Language Independent Sentence-Level Subjectivity Analysis with Feature Selection

Aditya Mogadala
Search and Information Extraction Lab
IIIT-H
Hyderabad
India
aditya.m@research.iiit.ac.in

Vasudeva Varma
Search and Information Extraction Lab
IIIT-H
Hyderabad
India
vv@iiit.ac.in

Abstract

Identifying and extracting subjective information from News, Blogs and other user generated content has lot of applications. Most of the earlier work concentrated on English data. But, recently subjectivity related research at sentence-level in other languages has increased. In this paper, we achieve sentence-level subjectivity classification using language independent feature weighing and selection methods which are consistent across languages. Experiments performed on 5 different languages including English and South Asian language Hindi show that Entropy based category coverage difference criterion (ECCD) feature selection method with language independent feature weighing methods outperforms other approaches for subjective classification.

1 Introduction

Subjective text expresses opinions, emotions, sentiment and beliefs, while objective text generally report facts. So the task of distinguishing subjective from objective text is useful for many natural language processing applications like mining opinions from product reviews (M. Hu and B. Liu, 2004), summarizing different opinions(K. Ganesan.et.al, 2010), question answering (A. Balahur.et.al, 2009) etc.

But research work performed earlier on subjectivity analysis has been applied only on English and mostly at document-level and word-level. Some methods (Wiebe and Riloff, 2005) which concentrated at sentence-level to learn subjective and objective expressions are boot-strapping algorithms which lacks scalability. But, recently focus shifted to multilingual space (R. Mihalcea.et.al, 2007). Banea (C. Banea.et.al, 2008) worked on sentence-level subjectivity analysis using machine translation approaches by leveraging resources and tools available for English . Another approach (C. Banea.et.al, 2010) used multilingual space and meta classifiers to build high precision classifiers for subjectivity classification.

However, aforementioned work (C. Banea.et.al, 2008) concentrated more on language specific attributes due to variation in expression of subjectivity in different languages. This create a problem of portability of methods to different languages. Other approach (C. Banea.et.al, 2010) which tried achieving language independence created large feature vectors for subjectivity classification. Different languages parallel sentences are taken into consideration to build high-precision classifier for each language. This approach not only increases the complexity and time for classification but also completely dependent on parallel corpus to get good accuracies. A weakly supervised method (C. Lin.et.al, 2011) for sentence-level subjectivity detection using subjLDA tried to reduce training data is available only for English. There are some experiments conducted for Japanese (H. Kanayama.et.al, 2006), Chinese (T. Zagibalov.et.al, 2008), Romanian (C. Banea.et.al, 2008; R. Mihalcea.et.al, 2007) languages data. But these approaches are performed at document level and not language independent.

In this paper, we try to address three major problems highlighted from earlier approaches. First, can

26th Pacific Asia Conference on Language,Information and Computation pages 171–180

language portability problem be eliminated by selecting language independent features. Second, can language specific tools like POS taggers, Named Entity recognizers dependency can be minimized as they vary with language. Third, can accuracy of subjective classification is maintained after feature reduction using feature selection methods which are consistent across languages.

Remainder of this paper is organized into following sections. Related work is mentioned in the Section 2. Next Section 3 discuss about our approach for feature weighing and selection methods. While the experimental setup Section 4 describes collection and evaluation metrics used to analyze the accuracy of approach. Experimental Section 5 explains experiments performed on different languages, while results and performance between SVM and NBM is analyzed in Section 6 and Section 7 respectively. Conclusion and future work is discussed in Section 8.

2 Related Work

We divide the subjective and objective classification task into unsupervised, multilingual and supervised methods.

2.1 Unsupervised

Sentiment classification and opinion analysis can be considered as a hierarchical task of subjectivity detection. Improvement of precision in subjectivity detection can benefit the later. Therefore, lot of work is done for subjective sentence detection to achieve later. (G. Murray.et.al, 2009) proposed to learn subjective expression patterns from both labeled and unlabeled data using n-gram word sequences. Their approach for learning subjective expression patterns is similar to (T. Wilson.et.al, 2008) which relies on n-grams, but goes beyond fixed sequences of words by varying levels of lexical instantiation.

2.2 Multilingual

In the multilingual space good amount of work is done in Asian and European languages. Several participants in the Chinese and Japanese Opinion Extraction tasks of NTCIR-6 (Y. Wu.et.al, 2007) performed subjectivity and sentiment analysis in languages other than English. (C. Banea.et.al, 2008; R. Mihalcea.et.al, 2007) performed subjectivity analysis in Romanian. While, (C. Banea.et.al, 2010) performed subjectivity analysis in French, Spanish, Arabic, German, Romanian languages.

2.3 Supervised

Furthermore, tools developed for English were used to determine sentiment or subjectivity labeling for a given target language by transferring the text to English and applying an English classifier on the resulting data. The labels were then transfered back into the target language (M. Bautin.et.al, 2008). These experiments are carried out in Arabic, Chinese, English, French, German, Italian, Japanese, Korean, Spanish, and Romanian. (X. Wan, 2009) who constructs a polarity co-training system by using the multi-lingual views obtained through the automatic translation of product-reviews into Chinese and English.

3 Approach

Subjective sentence classification is treated as a text classification task. (C. Banea.et.al, 2008) used unigrams at word level as features to classify the subjective and objective sentences in different languages. But, unigrams can occur in different categories (subjective and objective) with equal probability. This hampers the classification accuracy. Also, selecting all possible words in the sentences can create a large index size when considered as entire training set increasing the dimensionality of the feature vector for each sentence.

Feature selection can be applied to efficiently categorize the sentences. It is an important process which is followed for many text categorization tasks. Now to achieve our major objective of language independent subjective classification. We use feature extraction, weighing and selection methods that are language independent.

3.1 Feature Extraction and Weighing

Features are categorized into syntactic, semantic, link-based, and stylistic features (G. Forman, 2003) from the previous subjective and sentiment studies. Here, we concentrate more on feature weighing methods based on syntactic and stylistic properties of the text to maintain language independence. Unigrams and Bigrams extracted as features are weighed as given below.

Syntactic Feature Weighing

Syntactic features used in earlier works (M. Gamon, 2004) where word n-grams and part-of-speech (POS) tags. But, POS tagging create dependency on language specific tools. In order to eliminate the language specific dependencies we will use only word n-grams.

Sentence Representation with Unigram (UF.ISF)

This feature extraction is inspired from vector space model (G. Salton, 1975) used for flat documents. **UF** represents the unigram frequency at word level in a sentence. While **ISF** represent the inverse sentence frequency of the unigram. For a given collection S of subjective and objective sentences, an Index $I = \{u_1, u_2, ..., u_{|I|}\}$, where $|I|$ denotes the cardinal of I, gives the list of unigrams u encountered in the sentences S.

A sentence s_i of S is then represented by a vector $\vec{s_i} = (w_{i,1}, w_{i,2}, ..., w_{i,I})$ followed by the subjective or objective label. Here, $w_{i,j}$ represents the weight of unigram u_j in the sentence s_i. Now to calculate the weight $w_{i,j}$ we use the formula similar to TF.IDF.

$$w_{i,j} = \frac{c_{i,j}}{\Sigma_l c_{i,l}} * log \frac{|S|}{|\{s_i : u_j \in s_i\}|} \quad (1)$$

where $c_{i,j}$ is the number of occurrences of u_j in the sentence s_i normalized by the number of occurrences of other unigrams in sentence s_i, $|S|$ is total number of sentences in the training set and $|\{s_i : u_j \varepsilon s_i\}|$ is number of sentences in which the unigram u_j occurs at-least once.

Sentence Representation with Bigram (BF.ISF)

This feature extraction is similar to UF.ISF mentioned in the earlier section, but we extract co-occurring words. **BF** represents the Bigrams frequency at word level in a sentence. While **ISF** represent the inverse sentence frequency of the Bigram. For a given collection S of subjective and objective sentences, an Bigram Index $BI = \{b_1, b_2, ..., b_{|BI|}\}$, where $|BI|$ denotes the cardinal of BI, gives the list of bigrams b encountered in the sentences S.

A sentence s_i of S is then represented by a vector $\vec{s_i} = (wb_{i,1}, wb_{i,2}, ..., wb_{i,BI})$ followed by the subjective or objective label. Here, $wb_{i,j}$ represents the weight of bigram b_j in the sentence s_i. Now to calculate the weight $wb_{i,j}$ we use the formula similar to UF.ISF.

$$wb_{i,j} = \frac{c_{i,j}}{\Sigma_l c_{i,l}} * log \frac{|S|}{|\{s_i : b_j \in s_i\}|} \quad (2)$$

where $c_{i,j}$ is the number of occurrences of b_j in the sentence s_i normalized by the number of occurrences of other bigrams in sentence s_i, $|S|$ is total number of sentences in the training set and $|\{s_i : b_j \varepsilon s_i\}|$ is number of sentences in which the bigram b_j occurs at least once.

Stylistic Feature Weighing

Structural and lexical style markers can be considered as stylistic features which has shown good results in Web discourse (A. Abbasi.et.al, 2008). However, style markers have seen limited usage in sentiment analysis research. Some (M. Gamon, 2004) tried in this direction.

Sentence representation with Normalized Unigram Word Length (NUWL)

This feature extraction considers length of unique unigram words in the sentence. Length of unigram is calculated by the number of characters present in the word. For a given collection S of subjective and objective sentences, an Word Index $WI = \{uw_1, uw_2, ..., uw_{|WI|}\}$, where $|WI|$ denotes the cardinal of WI, gives the list of unigram words uw encountered in the sentences S.

A sentence s_i of S is then represented by a vector $\vec{s_i} = (lw_{i,1}, lw_{i,2}, ..., lw_{i,I})$ followed by the subjective or objective label. Here, $lw_{i,j}$ represents the weight of unigram word uw_j in the sentence s_i. Now to calculate the weight $lw_{i,j}$.

$$lw_{i,j} = \frac{L_{i,j}}{\Sigma_n L_{i,n}} * log \frac{|S|}{|\{s_i : uw_j \in s_i\}|} \quad (3)$$

where $L_{i,j}$ is the character count in the uw_j in the sentence s_i normalized by length of all the unigram words in sentence s_i. $|S|$ is total number of sentences in the training set and $|\{s_i : uw_j \varepsilon s_i\}|$ is number of sentences in which the unigram uw_j occurs atleast once.

3.2 Feature Selection

Feature selection methods help in removing the features which may not be useful for categorization. To achieve it, feature selection techniques select subset of total features. But, it is important to reduce features without compromising on the accuracy of a classifier. Most methods like Information Gain(IG) (C. Lee.et.al, 2006), Correlation Feature Selection(CFS) (M.A. Hall, 1999), Chi-Squared (χ^2) (J. Bakus.et.al, 2006), Odds ratio (OR) (G. Forman, 2003) does not consider the frequency of the text or term between the categories which leads in reduction of accuracy of a classifier.

In-order to overcome this problem, we used Entropy based category coverage difference (ECCD) (C. Largeron.et.al, 2011) feature selection method which uses the entropy of the text or term. f_j is used to represent the text feature extracted (unigram or bigram), c_k for category of the class and c_k^- for the complement of the class. Where j represent number of features and k represents two classes either subjective or objective.

Entropy based category coverage Difference(ECCD)
This feature selection method (C. Largeron.et.al, 2011) was proposed to mine INEX XML documents. We use this approach for improving the subjective and objective sentence classification. Let T_j^k be number of occurrences of text feature f_j in the category c_k sentence and, fq_j^k is the frequency of f_j in that category c_k given by Equation 4.

$$fq_j^k = \frac{T_j^k}{\sum_k T_j^k} \tag{4}$$

So Entropy $Ent(f_j)$ of text feature f_j is given by Equation 5

$$Ent(f_j) = \sum_{k=1}^{r} (fq_j^k) * log_2(fq_j^k) \tag{5}$$

Entropy equals 0, if the text feature f_j appears only in one category. It means that feature has good discrimination ability to classify the sentences. Similarly, entropy of the text feature will be high if the feature is represented in two classes. If Ent_m represent the maximum entropy of the feature f_j, ECCD(f_j, c_k) is given by following equation 6.

$$ECCD(f_j, c_k) = P(f_j|c_k) - P(f_j|c_k^-) * \frac{Ent_m - Ent(f_j)}{Ent_m} \tag{6}$$

Where $P(f_j|c_k)$ and $P(f_j|c_k^-)$ are probability of observing the text feature f_j in a sentence belonging to category c_k and c_k^- respectively.

The advantage of ECCD is that higher the number of sentences of category c_k containing feature f_j and lower the number of sentences in other category containing f_j, we get higher value for equation 6. It means f_j becomes the characteristic feature of that category c_k which helps in better feature selection. Feature selection method which is similar to ECCD is mentioned below.

Categorical Proportional Difference (CPD)
CPD (M. Simeon.et.al, 2008) is a measure of the degree to which a text feature contributes to differentiating a particular category from other categories in a text corpus. We calculate the CPD for a text feature f_j by taking a ratio that considers the number of sentences in subjective category c_k in which the text feature occurs and the number of sentences in objective category c_k^- in which the text f_j also occurs. Equation 7 shows the details. Certain threshold of CPD score is kept to reduce the number of features.

$$CPD(f_j, c_k) = \frac{P(f_j, c_k) - P(f_j, c_k^-)}{P(f_j, c_k) + P(f_j, c_k^-)} \tag{7}$$

3.3 Contingency Table Representation of features

Feature selection methods mentioned in earlier section is estimated using a contingency table. Let A, be the number of sentences in the subjective category containing feature f_j. B, be the number of sentences in the objective category containing f_j. C, be the number of sentences of subjective category which do not contain feature f_j given by f_j^- and D, be the number of sentences in objective category c_k^- which do not contain f_j. Let $(M = A+B+C+D)$ be the total possibilities. Table 1 represents the above mentioned details. Using the Table 1 each of the feature selection methods can be estimated. Table 2 show the details.

	Subjective	Objective
f_j	A	B
f_j^-	C	D

Table 1: Contingency Table

FS	Representation
$IG(f_j, c_k)$	$-\frac{A+C}{M}log(\frac{A+C}{M})+$ $\frac{A}{M}log(\frac{A}{A+B})+\frac{C}{M}log(\frac{C}{C+D})$
$\chi^2(f_j, c_k)$	$\frac{M(A*D-B*C)^2}{(A+B)*(A+C)*(B+D)*(C+D)}$
$OR(f_j, c_k)$	$\frac{D*A}{C*B}$
$CPD(f_j, c_k)$	$\frac{A-B}{A+B}$
$ECCD(f_j, c_k)$	$\frac{(A*D-B*C)*Ent_m-Ent(f_j)}{(A+C)*(B+D)*Ent_m}$

Table 2: Estimation Table

4 Experimental Setup

In-order to achieve subjective and objective classification at sentence level for different languages. We performed our experiments using different datasets.

4.1 Datasets

Translated MPQA corpus provided in (C. Banea.et.al, 2010) containing subjective and objective sentences[1] of French, Arabic, and Romanian languages were used for experiments. For English, MPQA corpus[2] containing subjective and objective sentences used for translation of above mentioned corpus is used. Hindi experiments were performed using sentences from the news corpus (A. Mogadala.et.al, 2012) tagged with positive, negative and objective sentences. Positive and negative sentences are further clubbed into subjective sentences to do subjectivity analysis.

4.2 Evaluation

To evaluate various feature selection methods, we use F-measure scores which combines precision and recall. Precision (P_s) measures the percentage of sentences correctly assigned to subjective category, and recall (R_s) measures the percentage of sentences that should have been assigned to subjective category but actually assigned to subjective category. Using P_s and R_s subjective F-measure F_s is calculated. Similarly, Objective F-measure F_o is calculated using P_o and R_o. After F-measure is determined for both subjective and objective class, the macro-average F-measure $F_{macro-avg}$ is determined by the following Equation 8.

$$F_{macro-avg} = \frac{\Sigma_{i=o,s}F_i}{2} \quad (8)$$

5 Experiments

Initially, 1500 subjective and 1500 objective sentences of English, Romanian, French and Arabic languages are used to perform the experiments. While for Hindi, entire corpus constituting 786 subjective and 519 objective sentences was used. Different feature weighing and selection methods are evaluated with 2 different classifiers to obtain best combination for each language. Table 8 to Table 12 show the Macro-Average ($F_{macro-avg}$) scores obtained after 10 cross-validation using sequential minimal optimization algorithm (J. Platt, 1998) for training a support vector machine(SVM) using polynomial kernel and Naive Bayes Multinomial(NBM) classifiers. Feature space obtained after application of feature selection methods for each language are mentioned in Tables 3, 4, 5, 6, 7.

Once the best combination is obtained for each language. It is compared with multilingual space classifier proposed in (C. Banea.et.al, 2010)[3] along with the baseline constituting simple Naive Bayes classifier with unigram features. Multilingual space constitutes words as features from all languages used for experiments except Hindi.

Scalability of feature selection methods is an issue. In-order to understand the performance of ECCD feature selection method with classifiers. In every iteration 500 sentences are added to each class of initial 1500 subjective and objective sentences limiting to maximum of 3500 to get average scores. Table 13 show the comparison of average scores obtained for each language. Figures 1, 2, 3, 4 show precision and recall for subjective sentences obtained using different methods for English, Romanian, French and Arabic respectively using different number of sentences.

[1] http://lit.csci.unt.edu/index.php/Downloads

[2] http://www.cs.pitt.edu/mpqa

[3] Note that paper used the entire dataset which had unequal subjective and objective sentences. We used equal number of subjective and objective sentences taken each time from dataset. So, experiments using this method are again performed on our dataset.

[SVM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.705	0.660	0.705
CFS	0.710	0.670	0.705
IG,OR,χ^2	0.680	0.665	0.685
CPD	**0.840**	0.805	0.835
ECCD	0.830	0.805	0.830

[NBM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.745	0.690	0.740
CFS	0.730	0.685	0.725
IG,OR,χ^2	0.735	0.690	0.730
CPD	0.855	**0.925**	0.890
ECCD	0.850	**0.925**	0.875

Table 8: $F_{macro-avg}$ - English

[SVM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.685	0.665	0.680
CFS	0.715	0.675	0.695
IG,OR,χ^2	0.695	0.635	0.690
CPD	0.845	0.815	**0.850**
ECCD	0.845	0.815	0.845

[NBM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.740	0.685	0.730
CFS	0.735	0.690	0.740
IG,OR,χ^2	0.745	0.685	0.725
CPD	0.865	0.935	0.890
ECCD	0.865	**0.940**	0.885

Table 9: $F_{macro-avg}$ - Romanian

[SVM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.695	0.685	0.685
CFS	0.710	0.695	0.685
IG,OR,χ^2	0.690	0.685	0.685
CPD	**0.855**	0.825	0.850
ECCD	0.845	0.820	0.835

[NBM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.730	0.705	0.725
CFS	0.730	0.710	0.725
IG,OR,χ^2	0.725	0.690	0.710
CPD	0.860	0.940	0.900
ECCD	0.845	**0.950**	0.885

Table 10: $F_{macro-avg}$ - French

[SVM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.670	0.660	0.675
CFS	0.710	0.665	0.680
IG,OR,χ^2	0.665	0.645	0.660
CPD	0.855	0.825	0.850
ECCD	0.850	0.830	**0.860**

[NBM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.720	0.690	0.710
CFS	0.730	0.695	0.725
IG,OR,χ^2	0.720	0.690	0.710
CPD	0.910	**0.915**	0.910
ECCD	**0.915**	**0.915**	**0.915**

Table 11: $F_{macro-avg}$ - Arabic

[SVM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.665	0.655	0.650
CFS	0.635	0.655	0.600
IG,OR,χ^2	0.665	0.660	0.650
CPD	0.760	**0.845**	0.755
ECCD	0.735	**0.845**	0.735

[NBM]

Feature Selection	UF.ISF	BF.ISF	NUWL
None	0.615	0.655	0.635
CFS	0.460	0.440	0.460
IG,OR,χ^2	0.580	0.655	0.605
CPD	0.590	0.845	0.660
ECCD	0.555	**0.850**	0.655

Table 12: $F_{macro-avg}$ - Hindi

6 Result Analysis

It is observed from the Table 8 to Table 12 that ECCD feature selection and BF.ISF feature weighing method with NBM classifier performs consistently across languages. This behavior is observed due to capability of ECCD in efficiently discrimi-

Language (Method)		P_s	R_s	F_s	P_o	R_o	F_o	$P_{macro-avg}$	$R_{macro-avg}$	$F_{macro-avg}$
English	Baseline	0.720	0.830	0.770	0.800	0.676	0.733	0.760	0.753	0.751
	NBM + BF.ISF + ECCD	**1.000**	0.865	**0.925**	**0.875**	**1.000**	**0.935**	**0.937**	**0.932**	**0.930**
	NB + MultiLingual Space (Banea,2010)	0.497	**0.927**	0.644	0.350	0.057	0.087	0.423	0.491	0.365
	Wiebe & Riloff (Wiebe and Riloff, 2005)	0.904	0.342	0.466	0.824	0.307	0.447	0.867	0.326	0.474
	Chenghua Lin (C. Lin.et.al, 2011)	0.710	0.809	0.756	0.716	0.597	0.651	0.713	0.703	0.703
Romanian	Baseline	0.713	0.830	0.766	0.796	0.663	0.723	0.755	0.746	0.745
	NBM + BF.ISF + ECCD	**1.000**	0.880	**0.940**	**0.890**	**1.000**	**0.940**	**0.945**	**0.940**	**0.940**
	NB + MultiLingual Space (Banea, 2010)	0.497	**0.913**	0.640	0.383	0.063	0.096	0.440	0.488	0.368
French	Baseline	0.703	0.826	0.760	0.790	0.643	0.713	0.746	0.736	0.736
	NBM + BF.ISF + ECCD	**1.000**	0.905	**0.950**	**0.915**	**1.000**	**0.955**	**0.957**	**0.952**	**0.952**
	NB + MultiLingual Space (Banea,2010)	0.490	**0.913**	0.636	0.370	0.056	0.096	0.430	0.485	0.366
Arabic	Baseline	0.703	0.800	0.750	0.770	0.666	0.713	0.736	0.733	0.731
	NBM + BF.ISF + ECCD	**1.000**	0.845	**0.915**	**0.865**	**1.000**	**0.925**	**0.932**	**0.922**	**0.920**
	NB + MultiLingual Space (Banea,2010)	0.497	**0.983**	0.656	0.293	0.006	0.016	0.353	0.495	0.336
Hindi	Baseline	0.680	0.900	0.770	0.690	0.350	0.460	0.685	0.625	0.615
	NBM + BF.ISF + ECCD	**0.810**	**1.000**	**0.900**	**1.000**	**0.650**	**0.790**	**0.905**	**0.825**	**0.850**

Table 13: Comparison of Average scores between proposed and other approaches

Feature Selection	Unigrams (UF.ISF,NUWL)	Bigrams (BF.ISF)
None	100.0	100.0
CFS	1.8	8.4
IG,OR,χ^2	60.0	60.0
CPD	66.2	90.0
ECCD	65.7	90.0

Table 3: Feature Space Used(%) - English

Feature Selection	Unigrams (UF.ISF,NUWL)	Bigrams (BF.ISF)
None	100.0	100.0
CFS	1.7	7.2
IG,OR,χ^2	60.0	60.0
CPD	65.5	90.1
ECCD	65.0	90.0

Table 4: Feature Space Used(%) - Romanian

Feature Selection	Unigrams (UF.ISF,NUWL)	Bigrams (BF.ISF)
None	100.0	100.0
CFS	1.4	5.7
IG,OR,χ^2	60.0	60.0
CPD	68.5	88.6
ECCD	68.0	88.5

Table 5: Feature Space Used(%) - French

Feature Selection	Unigrams (UF.ISF,NUWL)	Bigrams (BF.ISF)
None	100.0	100.0
CFS	1.2	3.9
IG,OR,χ^2	60.0	60.0
CPD	71.8	92.3
ECCD	71.4	92.2

Table 6: Feature Space Used(%) - Arabic

Feature Selection	Unigrams (UF.ISF,NUWL)	Bigrams (BF.ISF)
None	100.0	100.0
CFS	2.3	0.7
IG,OR,χ^2	60.0	60.0
CPD	58.1	81.1
ECCD	57.4	80.9

Table 7: Feature Space Used(%) - Hindi

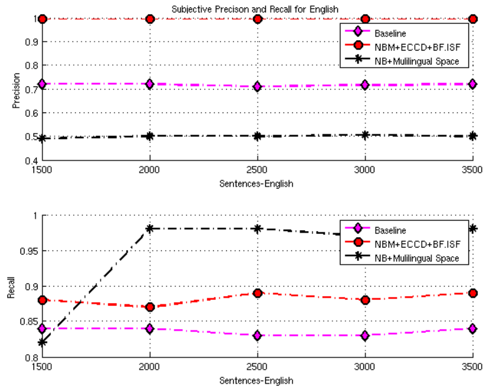

Figure 1: Subjective Precision and Recall (English)

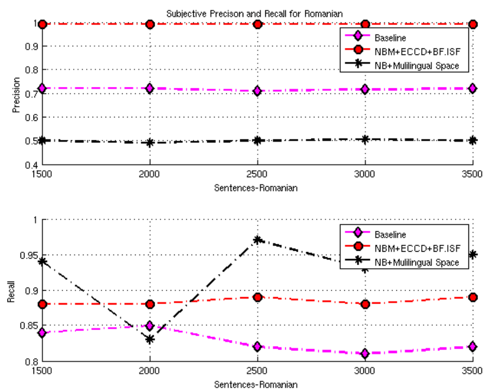

Figure 2: Subjective Precision and Recall (Romanian)

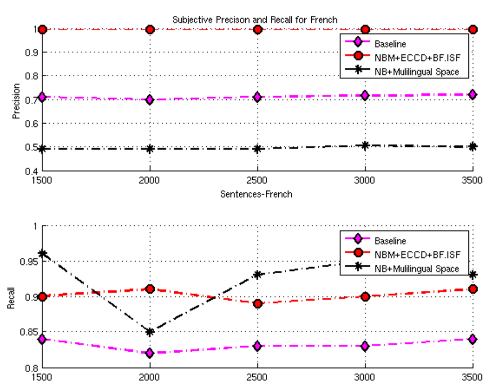

Figure 3: Subjective Precision and Recall (French)

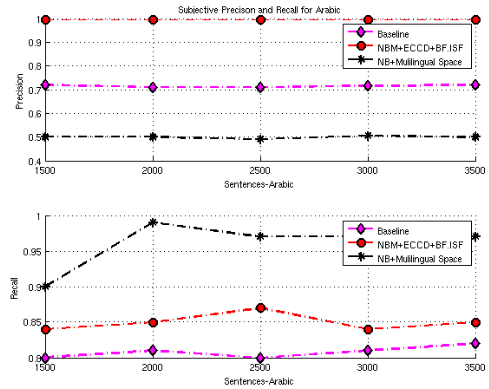

Figure 4: Subjective Precision and Recall (Arabic)

nating the features belonging to a particular class. Although, UF.ISF and NUWL with ECCD and CPD using SVM classifier has more scores than BF.ISF, it is not significantly contrasting with the results of NBM classifier. So, NBM classifier with ECCD feature selection and BF.ISF feature weighing method which obtained high $F_{macro-avg}$ scores is selected for comparison with other approaches in Table 13. The proposed method not only outperforms on $F_{macro-avg}$ compared to other approaches but also on $P_{macro-avg}$ and $R_{macro-avg}$ in all languages.

For English, proposed methods gains 23.8% over baseline in $F_{macro-avg}$ and 8.0% on $P_{macro-avg}$ on (Wiebe and Riloff, 2005). Similarly, from Table 13 it can be deducted that proposed method for Romanian attains 26.1% more $F_{macro-avg}$ than baseline and 155.4% more compared to Multilingual space classifier (C. Banea.et.al, 2010). Similar observations can be made for other languages. Even though (C. Banea.et.al, 2010) attains high recall for every language. It fails to attain high precision due to presence of large number of frivolous word features which are common for both classes. This being major drawback, ECCD feature selection method eliminates features which attains zero entropy. This reduces the randomness of features and leave only those features which are more eligible for discriminating the classes. Combining ECCD with BF.ISF, a language independent weighing method for Bi-gram features extracted from the sentences. We are able to attain a best classification accuracies which are consistent across languages.

Figures 1, 2, 3, 4 also show that increase in number of sentences does not effect the precision of the proposed method, as it still outperforms other methods. But, scalability problem persists for ECCD with BF.ISF for larger datasets, as it may not eliminate less random features due to noise and other constraints. Also, feature selection methods ensure the performance of classifiers is maintained by reducing number of features. But, it does not ensure reduction in fixed percentage of features. As observed our best performing feature selection method ECCD reduces feature size by 10% only.

7 Performance Analysis between SVM and NBM

From the Table 8 to 12 it is observed that SVM with some feature selection and weighing methods performs equivalent to the NBM. However, as the number of documents increases the performance of SVM may degrade. It can be derived that, as the training data size increases, it is rare to see SVM performing better than NBM.

7.1 Training Time behavior

SVM is in a clear disadvantage compared to NBM when processing time is considered. The training time of the SVM is particularly high, especially for larger feature spaces. It is probably attributed to the time taken in finding the proper separating hyperplane.

7.2 Features behavior

Large feature spaces do not necessarily lead to best performance. So feature selection methods are used to create small feature spaces to build SVM and NBM classifiers. Sometimes, small feature space sizes make SVM perform equivalent to NBM as observed in Table 12. Thus, this would explain why SVM is outperformed for small training set sizes and for small feature spaces with large training sets.

8 Conclusion and Future Work

In this paper, subjective classification is achieved using combination of feature selection and weighing methods which are consistent across languages. We found that our proposed method which combines ECCD feature selection and BF.ISF feature weighing method used along with NBM classifier perform across languages. It not only outperforms other feature selection methods but also achieve better scores compared to other approaches. In future, we want to apply this approach on bigger datasets and also extend it to multiple class problems.

References

M. Hu and B. Liu. 2004. *Mining opinion features in customer reviews.* Proceedings of the National Conference on Artificial Intelligence, 755–760.

K. Ganesan, C.X. Zhai and J. Han. 2010. *Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions.* Proceedings of the 23rd International Conference on Computational Linguistics, 340–348.

A. Balahur and E. Boldrini and A. Montoyo and P. Martínez-Barco. 2009. *Opinion and Generic Question Answering systems: a performance analysis.* Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, 157–160.

J. Wiebe and E. Riloff. 2005. *Creating subjective and objective sentence classifiers from unannotated texts.* Computational Linguistics and Intelligent Text Processing, 486–497, Springer.

R. Mihalcea, C. Banea and J. Wiebe. 2007. *Learning multilingual subjective language via cross-lingual projections.* ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS , Volume 45, 976.

C. Banea, R. Mihalcea, J. Wiebe and S. Hassan. 2008. *Multilingual subjectivity analysis using machine translation.* Proceedings of the Conference on Empirical Methods in Natural Language Processing, 127–135, ACL.

C. Banea, R. Mihalcea and J. Wiebe. 2010. *Multilingual subjectivity: are more languages better.* Proceedings of the 23rd International Conference on Computational Linguistics, 28–36, ACL.

G. Murray and G. Carenini. 2009. *Predicting subjectivity in multimodal conversations.* Proceedings of Empirical Methods in Natural Language Processing: Volume 3, 1348–1357, ACL.

T. Wilson and S. Raaijmakers. 2008. *Comparing word, character, and phoneme n-grams for subjective utterance recognition.* Ninth Annual Conference of the International Speech Communication Association.

C. Lin, Y. He and R. Everson. 2011. *Sentence subjectivity detection with weakly-supervised learning.* Proceedings of International Joint Conference on Natural Language Processing (IJCNLP), Volume 2, 2.

H. Kanayama and T. Nasukawa. 2006. *Fully automatic lexicon expansion for domain-oriented sentiment analysis.* Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 355–363, ACL.

T. Zagibalov and J. Carroll. 2008. *Unsupervised classification of sentiment and objectivity in Chinese text.* IJCNLP, 304–311.

Y. Wu and D.W. Oard. 2007. *NTCIR-6 at Maryland: Chinese opinion analysis pilot task.* Proceedings of the 6th NTCIR Workshop on Evaluation of Information Access Technologies.

M. Bautin, L. Vijayarenu and S. Skiena. 2008. *International sentiment analysis for news and blogs.* Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

X. Wan. 2009. *Co-training for cross-lingual sentiment classification.* Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1, 235–243, ACL.

G. Forman. 2003. *An extensive empirical study of feature selection metrics for text classification.* The Journal of Machine Learning Research, Volume 3, 1289–1305.

M. Gamon. 2004. *Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis.* Proceedings of the 20th international conference on Computational Linguistics, 841, ACL.

G. Salton, A. Wong and C.S. Yang. 1975. *A vector space model for automatic indexing.* Communications of the ACM, Volume 18, 613–620.

A. Abbasi, H. Chen and A. Salem. 2008. *Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums.* ACM Transactions on Information Systems (TOIS), Volume 26, 12, ACM.

C. Largeron, C. Moulin and M. Géry. 2011. *Entropy based feature selection for text categorization.* Proceedings of ACM Symposium on Applied Computing, 924–928, ACM.

M.A. Hall. 1999. *Correlation-based feature selection for machine learning.* Phd Thesis, The University of Waikato.

C. Lee and G.G. Lee. 2006. *Information gain and divergence-based feature selection for machine learning-based text categorization.* Information processing & management, Volume 42, 155–165, Elsevier.

J. Bakus and M.S. Kamel. 2006. *Higher order feature selection for text classification.* Knowledge and Information Systems, Volume 9, 468–491, Springer.

M. Simeon and R. Hilderman. 2008. *Categorical proportional difference: A feature selection method for text categorization.* Proceedings of the Seventh Australasian Data Mining Conference (AusDM), Volume 87, 201–208.

A. Mogadala and V. Varma. 2012. *Retrieval approach to extract opinions about people from resource scarce language News articles.* Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), ACM KDD.

J. Platt. 1998. *Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods - Support Vector Learning.* B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press.