

Sentiment Classification across Domains

Dinko Lambov¹, Gaël Dias¹, Veska Noncheva²

¹ Centre of Human Language Technology and Bioinformatics, University of Beira Interior,
Covilhã, Portugal

² Plovdiv University Paisii Hilendarski, Plovdiv, Bulgaria
d_lambov@mail.bg, ddg@di.ubi.pt, wesnon@uni-plovdiv.bg

Abstract. In this paper we consider the problem of building models that have high sentiment classification accuracy without the aid of a labeled dataset from the target domain. For that purpose, we present and evaluate a novel method based on level of abstraction of nouns. By comparing high-level features (e.g. level of affective words, level of abstraction of nouns) and low-level features (e.g. unigrams, bigrams), we show that, high-level features are better to learn subjective language across domains. Our experimental results present accuracy levels across domains of 71.2% using SVMs learning models.

1 Introduction

Over the past years, there have been an increasing number of publications focused on the detection and classification of sentiment and subjectivity in texts. Detection of sentiment in text offers enormous opportunities for various applications. It is likely to provide powerful functionality for competitive analysis and marketing analysis through topic tracking and detection of unfavorable rumors.

However, as stated in [1][2][3][4], most research have focused on the construction of models within particular domains and have shown difficulties in crossing domains. But, if sentiment is orthogonal to topic, we should be able to build classifiers that perform well on topics other than the one used to build classifier. As a consequence, our aim in constructing a classifier is to maximize accuracy across topics. For that purpose, we propose to use high-level features (e.g. level of affective words, level of abstraction of nouns) rather than low-level features (e.g. unigrams, bigrams) to learn a model of subjectivity which may apply to different domains: movie reviews, newspaper articles and automatically annotated texts downloaded from Wikipedia and Web Blogs. Since sentiment in different domains can be expressed in different ways [1][5][6], supervised classification techniques require large amount of labeled training data. However, the acquisition of these labeled data can be time-consuming and expensive. Moreover, most training sets are built for English and little has been done for other languages. From that assumption, we propose to automatically produce learning data from web resources. To do so, we propose to compare Wikipedia and Web Blogs texts to reference objective and subjective corpora and show that Wikipedia texts are representative of objectivity and Web Blogs are representative of subjectivity. As a consequence, learning data will be easy to build for many languages

without manual annotation alleviating intensive and time-consuming labor. Our methodology uses state-of-the-art characteristics that have been used to classify opinionated texts and proposes a new feature to classify sentiment texts, based on the level of abstraction of nouns. Before classification is performed, feature selection and visualization are performed to evaluate how well the datasets can be represented in the given space of characteristics. Finally, an exhaustive evaluation shows that (1) the level of abstraction of nouns is a strong clue to identify subjective texts which crosses domains, (2) high-level features allow cross-domain learning models and (3) automatically labeled dataset extracted from Wikipedia and Web Blogs get near, on average, to the best cross-domains classifiers reaching accuracy levels of 69.3% compared to 71.2% for manually annotated texts.

2 Related Work

The subjectivity and polarity¹ of language has been investigated at some length. Many features have been used to characterize opinionated texts at different levels (e.g. words [6], sentences [7] and texts [7][8][9][10]). In this section, we propose to enumerate works which focus on document.

At document level, [8] show that the unigram model with Support Vector Machines (SVM) reaches best results compared to more complex models in the domain of movie reviews. [7] propose a similar experiment based on Multiple Naïve Bayes classifiers. [8] propose a unigram SVM classifier which is applied to just subjective portions of a document. A sentence cut-based classifier is used to identify subjective parts in texts which are then used for text classification. [9] derive a variety of subjectivity cues (frequencies of unique words in subjective-element data; collocations with one or more positions filled by a unique word; distributional similarity of adjectives and verbs) from corpora and demonstrate their effectiveness on classification tasks. They determine a relationship between low frequency terms and subjectivity and find that their method for extracting subjective n-grams is enhanced by examining those that occur with unique terms. Finally, [10] present a method using verb class information, and an online resources, the Wikipedia dictionary, for determining the polarity of adjectives. They use verb-class information in the sentiment classification task, since exploiting lexical information contained in verbs has shown to be a successful technique for classifying documents.

Other research in the sentiment classification field regards cross-domain classification. How can we learn classifiers on one domain and use them on another domain? Tests have been done by [1][2][3][4]. Overall, they show that sentiment analysis is a domain-specific problem, and it is hard to create a domain independent classifier. One possible approach is to train the classifier on a domain-mixed set of data instead of training it on one specific domain [2][3][4]. Another possibility is to propose high-level features which do not depend so much on topics such as part-of-speech statistics as in [2]. In this case, the part-of-speech representation does not reflect the topic of the document, but rather the type of text used in the document. Just

¹ Most papers deal with polarity as the essence of subjectivity. Although, we will see that subjectivity can be expressed in different ways.

by looking at part-of-speech statistics, improved results can be obtained comparatively to unigram models (low-level models) when trying to cross domains. [1] use pivot features based on domain mutual information to relate training and target domains. They also measure domain adaptability by estimating the classification accuracy loss by adapting one domain to another without the use of a labeled target dataset.

3 Characterizing Subjectivity

Subjectivity can be expressed in different ways as summarized in [4] who identify the following dimensions: evaluation (positive or negative), potency (powerful or unpowerful), proximity (near or far), specificity (clear or vague), certainty (confident or doubtful) and identifiers (more or less), direct expressions, elements of actions and remarks. Based on these assumptions, our methodology aims at classifying texts at the subjectivity level (i.e. subjective vs. objective and not (positive, negative) vs. objective nor positive vs. negative) based on high-level features which can apply to different domains.

3.1 High-level features

Intensity of Affective Words: In most of previous works, sentiment expressions mainly depend on some words which can express subjective sentiment orientation. [11] have used a set of words extracted from WordNet Affect [12] to annotate the emotions in a text simply based on the presence of words from the WordNet Affect lexicon. WordNet Affect contains all information about the affective domain labels and the kind of mapping between Princeton WordNet 1.6 synsets and their corresponding affective domains. For example “*horror*” and “*hysteria*” express negative fear, “*enthusiastic*” expresses positive emotion, “*glad*” expresses joy. So, we propose to evaluate the level of affective words in texts as shown in Equation 1.

$$K_1 = \frac{\text{total affective words in text}}{\text{total words in text}}. \quad (1)$$

Dynamic Adjectives and Semantically Oriented Adjectives: [13] consider two features for the identification of opinionated sentences: (1) semantic orientation, which represents an evaluative characterization of a word's deviation from the norm for its semantic group and (2) dynamic adjectives which characterizes a word's ability to express a property in varying degrees. They noted that all sets involving dynamic adjectives and adjectives with positive or negative polarity are better predictors of subjective sentences than the class of adjectives as a whole. Here, we use the set of all adjectives identified in a reference corpus i.e. the set of dynamic adjectives, manually identified by [13] and the set of semantic orientation labels assigned as in [6]. So, we propose to evaluate the level of these adjectives in texts as shown in Equation 2.

$$K_2 = \frac{\text{total specific adjectives in text}}{\text{total adjectives in text}}. \quad (2)$$

Classes of Verbs: [10] present a method using verb class information. Their verb classes express objectivity and polarity. To obtain relevant verb classes, they use InfoXtract [14], an automatic text analyzer which groups verbs according to classes that often correspond to their polarity. As InfoExtract is not freely available, we reproduce their methodology by using the classification of verbs available in Levin’s English Verb Classes and Alternations [15]. So, we propose to evaluate the level of each class of verbs (i.e. conjecture, marvel, see and positive) in texts as in Equation 3.

$$K_3 = \frac{\text{total specific verbs in text}}{\text{total verbs in text}}. \quad (3)$$

Level of Abstraction of Nouns: There is linguistic evidence that level of generality is a characteristic of opinionated texts, i.e. subjectivity is usually expressed in more abstract terms than objectivity [4][16]. Indeed, descriptive texts tend to be more precise and more objective and as a consequence more specific. [4] define specificity as “[...] the extent to which a conceptualized object is referred to by name in a direct and clear way; or is only implied, suggested, alluded to, generalized, or otherwise hinted at”. In other words, a word is abstract when it has few distinctive features and few attributes that can be pictured in the mind. One way of measuring the abstractness of a word is by the hypernym relation in WordNet [17]. In particular, a hypernym metric can be the number of levels in a conceptual taxonomic hierarchy above a word (i.e. superordinate to). For example, “chair” (as a seat) has 7 hypernym levels: chair => furniture => furnishings => instrumentality => artifact => object => entity. So, a word having more hypernym levels is more concrete than one with fewer levels. So, we propose to evaluate the hypernym levels of all the nouns in texts as in Equation 4².

$$K_4 = \frac{\text{total hypernym levels for all nouns in text}}{\text{total nouns in text}}. \quad (4)$$

3.2 Low-level features

Perhaps the most common set of features used for text classification tasks is information regarding the occurrence of words or word ngrams, in the text. A vast

² Calculating the level of abstraction of nouns should be preceded by word sense disambiguation. Indeed, it is important that the correct sense is taken as a seed for the calculation of the hypernym level in WordNet. However, in practice, taking the most common sense of each word gave similar results as taking all the senses on average. As a consequence, we believe that word sense disambiguation can be avoided for this task.

majority of text classification systems treat documents as simple “bags-of-words” and use the word counts or presence as features. In this paper, the surface features we use are lemma unigrams and lemma bigrams associated to their TFIDF weights where w_{ij} is weight of term t_j in document d_i , N is the total number of documents in collection, and n is number of documents where term t_j occurs at least once (See Equation 5).

$$w_{ij} = \text{tf}_{ij} * \log_2 \frac{N}{n}. \quad (5)$$

4 Feature Selection and Visualization

Before performing any classification task, it is useful to evaluate to what extent the given high-level features are discriminative and allow representing distinctively the datasets in the given space of characteristics. For that purpose, we propose to do feature selection by applying the Wilcoxon rank-sum test and to visualize the datasets using multidimensional scaling.

4.1 Corpora

To perform these experiments, we used three manually annotated standard corpora and built one corpus based on Web resources and automatically annotated.

The Multi-Perspective Question Answering (Mpqa) Opinion Corpus³ contains 10,657 sentences in 535 documents from the world press on a variety of topics. All documents in the collection are marked with expression-level opinion annotations. The documents are from 187 different news sources in a variety of countries and date from June 2001 to May 2002. The corpus has been collected and manually annotated with respect to subjectivity as part of the summer 2002 NRRC Workshop on Multi-Perspective Question Answering. Based on the work done by [8] who propose to classify texts based only on their subjective/objective parts, we built a corpus of 100 objective texts and 100 subjective texts by randomly selecting sentences containing only subjective or objective phrases. This case represents the “ideal” case where all the sentences in texts are either subjective or objective.

The second corpus (Rotten/Imdb) is the subjectivity dataset v1.0⁴ which contains 5000 subjective and 5000 objective sentences collected from movie reviews data [8]. To gather subjective sentences, [8] collected 5000 movie review snippets from www.rottentomatoes.com. To obtain (mostly) objective data, they took 5000 sentences from plot summaries available from the Internet Movie Database www.imdb.com. Similarly to what we did for the MPQA corpus, we built a corpus of 100 objective texts and 100 subjective texts by randomly selecting 50 sentences containing only subjective or objective phrases.

³ <http://www.cs.pitt.edu/mpqa/>

⁴ <http://www.cs.cornell.edu/People/pabo/movie-review-data/>

The third corpus (Chesley) has been developed by [10] who manually annotated a dataset of objective and subjective documents⁵. It contains 496 subjective and 580 objective documents. Objective feeds are from sites providing content such as world and national news (CNN, NPR, etc.), local news (Atlanta Journal and Constitution, Seattle Post-Intelligencer, etc.), and various sites focused on topics such as health, science, business, and technology. Subjective feeds include content from newspaper columns (Charles Krauthammer, E. J. Dionne, etc.), letters to the editor (Washington Post, Boston Globe, etc.), reviews (dvdverdict.com, rottentomatoes.com, etc.), and political blogs (Powerline, Huffington Post, etc.).

The fourth corpus is the Wiki/Blog. We downloaded part of the static Wikipedia dump archive⁶ and automatically spidered Web Blogs from different domains. The final corpus contains 200 Mb of downloaded articles from Wikipedia and 100 Mb of downloaded texts from different Web Blogs. These texts are in English and cover many different topics. As we propose to compare Wikipedia and Web Blogs texts to reference objective and subjective corpora and show that Wikipedia texts are representative of objectivity and Web Blogs are representative of subjectivity, we automatically label texts as objectives if they come from Wikipedia and subjective if they come from Web Blogs.

4.2 Wilcoxon Rank-Sum Test

We use the Wilcoxon test to show that the comparative results are statistically significant. The Wilcoxon rank-sum test is a nonparametric alternative to the two sample t-test used to compare the locations of two populations i.e. to determine if one population is shifted with respect to another. The two sample Wilcoxon test with one-sided alternative is carried out for all experiments. The samples contain 200 values for each of the sets (100 objective texts and 100 subjective) and the exact p-value is computed. If the result of the test shows that the values of subjective (resp. objective) data are shifted to the right of the values of objective (resp. subjective) for all of the sets, we can say that the distributions of subjective and objective differ by a positive shift. The estimations for the differences of the location parameters are also computed for each of the sets. The exact 95% confidence interval for the difference of the location parameters of each of the sets is obtained by the algorithm described in [18] for which the Hodges-Lehmann estimator is employed. As a consequence, for each of the sets, we are 95% confident that the interval contains the actual difference between the features values of subjective and objective texts. The observed results are consistent with the hypothesis that most of the high-level features exposed in section 3 (intensity of affective words, dynamic and semantically oriented adjectives, classes of verbs and level of abstraction of nouns) have good discriminative properties for subjectivity/objectivity identification. As illustrated in Table 1, we can see that only the level of positive verbs does not significantly separate the objective sample from the subjective one over training corpora. This is mainly due to the fact that positive verbs do not occur frequently in texts thus biasing the statistical test. As a

⁵ <http://www.tc.umn.edu/~ches0045/data/>

⁶ <http://download.wikimedia.org/enwiki/>

consequence, we discarded this feature from our classification task. We can also see that Chesley’s dataset shows an uncharacteristic behavior. This is mainly due to the fact that Chesley’s texts are either objective or, positive or negative. Unlike the other corpora, Chesley’s corpus focuses on subjectivity exclusively based on polarity.

Table 1. Computed p-values using Wilcoxon test

Corpus:	Mpqa	Rotten/Imdb	Chesley	Wiki/Blog
Feature:				
Affective words	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Dynamic adj.	< 0,0001	< 0,0001	0,014	< 0,0001
Semantic adj.	< 0,0001	< 0,0001	0,045	< 0,0001
Conjecture verbs	0,00024	< 0,0001	0,021	< 0,0001
Marvel verbs	< 0,0001	< 0,0001	0,44	< 0,0001
See verbs	< 0,0001	< 0,0001	0,006	< 0,0001
Positive verbs	< 0,0001	0,00011	0,075	0,00061
Level of abstraction	0,003	< 0,0001	< 0,0001	< 0,0001

4.3 Multidimensional Scaling

In this subsection, we propose a visual analysis of the distribution of the datasets in the space of high-level features.

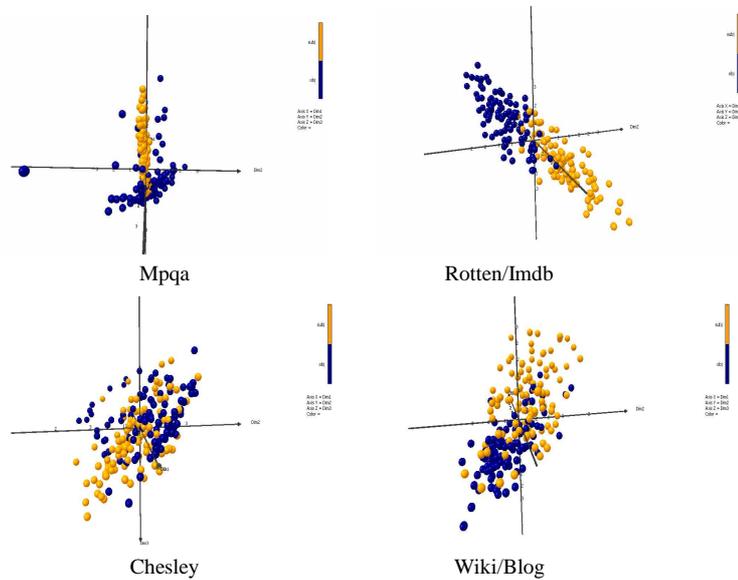


Figure 1: MDS over all training datasets

The goal of this study is to give a visual interpretation of the data distribution to assess how well classification may perform. If objective and subjective texts can be represented in a distinct way in the reduced space of features, one may expect good

classification results. To perform this study, we use a Multidimensional Scaling (MDS) process which is a traditional data analysis technique. MDS [19] allows to displaying the structure of distance-like data into an Euclidean space. In practice, the projection space we build with the MDS from such a distance is sufficient to have an idea about whether data is organized into classes or not. For our purpose, we perform the MDS process on all corpora trying to visualize subjective texts from objective ones. The obtained visualizations (Figures 1) show distinctly that a particular data organization can be drawn from the data. This visualization clearly shows that exclusively objective and subjective texts (i.e. Mpqa and Rotten/Imdb) may lead to improved results as data is well separated in the reduced 3-dimension space. In the case of Chesley's corpus and Wiki/Blog separating data in the space seems more difficult. Indeed, as these texts are not composed exclusively of subjective or objective sentences, the overlap in the space is inevitable. As [9] state, 75% (resp. 56%) of the sentences in subjective (resp. objective) texts are subjective (resp. objective)⁷. However, a pattern in the space seems to emerge comforting us in the choice of our high-level features for our classification task.

5 Experiments

In this section, we report the results of machine learning experiments for which we used Support Vector Machines (SVMs) to learn models of subjectivity over four different domains. SVMs have consistently shown to perform better than other classification algorithms for topic text classification in general [20], and for sentiment classification in particular [8]. All experiments have been performed on a leave-one-out 5 cross validation basis. In particular, we used Joachim's SVMlight package⁸ [20] for training and testing with SVM. As part-of-speech tagger we used the MontyTagger module of the free, common sense-enriched Natural Language Understander for English MontyLingua⁹ [21].

5.1 Results for In-Domain Data

In order to evaluate the difference between high level features with low level ones, we performed a comparative study on our four test sets. For the high level features, we took into account 7 features (affective words, dynamic and semantically oriented adjectives, conjuncture verbs, see verbs, marvel verbs and level of abstraction of nouns)¹⁰. For the Unigram and Bigram models, we used all the lemmas inside the corpora withdrawing their stop words. The results of these experiments are shown in Table 2 and 3 and evidence an important gain with low level features compared to high level features. Indeed, as it has already been stated in other works, objective language and subjective language hardly intersect which means that one word,

⁷ These data are extracted from the Wall Street Journal.

⁸ <http://svmlight.joachims.org/>

⁹ <http://web.media.mit.edu/~hugo/montylingua/>

¹⁰ An exhaustive study of the level of abstraction is made in the next section.

specific to one domain (i.e. which does not represent any subjective value outside it) can easily distinguish between objective and subjective texts.

Table 2: Results for High Level Features – In Domain (in percentage)

		All	Subjective		Objective	
		Accuracy	Precision	Recall	Precision	Recall
High-Level Features	Mpqa	71,2	81,1	55,2	66	87,2
	Rotten/Imdb	86,8	84,3	90,4	89,6	83,2
	Chesley	64,4	64,8	62,8	63,9	66
	Wiki/Blog	76,2	93,9	56	68,6	96,4

Table 3: Results for Low Level Features – In Domain (in percentage)

		All	Subjective		Objective	
		Accuracy	Precision	Recall	Precision	Recall
Unigrams	Mpqa	80,6	100	61,2	72	100
	Rotten/Imdb	97	97,9	96	96	98
	Chesley	72,6	64,7	99,2	98,3	46
	Wiki/Blog	94,6	90,2	100	100	89,2
Bigrams	Mpqa	67,8	100	35,6	64,6	100
	Rotten/Imdb	98	97,9	96	96,1	100
	Chesley	70,6	62,9	100	100	41,2
	Wiki/Blog	80	74,7	90,8	88,3	69,2

5.2 Results for Cross-Domain Data

In order to test models across domains, we propose to train different models based on one domain only at each time and test the classifiers over all domains together. Before doing the experiments it is important to visualize the data to understand how difficult the task may be. Indeed, when gathering texts from different domains, the best case is when all subjective texts represent one unique cloud in the space which does not intersect with a unique cloud formed by objective texts. However, it is not usually the case. In Figure 2, we show two different situations which illustrate that positive results may be obtained (e.g. Rotten/Imdb and Wiki/Blogs) and other cases where results may not be expected (e.g. Rotten/Imdb and Mpqa) when using high level features to characterize each text. Indeed, in the first picture in Figure 2 (Rotten/Imdb and Wiki/Blogs), we can see that subjective texts (blue and red) and objective texts (green and yellow) form two clouds separated by the axe *Dim 1* although many blogs (blue) are spread over the space. Unlikely, the second picture (Mpqa and Rotten/Imdb) evidences four clouds with almost no intersection which clearly indicates that subjectivity can be expressed in different ways from domain to domain, thus complicating the learning process. We do not present the visualization results based on the vector space model representation of texts (i.e. low level features) as they evidence highly intersected clouds. As a consequence, unreliable results may be expected for low level feature across domains. So, in Table 4 and 5, we present the results for domain transfer for high level features and low level features. Each percentage can be expressed as the average results over all datasets. Best results

overall are obtained for high level features for the Chesley dataset with accuracy of 71.2% .

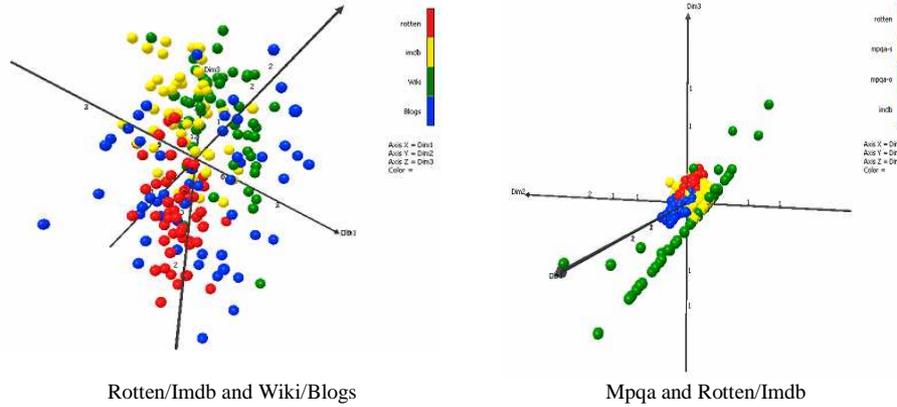


Figure 2: MDS over mixed datasets

Table 4: Results for High Level Features – Cross Domain (in percentage)

		All	Subjective		Objective	
		Accuracy	Precision	Recall	Precision	Recall
High-Level Features	Mpqa	58,8	47,7	81,9	33,9	35,8
	Rotten/Imdb	68,9	81	49,7	66,1	88,1
	Chesley	71,2	79,4	59,6	68	82,8
	Wiki/Blog	69,3	76,5	61,8	71,1	76,8

Table 5: Results for Low Level features – Cross Domain (in percentage)

		All	Subjective		Objective	
		Accuracy	Precision	Recall	Precision	Recall
Unigrams	Mpqa	53,8	25,9	15,6	53,2	92
	Rotten/Imdb	63,9	66,2	28,6	63,3	99,3
	Chesley	59,9	68,9	97,3	46	22,5
	Wiki/Blog	61,1	60	100	25	22
Bigrams	Mpqa	54,4	25	8,9	53,6	100
	Rotten/Imdb	67,1	66,3	90,3	65,2	44
	Chesley	55	53,2	99,8	25	10,3
	Wiki/Blog	57,5	56,2	97,7	22	17,3

5.3 Cross-Domain and Level of Abstraction

In order to evaluate the importance of the level of abstraction of nouns as a clue for subjectivity identification, we propose to study the six state-of-the-art features without the level of abstraction of nouns and then compare with the full set of seven features.

Then, we present the importance of each class of features individually to assess how discriminative each class of features is. For that purpose, we defined four classes of features: affective words, adjectives (semantically oriented and dynamic), verbs (conjecture, marvel and see) and level of abstraction of nouns. The results are illustrated in Table 6 for leave-one-out 5 cross validation for cross-domain and evidence that the level of abstraction of nouns is the best feature to cross domains except for the case of the Mpqa dataset. This is mainly due to the over-evaluation of strong features by the SVM. Indeed, in Table 6, as affective words are the best feature for the Mpqa dataset, best results are obtained without the level of abstraction. In the other cases, the level of abstraction of nouns is the best feature thus implying best results with seven features.

Table 6: Results for High-Level Features – Cross Domain (in percentage)

	Mpqa	Rotten/Imdb	Chesley	Wiki/Blog
7 features	58,8	68,9	71,2	69,3
6 features	64,4	67,7	67,4	68,7
Affective words only	66,2	59,9	67,1	66,2
Adjectives only	60,1	67,2	65,3	68,5
Verbs only	64,8	69,3	59,9	68,1
Level of Abstraction only	51,7	70,9	72	71

6 Conclusions

Sentiment classification is a domain specific problem i.e. classifiers trained in one domain do not perform well in others. At the same time, sentiment classifiers need to be customizable to new domains in order to be useful in practice. In this paper, we proposed new experiments based on high-level features to learn subjective language across domains. Best results showed accuracy of 71.2% across domains. Indeed SVM steadily perform best for high-level features than for low-level features. Moreover, our experimental results show that using levels of abstraction of nouns as a feature leads to improved performance on subjectivity classification tasks. Also, the results produced via automatic construction of data sets (Wiki/Blogs) get near, on average, to the best cross-domains classifiers reaching accuracy levels of 69.3% compared to 71.2% for manually annotated texts. A direct application of this study is to automatically produce data sets for other languages than English and allow classification of multilingual opinionated texts. Indeed, building data sets for the classification of opinionated texts can be done automatically, at the document level, just by downloading Web Blogs for the subjective part and Wikipedia texts for the objective part. Thus labor-intensive and time-consuming work can be leveraged.

Acknowledgments. This work is supported by the VIPACCESS project funded by the Portuguese Agency for Research (Fundação para a Ciência e a Tecnologia) with the reference PTDC/PLP/72142/2006.

References

1. Blitzer, J., Crammer, K., Kulesza, A., Pereira, F. and Wortman, J: Biographies, Bollywood, Boom-boxes, and Blenders: Domain Adaptation for Sentiment Classification. In Proceedings of Association for Computational Linguistics (2007)
2. Finn, A. and Kushmerick, N.: Learning to Classify Documents According to Genre. J. American Society for Information Science and Technology, Special issue on Computational Analysis of Style, 57(9) (2006)
3. Aue, A. and Gamon, M.: Customizing Sentiment Classifiers to New Domains: a Case Study. In Proceedings of International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria (2005)
4. Boiy, E., Hens, P., Deschacht, K. and Moens, M-F.: Automatic Sentiment Analysis of Online Text. In Proceedings of the 11th International Conference on Electronic Publishing, Openness in Digital Publishing: Awareness, Discovery & Access, June 13-15 (2007)
5. Engström, C.: Topic Dependence in Sentiment Classification. Master Thesis. University of Cambridge (2004)
6. Hatzivassiloglou, V. and McKeown, K.R.: Predicting the Semantic Orientation of Adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics (1997)
7. Yu, Hong and Hatzivassiloglou, V.: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In EMNLP (2003)
8. Pang, B. and Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (2004)
9. Wiebe, J., Wilson, T., Bruce, R., Bell, M. and Martin, M.: Learning Subjective Language. In Computational Linguistics, 30(3), 277–308 (2004)
10. Chesley, P., Vincent, B., Xu, L. and Srihari, R.: Using Verbs and Adjectives to Automatically Classify Blog Sentiment. In Proceedings of AAAI Spring Symposium (2006)
11. Strapparava, C. and Mihalcea, R.: Learning to Identify Emotions in Text. In Proceedings of the Symposium on Applied Computing. 1556-1560 (2008)
12. Strapparava, C. and Valitutti, A.: WordNet-Affect: An Affective Extension of WordNet. In Proceedings of the Language Resources and Evaluation International Conference (2004)
13. Hatzivassiloglou, V. and Wiebe, J.: Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In Proceedings of International Conference on Computational Linguistics (2000)
14. Srihari, R., Li, W., Niu, C. and Cornell, T.: InfoXtract: A Customizable Intermediate Level Information Extraction Engine. In Proceedings of the HLT-NAACL 2003 Workshop on Software Engineering and Architecture of Language Technology Systems (2003)
15. Levin, B.: English Verb Classes and Alternations. University of Chicago Press (1993)
16. Osgood, C.E., Suci, G.J. and Tannebaum, P.H.: The Measurement of Meaning. University of Illinois Press (1971)
17. Miller, G.A.: Wordnet: A Lexical Database. In Communications of the ACM 38 (1995)
18. Bauer, D.F.: Constructing Confidence Sets Using Rank Statistics. In Journal of the American Statistical Association 67, 687–690 (1972)
19. Kruskal J. B. and Wish M.: Multidimensional Scaling. Sage Publications. Beverly Hills. CA (1977)
20. Joachims, T.: Learning to Classify Text Using Support Vector Machines. Dissertation, Kluwer (2002)
21. Liu., H.: Montylingua: An End-to-End Natural Language Processor with Common Sense. Available at: web.media.mit.edu/~hugo/montylingua (2004)