properties (Liu, 2010). Objective information is typically fact-based, measurable, observable and verifiable. For example, (S1) (Table1) is a factual sentence. In contrast, most opinion sentences express some sentiment, usually having a positive or negative polarity. However, some opinion sentences are *neutral* i.e., they do not explicitly express any particular sentiment. For example, (S2) is an opinion sentence that expresses some sentiment (mostly positive), whereas (S3), (S4) and (S5) are opinion sentences that do not express any particular sentiment. Some factual sentences may be *mixed* i.e., they may also contain subjective expressions, with or without the presence of an explicit sentiment polarity. Sentences (S6) and (S7) are mainly factual, but (S6) contain an opinion along with a positive sentiment whereas (S7) expresses an opinion without much sentiment. Any expression about the private (internal) state of mind of a person is, by definition, a subjective sentence; e.g., (S8). Opinion sentences often contain subjective expressions other than sentiments, such as opinions, points of view (S4), judgements (S5), predictions (S9), interpretations, comparisons (S10) etc.

In this paper, we are interested in automatically identifying *subjective* sentences, which contain subjective expressions, but which may or may not contain explicit sentiment markers. We are also interested in detecting mixed sentences, which contain both facts and opinions. In this paper, we include mixed sentences in the class of subjective sentences. The remaining sentences (i.e., of the other class) either express pure facts or they may express neither facts nor opinions (e.g., they may be instructions or questions).

Categorization of sentences as subjective or



*D S Sharma, R Sangal and A K Singh. Proc. of the 13th Intl. Conference on Natural Language Processing*, pages 239–248,
Varanasi, India. December 2016. 

| | |
|---|---|
| S1 | `On April 24, 1975, the West German embassy in Stockholm was seized by members of the RAF; two of the hostages were murdered as the German government under Chancellor Helmut Schmidt refused to give in to their demands.` |
| S2 | `Windows 7 is quite simply faster, more stable, boots faster, goes to sleep faster, comes back from sleep faster, manages your files better and on top of that it's beautiful to look at and easy to use.` |
| S3 | `More than 98 bank entities are expected to be serviced under the Web based PaySys application.` |
| S4 | `With Spy Kids 2, the Spy Kids franchise establishes itself as a part of the movie landscape for children.` |
| S5 | `The security cover for deputy ministers can be reduced.` |
| S6 | `The ninth seed Sindhu took 40 minutes to dismantle the eighth seeded Tai and gave her medal chances a boost in her maiden Olympic appearance.` |
| S7 | `Google Duo isn't much different from the other video chatting services, except that it gives a glimpse at who is making the call, helping the recipient decide whether to answer.` |
| S8 | `John found his sorrow receding in the cool breeze.` |
| S9 | `The Independence Day holiday will boost the film's biz further.` |
| S10 | `The Mi 5 camera in performs better in most outdoor situations than OnePlus 3.` |
| S11 | `Baseball writer Bill James, in the The New Bill James Historical Baseball Abstract, ranked Robinson as the 32nd greatest player of all time strictly on the basis of his performance on the field, noting that he was one of the top players in the league throughout his career.` |

Table 1: Some example sentences.

non-subjective has many practical applications. Information extraction, question-answering and document summarization are often interested in factual sentences, and hence would use the classification of sentences to remove subjective sentences, which can improve the system performance and effectiveness. Some real-life applications explicitly need to mark and summarize sentiments and opinions (which are inherently subjective); for example, identifying features facing criticism from product reviews, identifying extreme positions from political views, and marketing oriented anaysis of web contents.

Classification of sentences from different perspectives is a well-explored problem. Syntactically, a sentence is typically classified into classes such as DECLARATIVE, IMPERATIVE, INTERROGATIVE, EXCLAMATIVE, COMMUNICATIVE, INFORMATIVE etc., with further sub-classes. Other structurally-oriented sentence classes include MAJOR (has subject and predicate), MINOR (without a finite verb; e.g., `The more, the merrier.`), PERIODIC (meaning is not complete until the final clause or phrase; e.g., `Silent and soft, and slow, descends the snow.`) etc. Semantically classifying sentences (based on the sentence's purpose) is a much harder task, and is gaining increasing attention from linguists and NLP researchers (Zhou et al., 2004), (Wang et al., 2005), (McKnight and Srinivasan, 2003), (Cohen et al., 2004), (Corston-Oliver et al., 2004), (Ivanovic, 2006), (Khoo et al., 2006), (Yamamoto and Takagi, 2005), (Hachey and Grover, 2004), (He et al., 2006), (Naughton et al., 2010), (Kadoya et al., 2005), (Momtazi and Klakow, 2009). Most work in this area has used supervised learning approaches (e.g., using SVM, decision trees, maximum entropy based classifier, naive Bayes etc.), with the exception of (Ito et al., 2004) (semi-supervised) and (Teufel and Moens, 1998), (Ibekwe-SanJuan et al., 2008) (knowledge-based) and (Deshpande et al., 2010) (rule-based). Sentence classification has been applied to tasks such as summarization, information extraction, IR, automatic ontology creation (Hearst, 1998) and text entailment (Zanzotto and DellArciprete, 2009). Sentence classification has been used on documents in several practical application domains such as biomedical papers, legal judgements, product reviews, customer complaints in help-desk, emails etc. The sentences classes have also been more domain dependent (Table 2).

In this paper, we use a variety of standard classifiers to build models of SUBJECTIVE vs. NON-SUBJECTIVE sentences. We demonstrate that some of the standard classifiers outperform the approaches reported in the literature on several interesting datasets, including one real-life dataset from us[1]. We discuss two novel applications of this work, which have not been reported in the literature to our knowledge. We perform predic-

[1]Please contact the authors for obtaining this dataset.

tive identification (and a subsequent analysis) of the prevalence of subjective sentences in two interesting domains: the first in performance appraisal text and the other in scientific papers.

The rest of the paper is organized as follows. We review the related work in Section 2. Then, we describe the features used for a classification based approach to identify subjective sentences in Section 3. The experiments performed and the results are described in Section 4. In Section 5 we apply the classifiers learned earlier to identify subjective sentences in two novel domains: (i) performance appraisal text; and (ii) scientific papers. Finally, we discuss the conclusions of our work and identify some further work in Section 6.

## 2 Related Work

There has been a considerable amount of research work in the field of Sentiment Analysis and Opinion Mining. We focus here on predictive identification of subjective text, which can be done at document, sentence and phrase level by using features like presence of a pronoun, adjective, semantic orientation etc. Often the other class is objective or factual text (as against a general non-subjective class, as in this paper).

(Yu and Hatzivassiloglou, 2003) presented a Bayesian classifier for discriminating between Opinion or Factual articles. Then, three approaches to classify opinions from facts at the sentence level were proposed. The first approach was based on the similarity of the given sentence to the opinion or fact documents. In the second approach, a Naive Bayes classifier was trained using the sentences in opinion and fact documents. The features included words, bigrams, trigrams and the parts of speech, as well as the presence of semantically oriented words combined with the polarity information and the average semantic orientation of the words in a sentence. In the third approach multiple Naive Bayes classifiers were trained using iterative training process relying on different set of features.

In a classic paper (Pang and Lee, 2004), the authors have used Naive Bayes and SVM to detect subjective sentences and used these sentences for sentiment analysis and proposed a novel graph-cut based method for assigning the overall polarity to a document, based on the sentiments in these individual sentences.

(Kim and Hovy, 2005) developed a collection of opinion bearing and non-opinion bearing words using multiple collections like Wordnet, WSJ Data, Columbia wordlist and then combined these lists into a single collection with averaging the score of individual lists. Score of a sentence was computed using these scores in two ways. In the first approach the scores of all the words in the sentence were added and in second approach presence of a single strong valence word was used as the score.

(Stepinski and Mittal, 2007) classified news articles into facts and opinions by training a classifier on opinion/fact stories. Sentences in an article were labeled as factual and opinion and an overall score was computed as the average of these labels. Label was weighed based on the confidence of the classification. Then a Passive-Aggressive algorithm (Crammer et al., 2006) was trained on unigram, bigram and trigram features. An iterative training process in which a classifier is trained using a subset of the desired features was used. Based on the results of classifying the training set with the classifier, the misclassified sentences were removed and the classifier was trained again using a larger set of features.

(Kim and Myaeng, 2007) introduced a lexical information based methodology for opinion analysis. The task of classifying sentences into Subjective or Objective was done by using a combination of rule-based approach and a machine learning based method. First, a training set was obtained using the rules based on lexical clues and then a SVM-based classifier was trained on the same. Polarity or semantic orientation of the subjective sentences was determined and the opinion holders were also found by using Named-Entity Recognition with extracted lexical clues.

(Godbole et al., 2007) use co-occurrence of an entity and a sentiment word in the same sentence to mean that the sentiment is associated with that entity. Mentioning that it can lead to inaccuracies they claim that due to the volume text that is processed by them enables them to generate accurate sentiment words. They create two scores for each entity polarity score and subjectivity score. Polarity indicates percentage of positive sentiment references among total sentiment references, while Subjectivity indicates proportion of sentiment to frequency of occurrence.

(Pak and Paroubek, 2010) have presented a method for automatic collection of corpus that can

| Domain | Sentence Classes |
|---|---|
| research papers | BACKGROUND, TOPIC, RELATED-WORK, PURPOSE/PROBLEM, HYPOTHESIS, AIM, SOLUTION/METHOD, RESULT, CONCLUSION/CLAIM, FUTURE-WORK (Yamamoto and Takagi, 2005), (Ibekwe-SanJuan et al., 2008), (Teufel and Moens, 1998), (McKnight and Srinivasan, 2003) |
| movies | OPINIONATIVE, FACTOID (Momtazi and Klakow, 2009) |
| product reviews | RECOMMEND, NOT-RECOMMEND (Wang et al., 2005) |
| help-desk | REQUEST, QUESTION, APOLOGY, INSTRUCTION, SUGGESTION, STATEMENT, SPECIFICATION, THANKS, APOLOGY, RESPONSE-ACK (Khoo et al., 2006) |
| legal | FACT, PROCEEDINGS, BACKGROUND, FRAMING, DISPOSAL (Hachey and Grover, 2004) |
| emails | REQUEST, PROPOSE, AMEND, DELIVER, COMMIT, MEETING, DATA, TASK, CHITCHAT, FAREWELL (Cohen et al., 2004), (Corston-Oliver et al., 2004) |
| biography | BIO, FAME, PERSONALITY, SOCIAL, EDUCATION, NATIONALITY, SCANDAL, PERSONAL, WORK (Zhou et al., 2004) |

Table 2: Examples of Sentence Classes (modified from (Deshpande et al., 2010)).

be used to train a sentiment classifier. The corpus contains both positive and negative sentiments as well as objective text (no sentiments). For positive and negative sentiments help was taken from emoticons used in the tweets and for objective text the messages from twitter accounts of popular newspapers and magazines were considered. They first conduct basic statistical analysis of the collected corpus and then run multinomial naive based classifier that uses N-Gram and POS Tags as features on the data collected.

## 3 Classification of Subjective Sentences

We adopt a feature-based classification approach to classify the sentences into facts and opinions. We extract several features from each sentences and thus represent each sentence as a numeric feature vector. We then use several standard classifiers for the task of classifying a sentence as SUBJECTIVE or NON-SUBJECTIVE and compare their results.

**Number of adjectives:** An interesting relation between presence of adjectives in a sentence and its subjectivity has been explored in the literature (Wiebe et al., 1999), (Hatzivassiloglou and Wiebe, 2000), (Wiebe, 2000). (Wiebe et al., 1999) demonstrated that adjectives are statistically significantly, and positively correlated with subjective sentences in the corpus on the basis of the log-likelihood ratio. The probability that a sentence is subjective, given that there is at least one adjective in the sentence, is 0.545, even though there are more objective than subjective sentences in the corpus. Therefore, we use number of adjectives in a sentence as a feature. Note that if the same adjective occurs multiple times in a sentence, then we count each occurrence separately.

**Number of nouns:** (Riloff et al., 2003) reported the effectiveness of nouns for identification of subjective sentences. Hence, we use number of nouns, including proper nouns, as a feature.

**Word count:** Opinions and subjective sentences tend to be more elaborate and hence longer. So we use number of words in the sentences, excluding stop words and punctuations, as a feature.

**Number of strongly subjective words:** Most of the systems developed for subjectivity classification and sentiment analysis, such as (Liu, 2010), (Syed et al., 2010), use a lexicon of opinion-bearing words. Two of the most commonly used lexicons are the SentiWordNet (Baccianella et al., 2010) and the Subjectivity Lexicon (Wilson et al., 2005b). The Subjectivity Lexicon classifies a word as strongly subjective (e.g., `beautiful`) or weakly subjective (e.g., `benefit`). We use the number of strongly subjective words from the Subjectivity Lexicon present in a sentence divided by word count as a feature. We also use the number of words from SentiWordNet (having either nonzero positive polarity score or nonzero negative polarity score) present in a sentence divided by word count as another feature.

**Number of named entities:** A commonly observed characteristics of a factual or objective sentence is then it tends to use more number of named entities. Hence we use number of named entitied present in a sentence as a feature. We use Stanford NER tool for this purpose.

**Number of comparatives:** This feature represents the number of comparative words (e.g., `faster`) used in the sentence.

**Number of superlatives:** This feature refers to the number of superlative words (e.g., `fastest`) used in the sentence.

| Dataset | #Sentences | #SubjSent | #NonSubjSent |
|---|---|---|---|
| D1 | 2074 | 1130 | 944 |
| D2 | 10000 | 5000 | 5000 |
| D3 | 293 | 135 | 158 |
| D4 | 613 | 281 | 332 |

Table 3: Dataset Description

**Tense:** Sentences in future tense tend to be more subjective. Hence we use the major tense (past, present or future) of the sentence as a feature.

**Number of adverbs:** This feature represents the number of adverbs (e.g., `simply, back`) in the sentence.

**Number of date, time and numbers:** The intuition behind this is that the factual information generally has lots of dates and numeric data. Such a sentence has a higher probability of being non-subjective than subjective. Therefore, we use number of date, time and number entities present in the sentence as a feature.

As an example, the feature vector for the non-subjective sentence S11 is $[3, 18, 25, 0.08, 0.32, 2, 0, 1, 0, 1, 1]$. This sentence has number of adjectives = 3 (`Historical, 32nd, top`), number of nouns = 18, word count = 25, fraction of strongly subjective words as per Subjectivity Lexicon = $2/25 = 0.08$ (`greatest, strictly`), fraction of sentiment words as per SentiWordNet $= 8/25 = 0.32$, number of named entities = 2 (`Bill James, Robinson`), number comparatives = 0, number of superlatives = 1 (`greatest`), tense = 0 (past), number of adverbs = 1 (`strictly`), number of date, time and numbers = 1 (`one`).

## 4 Experiments and Results

### 4.1 Datasets

Table 3 lists the datasets we have used in this paper. Dataset D1 contains 1130 subjective sentences taken from the opinosis dataset (Ganesan et al., 2010), to which we added 944 non-subjective sentences that contain factual information about various domains like science, history, sports etc. Dataset D2 is the subjectivity dataset published in (Pang and Lee, 2004), and contains 5000 subjective and 5000 non-subjective sentences. Dataset D3 and D4 are explained later.

### 4.2 Results

We train a number of classifiers from the WEKA toolset on dataset D2, using 10-fold cross validation. For stacking-based classification, we use Logistic as a meta-classifier (combiner of base classifiers). In Stacking1, we used Naive Bayes, SVM, MultilayerPerceptron and Random Forest as base classifiers. In Stacking2, we use Naive Bayes, SVM, Logistic regression, MultilayerPerceptron and Random Forest as base classifiers. In Stacking3, we used MultilayerPerceptron + bagging, Logistic + AdaBoost as base classifiers.

For comparison, we use OpinionFinder and TextBlob. OpinionFinder is a system that automatically identifies subjective sentences, as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. OpinionFinder has two classifiers for subjectivity classification, one is rule-based and the other is a model trained on MPQA corpus (Wilson et al., 2005a), (Riloff and Wiebe, 2003), (Wiebe and Riloff, 2005), (Wilson et al., 2005b), (Riloff and Wiebe, 2003). TextBlob is a Python library for processing textual data and provides an API for common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction and sentiment analysis (`https://textblob.readthedocs.io/en/dev/`).

The results given by various classifiers trained on dataset D2 using 10-fold cross validation are shown in Table 4. As seen, Stacking3 shows the best $F$-measure on D2, which is much better than the baselines of both the OpinionFinder subjectivity classifiers and TextBlob. The $F$-measure of Stacking2 and MLP + bagging classifier on D2 is also quite close to Stacking3. We then use the classifiers trained on D2 and apply them to D1 as the test dataset. Once again, Stacking3 shows the best $F$-measure on D1, which is much better than the baselines of both OpinionFinder subjectivity classifiers but, is very close to TextBlob. The $F$-measures of the MLP + bagging and Logistic + Bagging on D1 are also quite close to Stacking3.

## 5 Applications

### 5.1 Performance Appraisal Text

*Performance appraisal* (PA) is a crucial HR process that enables an organization to periodically measure and evaluate every employee's perfor-

| Classifier | D1 | | | D2 | | | D3 | | | D4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| J48 | 0.727 | 0.683 | 0.677 | 0.666 | 0.666 | 0.666 | 0.675 | 0.669 | 0.659 | 0.622 | 0.622 | 0.622 |
| Logistic regression | 0.748 | 0.699 | 0.692 | 0.684 | 0.683 | 0.683 | 0.681 | 0.672 | 0.66 | 0.671 | 0.672 | 0.67 |
| MultilayerPerceptron (MLP) | 0.736 | 0.694 | 0.688 | 0.681 | 0.681 | 0.681 | 0.685 | 0.679 | 0.67 | 0.66 | 0.656 | 0.656 |
| Naive Bayes (NB) | 0.726 | 0.688 | 0.682 | 0.666 | 0.664 | 0.663 | 0.696 | 0.693 | 0.687 | 0.655 | 0.656 | 0.651 |
| Random Forest (RF) | 0.65 | 0.62 | 0.614 | 0.645 | 0.645 | 0.644 | 0.646 | 0.642 | 0.629 | 0.611 | 0.613 | 0.611 |
| SVM | 0.719 | 0.685 | 0.68 | 0.685 | 0.684 | 0.684 | 0.655 | 0.652 | 0.643 | 0.638 | 0.63 | 0.63 |
| J48 + AdaBoost | 0.708 | 0.662 | 0.654 | 0.656 | 0.656 | 0.656 | 0.621 | 0.618 | 0.602 | 0.633 | 0.635 | 0.631 |
| Logistic + AdaBoost | 0.748 | 0.699 | 0.692 | 0.684 | 0.683 | 0.683 | 0.681 | 0.672 | 0.66 | 0.671 | 0.672 | 0.67 |
| MLP + AdaBoost | 0.739 | 0.696 | 0.69 | 0.681 | 0.681 | 0.681 | 0.69 | 0.683 | 0.673 | 0.663 | 0.659 | 0.66 |
| NB + AdaBoost | 0.726 | 0.688 | 0.682 | 0.666 | 0.664 | 0.663 | 0.696 | 0.693 | 0.687 | 0.655 | 0.656 | 0.651 |
| RF + AdaBoost | 0.664 | 0.628 | 0.62 | 0.647 | 0.647 | 0.647 | 0.636 | 0.631 | 0.617 | 0.626 | 0.628 | 0.624 |
| SVM + AdaBoost | 0.719 | 0.675 | 0.668 | 0.676 | 0.676 | 0.676 | 0.688 | 0.672 | 0.656 | 0.652 | 0.653 | 0.652 |
| J48 + bagging | 0.696 | 0.649 | 0.64 | 0.667 | 0.667 | 0.667 | 0.682 | 0.662 | 0.641 | 0.638 | 0.638 | 0.63 |
| Logistic + bagging | 0.751 | 0.7 | 0.693 | 0.684 | 0.684 | 0.684 | 0.686 | 0.676 | 0.664 | 0.673 | 0.674 | 0.671 |
| MLP + bagging | 0.752 | 0.702 | 0.694 | 0.688 | 0.688 | 0.687 | 0.71 | 0.693 | 0.679 | 0.663 | 0.664 | 0.661 |
| NB + bagging | 0.728 | 0.688 | 0.682 | 0.665 | 0.663 | 0.662 | 0.696 | 0.693 | 0.687 | 0.664 | 0.664 | 0.658 |
| RF + bagging | 0.684 | 0.641 | 0.632 | 0.656 | 0.656 | 0.656 | 0.639 | 0.631 | 0.613 | 0.618 | 0.62 | 0.615 |
| SVM + bagging | 0.716 | 0.684 | 0.68 | 0.682 | 0.682 | 0.682 | 0.647 | 0.645 | 0.637 | 0.647 | 0.638 | 0.638 |
| Stacking1 | 0.733 | 0.69 | 0.684 | 0.685 | 0.685 | 0.685 | 0.694 | 0.683 | 0.671 | 0.671 | 0.67 | 0.671 |
| Stacking2 | 0.74 | 0.696 | 0.689 | 0.687 | 0.687 | 0.687 | 0.7 | 0.686 | 0.673 | **0.677** | **0.677** | **0.677** |
| Stacking3 | **0.757** | **0.719** | **0.715** | **0.688** | **0.688** | **0.688** | **0.718** | **0.71** | **0.702** | 0.665 | 0.666 | 0.665 |
| OpinionFinder rule-based | 0.37 | 0.423 | 0.311 | 0.371 | 0.388 | 0.367 | 0.364 | 0.454 | 0.374 | 0.42 | 0.498 | 0.404 |
| OpinionFinder classifier | 0.722 | 0.599 | 0.559 | 0.58 | 0.577 | 0.572 | 0.625 | 0.604 | 0.563 | 0.704 | 0.643 | 0.595 |
| TextBlob | 0.712 | 0.713 | 0.712 | 0.599 | 0.595 | 0.591 | 0.597 | 0.59 | 0.591 | 0.642 | 0.643 | 0.642 |

Table 4: Results (Precision, Recall and F-Measure) of various classifiers on different datasets.

mance and also to drive performance improvements. While the use of IT-based PA systems is fairly common in modern organizations, the use of text-mining techniques to derive insights from PA data is relatively less explored in academic research (Apte et al., 2016), (Ramrakhiyani et al., 2016). In most PA processes, the communication contains 2 major steps: (i) in *self-appraisal*, where an employee records his/her achievements, activities, tasks handled etc.; and (ii) in *feedback*, where the supervisor provides the criticism, appreciation, judgement, evaluation and suggestions for improvement of performance etc.

We have selected a small set of sentences from feedback step in a real performance appraisal in a large IT organization. This dataset D3 has 293 sentences, which were manually tagged by two people independently. We trained the classifier Stacking3 on Dataset D2 and applied the learned model to test on this dataset D3. The results are shown in Table 4.

One may expect many sentences in the PA dataset to be subjective, full of opinions and sentiments, since these sentences are related to the evaluation of one person's work by another person. However, that does not quite seem to be the case - only 46% sentences are marked as SUBJECTIVE by annotators. Further, the human annotators seem to be using a somewhat broader notion of subjectivity here. Note that the annotators have marked 135 sentences as SUBJECTIVE whereas the classifier has predicted only 98 sentences as SUBJECTIVE. Moreover, out of 135 sentences marked as SUBJECTIVE by annotators, the classifier has marked 61 as NON-SUBJECTIVE.

The main observations from an analysis of the results are as follows. First, many sentences are suggestions or recommendations made by the supervisors, which might be inherently considered as subjective (A1, A2 in Table 5). Human annotators have marked 52 sentences as suggestions, out of which they have marked 48 as SUBJECTIVE. On the other hand, the classifier identified only 19 sentences (out of these 52) as SUBJECTIVE, indicating that it is weak in identifying suggestions. This may be because the training dataset (D2) did not have too many suggestions (the sentences are from movie reviews). Thus we might improve the accuracy by adding a Boolean feature, regarding whether a sentence is a suggestion or not. Work such as (Pawar et al., 2015) may be useful to compute this feature automatically.

Second, several sentences contain ambiguous fragments, such as `very good team player`, `very large teams`, `constantly engaged`, `the right stakeholders` etc.; see also sentences A3, A4 in Table 5. These text fragments are ambiguous in the sense that the extent of quantification is not clear, or may even be impossible. However, human annotators seem to be tolerant of such ambiguities, and many sentences which they have marked as NON-SUBJECTIVE contain such ambiguous fragments. However, the classifier has marked many such sentences as SUBJECTIVE, indicating that the role of ambiguous text in subjective sentences needs to be explored more from a linguistic perspective.

### 5.2 Scientific Abstracts

Factual observations and objective descriptions are important in scientific literature, and hence it is expected that it would not contain much subjectivity. To validate this hypothesis, we downloaded 1440 abstracts for biomedical literature taken from Medline (`www.ncbi.nlm.nih.gov/pubmed`). Total number of sentences in these abstracts is 18261. Average number of words per sentence is 19 and the average number of words per abstract is 239. Our dataset D4 consists of 613 sentences taken randomly from these abstracts and labeled manually. There were 281 (46%) SUBJECTIVE sentences and 332 (54%) NON-SUBJECTIVE sentences. Some of the SUBJECTIVE sentences are shown in Table 6. Table 4 shows the results for various classifiers on this dataset D4. Stacking2 classifier seem to give the best results. We then applied the Stacking3 classifier (trained on Dataset D2 where it gives the best results) to all the 18261 (unlabeled) sentences in the abstracts. It classified 3532 (19.3%) sentences as SUBJECTIVE. Thus preliminary results show that there is a significant prevalence of SUBJECTIVE sentences in scientific literature, contrary to popular belief. We now look closely at these SUBJECTIVE sentences.

Some sentences seem to be comments, judgements, opinions or evaluations (of some experiments, say), which are inherently subjective; see B1, B4, B7. Suggestions are another class of sentences often marked as SUBJECTIVE by our classifier; see B3, B4. The third class of marked SUBJECTIVE sentences is about conclusions or predictions; see B5, B6. Some sentences, such as B8, seem to be classified as SUBJECTIVE because of the use of ambiguous expressions such as `possible`, `common`, `potential`. A common scenario where the classifier seems to be going wrong is when sentences (e.g., B2) include words like `significant`, which actually refer to rigorous notion of statistical significance and not to any vague ideas about significance of something.

In the PA dataset D3, we naturally expect a high fraction of sentences to be SUBJECTIVE;D3 had 46% SUBJECTIVE sentences. Surprisingly, the scientific abstracts also seem to contain a fairly high fraction (19.3%) of SUBJECTIVE sentences. We have also observed that the context and domain play an important role in terming a sentences as SUBJECTIVE or NON-SUBJECTIVE. A single sentence can be termed as SUBJECTIVE when it is taken as an isolated sentence. However the context, i.e. accompanying sentences and domain, can be used to interpret the sentence as NON-SUBJECTIVE. For example, B1 is marked as SUBJECTIVE as an isolated sentence. The paragraph in which it appears is as follows: `The mucocele decreased in size and the postoperative course was uneventful. No recurrence was observed at 6 months' follow-up`. A possible reasons why this sentence is marked as subjective is because of vague word like `uneventful`. The next sentence provides a more objective basis, which can make this sentence NON-SUBJECTIVE.

## 6 Conclusion

Subjective sentences describe people's opinions, points-of-view, interpretations, comparisons, sentiments, judgments, appraisals or feelings toward entities, events and their properties. Identifying subjective sentences is the basis for opinion mining and sentiment analysis, and is important in practical applications such as political position analysis, social media analytics for marketing and product review analytics. Identifying subjective sentences is particularly challenging for neutral sentences i.e., when there is no particular sentiment expressed. Sentimentally neutral, but still subjective, sentences occur in many practical documents such as scientific papers, patents, financial reports and news. The notion of subjective expression or subjective communication is also crucial in philosophy of art. While there is much work in identifying sentiments and their polarity, there is relatively less work in identifying subjective sentences. In this paper, we used a variety of standard classifiers to build models of SUBJECTIVE vs. NON-SUBJECTIVE sentences. We demonstrated that some of the standard classifiers outperform the approaches reported in the literature on several interesting datasets. We discussed two novel applications of this work, which have not been reported in the literature to our knowledge: understanding the prevalence of subjective sentences in (i) performance appraisal text; and (ii)

| ID | Actual | Predicted | Sentence |
|---|---|---|---|
| A1 | S | NS | The assets can be extended to larger infra led offerings and cloud migration offerings. |
| A2 | S | NS | He also needs to accept and bring changes in his team to suit evolving needs of the account. |
| A3 | NS | S | Is able to effectively collaborate across the organisational groups for achieving the desired objective. |
| A4 | NS | S | Being well organised and having done multiple roles in xyzxyz - Delivery, Presales, Domain Consulting - he can run a small mid size unit quite comfortably. |

Table 5: Some example sentences from Dataset D3.

| | |
|---|---|
| B1 | The mucocele decreased in size and the postoperative course was uneventful. |
| B2 | Significant differences were detected in bacterial community structures and co-occurrence patterns between the wet and dry seasons. |
| B3 | New epidemiological and genetic studies are needed to identify possible common risk factors. |
| B4 | Findings suggest that specific types of stressors may influence eating behaviors differently. |
| B5 | This approach should be of special interest to those concerned about the impact of the presence of low-volatility organic liquids in waters of environmental and biological systems. |
| B6 | Hence, this may provide a new insight into understanding the mechanism of DR pathogenesis, as well as a potential therapeutic target for proliferative DR. |
| B7 | Some patients may experience a short term pain response. |
| B8 | This review is an up-to-date compilation on its traditional uses in context to phytochemical and pharmacological perspectives. |

Table 6: Examples sentences from scientific abstracts.

scientific papers. Rather surprisingly, we found that scientific papers also seem to contain a substantial fraction of subjective sentences. Finally, we compared the nature of subjective sentences in the human-centric text (such as performance appraisals) and scientific papers, and reported that there seem to be *different* reasons why some sentences in these domains are subjective.

For further work, we are developing additional techniques for subjectivity detection, based on co-training and label propagation. We are also exploring detection of subjectivity in different domains, such as financial reports and political news. It is already known, and we have also found in our work, that there is substantial variation and disagreement in the human annotators' perception of subjective sentences. That is, the basis of subjectivity may itself be subjective! It would be interesting to explore the nature of subjective expressions by taking into account context and human psychological factors. To get a deeper understanding of the notion of subjectivity, we propose the need to further investigate the linguistic and semantic basis for subjective sentences, and their variations across different domains.

## References


M. Apte, S. Pawar, S. Patil, S. Baskaran, A. Shrivastava, and G.K. Palshikar. 2016. Short text matching in performance management. In *Proc. 21st Int. Conf. on Management of Data (COMAD)*, pages 13–23.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. LREC*, pages 2200–2204.

W.W. Cohen, V.R. Carvalho, and T.M. Mitchell. 2004. Learning to classify email into "speech acts". In *Proc. Empirical Methods in Natural Language Processing (EMNLP-04)*, pages 309–316.

S. Corston-Oliver, E. Ringger, M. Gamon, and R. Campbell. 2004. Task-focused summarization of email. In *Proc. ACL-04 Workshop on Text Summarization Branches Out*, pages 43–50.

Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. 2006. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7:551–585.

S. Deshpande, G.K. Palshilkar, and G. Athiappan. 2010. An unsupervised approach to sentence classification. In *Proc. Int. Conf. on Management of Data (COMAD)*, pages 88–99.

Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In *Proc. 23rd Int. Conf. Computational Linguistics*, pages 340–348.

Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs. In *Proc. ICWSM*, pages 219–222.

B. Hachey and C. Grover. 2004. Sentence classification experiments for legal text summarisation. In *Proc. 17th Annual Conference on Legal Knowledge and Information Systems (Jurix-2004)*, pages 29–38.

Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proc. 18th Int. Conf. Computational Linguistics*, pages 299–305.

Q. He, K. Chang, and E.-P. Lim. 2006. Anticipatory event detection via sentence classification. In *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics (SMC-06)*, pages 1143–1148.

M. A. Hearst. 1998. Automated discovery of wordnet relations. In *in C. Fellbaum (Ed.), WordNet: An Electronic Lexical Database and Some of its Applications*. MIT Press.

F. Ibekwe-SanJuan, S. Fernandez, E. SanJuan, and E. Charton. 2008. Annotation of scientific summaries for information retrieval. In *Proc. ESAIR-08*.

T. Ito, M. Shimbo, T. Yamasaki, and Y. Matsumoto. 2004. Semi-supervised sentence classification for medline documents. *IEIC Technical Report*, 104(486):51–56.

E. Ivanovic. 2006. Using dialogue acts to suggest responses in support services via instant messaging. In *Proc. 2006 Australasian Language Technology Workshop (ALTW2006)*, pages 159–160.

Y. Kadoya, K. Morita, M. Fuketa, M. Oono, E.-S. Atlam, T. Sumitomo, and J.-I. Aoe. 2005. A sentence classification technique using intention association expressions. *Int. J. Computer Mathematics*, 82(7):777–792.

A. Khoo, Y. Marom, and D. Albrecht. 2006. Experiments with sentence classification. In *Proc. 2006 Australasian Language Technology Workshop (ALTW2006)*, pages 18–25.

Soo-Min Kim and Eduard Hovy. 2005. Automatic detection of opinion bearing words and sentences. In *Proc. Int. Joint Conf. Natural Language Processing (IJCNLP)*, pages 61–66.

Youngho Kim and Sung-Hyon Myaeng. 2007. Opinion analysis based on lexical clues and their expansion. In *Proc. NTCIR-6*, pages 347–354.

Bing Liu. 2010. Sentiment analysis and subjectivity. In *Handbook of natural language processing, vol. 2*, pages 627–666.

L. McKnight and P. Srinivasan. 2003. Categorization of sentence types in medical abstracts. In *Proc. American Medical Informatics Association Annual Symposium*, pages 440–444.

S. Momtazi and D. Klakow. 2009. Language model-based sentence classification for opinion question answering systems. In *Proc. International Multiconference on Computer Science and Information Technology*, pages 251–255.

M. Naughton, N. Stokes, and J. Carthy. 2010. Sentence-level event classification in unstructured texts. *Information Retrieval*, 13(2):132–156.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. LREC*, pages 1320–1326.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. 42nd annual meeting on Association for Computational Linguistics*, page 271.

S. Pawar, N. Ramrakhiyani, G. K. Palshikar, and S. Hingmire. 2015. Deciphering review comments: Identifying suggestions, appreciations and complaints. In *Proc. 20th Int. Conf. Applications of Natural Language to information Systems (NLDB), LNCS 9103*, pages 204–211.

N. Ramrakhiyani, S. Pawar, G.K. Palshikar, and M. Apte. 2016. Aspects from appraisals! a label propagation with prior induction approach. In *Proc. 21st Int. Conf. Applications of Natural Language to information Systems (NLDB), LNCS 9612*, pages 301–309.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proc. Conf. on Empirical methods in natural language processing (EMNLP)*, pages 105–112.

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proc. 7th Conf. Natural Language Learning (HLT-NAACL)*, pages 25–32.

Adam Stepinski and Vibhu Mittal. 2007. A fact/opinion classifier for news articles. In *Proc. 30th ACM Conf. Research and development in information retrieval (SIGIR)*, pages 807–808.

Afraz Z. Syed, Muhammad Aslam, and Ana Maria Martinez-Enriquez. 2010. Lexicon based sentiment analysis of urdu text using sentiunits. In *Proc. Mexican Int. Conf. on Artificial Intelligence*, pages 32–43.

S. Teufel and M. Moens. 1998. Sentence extraction and rhetorical classification for flexible abstracts. In *Proc. AAAI-98*.

C. Wang, J. Lu, and G. Zhang. 2005. A semantic classification approach for online product reviews. In *Proc. IEEE/WIC/ACM International Conference on Web Intelligence*, pages 276–279.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proc. Int. Conf. Intelligent Text Processing and Computational Linguistics*, pages 486–497.

Janyce M. Wiebe, Rebecca F. Bruce, and Thomas P. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proc. 37th annual meeting of the Association for Computational Linguistics*, pages 246–253.

Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proc. AAAI/IAAI*, pages 735–740.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005a. Opinionfinder: A system for subjectivity analysis. In *Proc. Conf. on Human Language Technology and Empirical Methods in Natural Language Processing interactive demonstrations*, pages 34–35.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005b. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. Conf. on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354.

Y. Yamamoto and T. Takagi. 2005. Experiments with sentence classification: A sentence classification system for multi biomedical literature summarization. In *Proc. 21st International Conference on Data Engineering Workshops*, pages 1163–1168.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proc. Conf. on Empirical methods in natural language processing (EMNLP)*, pages 129–136.

F.M. Zanzotto and L. DellArciprete. 2009. Efficient kernels for sentence pair classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*, pages 91–100.

L. Zhou, M. Ticrea, and E. Hovy. 2004. Multi-document biography summarization. In *Proc. Empirical Methods in Natural Language Processing (EMNLP-04)*, pages 434–441.