Multi-View AdaBoost for Multilingual Subjectivity Analysis

Min Xiao   Yuhong Guo
Department of Computer and Information Sciences
Temple University, Philadelphia, PA, USA
minxiao@temple.edu, yuhong@temple.edu

ABSTRACT
Subjectivity analysis has received increasing attention in natural language processing field. Most of the subjectivity analysis works however are conducted on single languages. In this paper, we propose to perform multilingual subjectivity analysis by combining multi-view learning and AdaBoost techniques. We aim to show that by boosting multi-view classifiers we can develop more effective multilingual subjectivity analysis tools for new languages as well as increase the classification performance for English data. We empirically evaluate our two multi-view AdaBoost approaches on the multilingual MPQA dataset. The experimental results show the multi-view AdaBoost approaches significantly outperform existing monolingual and multilingual methods.

1 Introduction

Subjectivity analysis has received increasing interest in natural language processing (NLP) area (Banea et al., 2010; Alm, 2011; Abdul-Mageed and Diab, 2011; Abdul-Mageed et al., 2011). Subjectivity refers to the expression of emotions, sentiments, opinions, beliefs, speculations, evaluations, as well as other private states (Banfield, 1982; Wiebe, 1994). Subjectivity classification aims to distinguish whether a given text expresses subjective or objective meaning (Wiebe and Cardie, 2005; Abdul-Mageed et al., 2011; Banea et al., 2008, 2010). Subjectivity analysis has been intensively studied, particularly motivated by the prevalent need for opinion-related applications, including mining opinions from product reviews (Pang et al., 2002; Hu and Liu, 2004) or political news (Abbott et al., 2011), and recognizing stances in online debates (Somasundaran and Wiebe, 2009, 2010). Moreover, many NLP tasks employ subjectivity analysis as an additional layering to filter data. Research that benefited from this phase ranges from conversation summarization (Seki et al., 2005; Carenini et al., 2008) and information extraction (Riloff et al., 2005) to text semantic analysis (Wiebe and Mihalcea, 2006) and question answering (Li et al., 2008; Yu and Hatzivassiloglou, 2003).

Although subjectivity analysis has been widely studied in NLP area, much work has only focused on English data. Recently, some researchers propose to carry out subjectivity analysis in a multilingual framework based on machine translation, where resources or tools of subjectivity analysis developed in one language are used to support developing resources or tools in another language (Mihalcea and Banea, 2007; Banea et al., 2008, 2010). The approaches in (Mihalcea and Banea, 2007; Banea et al., 2008) however only exploit the translated target-language-view of the data to develop a subjectivity analysis tool, which is a waste of resources in a multilingual setting since possible parallel views of the data are ignored. Banea et al. (2010) propose to overcome this shortcoming by conducting subjectivity analysis based on concatenated multilingual input feature vectors. This simple feature combination method nevertheless is still very preliminary in exploring the capacity of multi-view learning for subjectivity analysis on multilingual data.

In this paper, we propose to use multi-view AdaBoost approaches for multilingual subjectivity analysis, which combine the advantages of both multi-view learning and AdaBoost learning in one integrated framework. By exploring multi-view learning, we expect to exploit the complementary discriminative information in different language views. By incorporating the multi-view learning into an AdaBoost framework, we expect to further boost the classification accuracy of the integrated models. Based on different strategies of exploring multilingual information, we develop two approaches in this paper: Multi-View Majority Voting AdaBoost (*MVAB1*) and Multi-View Weighted Voting AdaBoost (*MVAB2*).

To demonstrate the effectiveness of the proposed approaches, we empirically evaluate them on a multilingual subjectivity analysis dataset, the Multilingual Multi-Perspective Question Answering (MPQA) corpus. To justify the robustness of our boosting framework, we conduct experiments using two types of base classifiers, Support Vector Machines (SVM) and Naïve Bayes (NB). The experimental results show that the proposed approaches can significantly outperform other comparison methods for multilingual subjectivity analysis. Overall, the contributions of this paper can be summarized as below:

- We propose two multi-view AdaBoost algorithms, Multi-View Majority Voting AdaBoost and Multi-View Weighted Voting AdaBoost, which can be widely used for multilingual classification tasks when parallel corpora or machine translation is available.

- Experimentally, we evaluate our approaches on Multilingual MPQA corpus and obtain a subjectivity classifier with *accuracy* as high as 78.19% and *macro F1* as high as 77.44% over all six languages.

The remainder of the paper is organized as follows. Related work is presented in Section 2. In Section 3, we present the multilingual subjectivity analysis problem and two proposed multi-view AdaBoost approaches. In Section 4, we present the experimental results and discussions. We then conclude the paper.

2 Related Work

The importance of subjectivity analysis has been widely acknowledged by language analysts, including computational linguists. Due to the availability of data resources, much work on subjectivity analysis has focused on English data alone. However, recently, some work tries to generate resources and develop tools for other languages by transferring labeled English subjectivity resources and corresponding analysis tools.

Mihalcea and Banea (2007) proposed to build subjectivity classifiers for Romanian data by leveraging the resources and tools available in English. They developed a lexicon-based approach and a corpus-based approach. For the lexicon-based approach, they first created a target-language subjectivity lexicon by translating the existing annotated English subjectivity lexicon via bilingual dictionaries and then trained a rule-based classifier relying on the translated lexicon. For the corpus-based approach, they first manually translated an automatically annotated English corpus into Romanian language and projected the subjectivity annotations correspondingly, and then trained a statistical classifier on the resulting corpus. They empirically evaluated their approaches on MPQA corpus and SemCor corpus (Miller et al., 1993), showing that the corpus translations preserve subjectivity more reliably than the lexicon translations. Nevertheless, the requirement for manual translation is a big restriction for potential usage of the proposed approaches.

Banea et al. (2008) then proposed to generate resources and tools for new languages (Spanish and Romanian) using machine translation and cross-lingual annotation projections. Specifically, they used machine translation to transfer the manually or automatically annotated training data from the source language (English) into the target languages, and projected the subjectivity annotations of the transferred data across language correspondingly. Then they employed statistical machine learning techniques such as Support Vector Machines and Naïve Bayes to produce a subjectivity classifier on the translated corpus in the target language. The advantage of this approach is that it does not need any original target language data for training. Thus it can be widely used for any new target language as long as a source-target-language machine translation engine is available. Nonetheless, the subjectivity analysis tool developed by this approach is dependent on the quality of machine translation since only translated data is used in training.

Banea et al. (2010) proposed to combine multiple language spaces altogether in an expanded feature space. Specifically they combined the original English feature vector of an instance and its translated feature vectors in different target languages together into one feature vector, and then used the training instances expressed in this expanded feature space to train multilingual subjectivity classifiers. They empirically evaluated their approach on MPQA corpus, by translating English sentences into five other languages. Their empirical results showed that multiple languages can complement each other to greatly increase subjectivity classification

performance for target languages as well as for English source data, comparing to training subjectivity classifiers on the target language alone. Nevertheless, the parallel texts in multiple languages can be approximately taken as label-conditionally independent multiple views of the same set of data objects, and their simple feature space expanding method is still far from fully exploiting this multilingual information. Thus, in this work we investigate new multi-view AdaBoost approaches to improve the performance of multilingual subjectivity analysis.

Multilingual views have also been exploited in sentiment analysis (Wan, 2009; Lu et al., 2011). Wan (2009) used co-training on bilingual views (Chinese and English) generated from machine translation to perform sentiment analysis on Amazon product reviews. Their approach however requires in-domain data from the target language for training. Lu et al. (2011) developed a maximum entropy based statistical model to jointly train two monolingual sentiment classifiers using an EM-algorithm. They also only studied the bilingual situation with experiments on English and Chinese. Combining multi-view learning and boosting has been studied in a few different ways on application problems outside of NLP field, including a semi-supervised boosting method for object category recognition and visual object tracking (Saffari et al., 2010), and an embedded two-view AdaBoost method for UCI data (Xu and Sun, 2010). But our work is the first one that combines multi-view learning and AdaBoost learning to address supervised subjectivity analysis with multiple languages.

3   Multi-View AdaBoost for Multilingual Subjectivity Analysis

In this section, we introduce two multi-view AdaBoost approaches, which combines the advantages of both multi-view learning and boosting learning to achieve better multilingual subjectivity classifiers. Below, we will first describe the general framework of multilingual subjectivity analysis and briefly introduce the AdaBoost algorithm, and then present the two proposed multi-view AdaBoost approaches.

3.1   Multilingual Subjectivity Analysis

Banea et al. (2010) pointed out that training a subjectivity classifier on the resulting target monolingual corpus alone, though works, is not good enough. There are two main reasons. First, in order to correctly predict labels based on statistical information, a sufficient amount of training data is needed, which may not be available in the monolingual corpus. Second, in the monolingual corpus alone, some discriminative features present in the test data may not appear in the training data and therefore their information cannot be used to generate an effective classifier. In both cases, multilingual subjectivity analysis can have advantages.

Below we demonstrate the problems of learning with monolingual corpus using examples from the multilingual MPQA dataset. Following (Banea et al., 2010)'s suggestions, we assume that only the words in italics carry potential subjective meaning and their surrounding contexts would be objective if without them. Therefore, their association with an either subjective or objective sense imparts the same label to the whole segment.

We explore the first data sparseness problem through the following two examples (En 1 and En 2) from the English version of the MPQA dataset as well as their respective translations in Spanish (Es 1 and Es 2):

> "En 1: The source said that the ministry would soon deliver copies of the report to the various ministries *concerned*, especially the Interior and Municipalities Ministry,

prior to relaying its observations to the State Department in Washington."

"Es 1: La fuente dijo que el ministerio pronto entregar copias del informe a los distintos ministerios *interesados*, en particular el Ministerio del Interior y de los Municipios, antes de transmitir sus observaciones al Departamento de Estado en Washington."

"En 2: Still, the overseers of the prison are *concerned* that detainees aren't getting enough pita bread with their meals, and they're planning to make the food spicier, just the way the prisoners like it back home. "

"Es 2: Sin embargo, los supervisores de la prisión tienen la *preocupación* de que los detenidos no están recibiendo suficiente pan pita con sus comidas, y que está pensando en hacer la comida spicier, sólo la forma en la que los reclusos como para volver a casa."

We focus on the word *concerned*. In the first example (En 1), it is used with an objective sense, which means a group of ministries defined earlier in the context. While in the second example (En 2), *concerned* serves as a subjective carrier. If we train a monolingual classifier on the English data alone, due to the data sparseness paradigm, our machine learning model may not distinguish between the word's subjective and objective senses when inferring a label for the whole sentence. However, the corresponding translations of *concerned* in Spanish are functionally different because of the surrounding context. We denote the respective translations in the Spanish context (Es 1 and Es 2) using the italic form. If we take the Spanish translation into consideration during training, we may obtain a classifier, which can potentially differentiate between the senses and predict the correct sentence label.

Next, we explore the second problem with two other examples below (En 3 and En 4) extracted from the MPQA dataset and their corresponding machine translations in Romanian:

"En 3: What is the point of engaging in dialogue with a government that is only interested in buying time while it fervently escalates a campaign of bludgeoning its citizens in the hope of frightening voters into *supporting* Mugabe? "

"Ro 3: Ce este pe punctul de angajarea in dialog cu un guvern doar ca este interesat de cumpararea timp in timp ce aceasta inflacarare e o campanie de forta cetatenilor sai in speranta de infricosator electoratul in *sprijinirea* lideri? "

"En 4: According to the sources, the EU which was barred by the Government from observing the election because some of its members were openly *supporting* the MDC, sent a Ms Maria Macchiaverna to "support the financial management of our assistance" to the Sadc Parliamentary Forum and ZESN."

"Ro 4: Potrivit surselor, UE care a fost oprita de guvernul de la observarea alegerilor pentru ca unii dintre membrii sai s-au deschis *sustinerea* mdc, a trimis un MS Maria macchiaverna sa "sprijin financiar de gestiune noastra de asistenta" la parlamentare sadc Forumului si zesn. "

In both of the two examples (En 3 and En 4), *supporting* carries subjective senses. However, the corresponding translations of *supporting* in Romanian for En 3 and En 4 are different: *sprijinirea* (in Ro 3) and *sustinerea* (in Ro 4). If we train a monolingual classifier on Romanian corpus alone, and the training set contains *sprijinirea* but not *sustinerea*, it is hard to infer the correct label for a context containing *sustinerea* by leveraging information from *sprijinirea*. However, if we adopt a multilingual classifier, we may be able to predict a correct label for the

context containing *sustinerea* by using the English information, or the associations between *supporting* and *sprijinirea*, as well as *supporting* and *sustinerea*.

As suggested in these examples above, exploiting multilingual information can compensate the shortcomings of learning from monolingual data. Multilingual subjectivity analysis has been previously studied in (Banea et al., 2008, 2010). We propose to conduct multilingual subjectivity analysis following the general framework suggested in these works. Assume we have manually annotated subjectivity corpus in English, and aim to develop subjectivity resources and tools for other languages, such as Arabic, French, German, Romanian, and Spanish, for which there are few text processing resources and tools to date. The multilingual subjectivity analysis framework contains three steps:

- Translating English sentences into target languages by using machine translation.

- Projecting subjectivity annotations from English to translated target languages.

- Producing subjectivity analysis tools on the resulting labeled corpora by using statistical machine learning techniques.

Banea et al. (2010) empirically studied the multilingual subjectivity analysis problem and provided a simple solution by expanding the feature space with multiple languages. In this work, we propose to further improve multilingual subjectivity analysis by exploiting multi-view learning in the framework of AdaBoost. In addition to the multilingual analysis problem discussed above, our approaches have the following two standpoints. First, Amini et al. (2009) provided theoretical bounds for multi-view classifiers and showed that additional views generated by machine translation may significantly improve classification performance. Second, within an AdaBoost framework, the algorithms can deal with hard examples which are difficult to be recognized using base multi-view learners. Moreover, an exponential loss function is guaranteed to be minimized over the multilingual data.

3.2 AdaBoost

Boosting is a general method for improving accuracy of any given learning algorithm by combining many weak or base learners. A well-known boosting algorithm is AdaBoost, which was introduced by (Freund and Schapire, 1997). AdaBoost takes a set of training instances $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ as input, where each x_i belongs to some instance space $\mathcal{X}$, and each label y_i is in some label space $\mathcal{Y}$. For a binary classification problem, we assume $\mathcal{Y} = \{+1, -1\}$. AdaBoost calls a given base learning algorithm repeatedly in a series of rounds $t = 1 \cdots T$ and maintains a set of weights or a distribution over the training instances during the rounds. All weights are initialized equally; but in each round, the weights of misclassified instances are increased so that the weak learner will be forced to focus on the hard instances in the training set in the next round.

In each round, the weak learner trains a base classifier $h_t : \mathcal{X} \to \mathcal{Y}$ over the training instances with the weighted distribution D_t. For the binary classification problem, the base learner's job is to minimize the error

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] \tag{1}$$

Once the base classifier h_t has been found, AdaBoost chooses a parameter α_t that intuitively

---

Algorithm 1 Multi-view Majority Voting AdaBoost

Input: A multi-view binary training set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, with $y_i \in \{+1, -1\}$.
Output: The final classifier H.
Initialize: $D_1(i) = \frac{1}{n}$.
for $t = 1$ **to** T **do**
  • Separately train a set of single view classifiers $\{h_{tv}\}$ using distribution D_t.
  • Compute the base classifier h_t using the majority voting scheme in Eq. (5).
  • Set $\epsilon_t = Pr_{i \sim D_t}[h_t(\mathbf{x}) \neq y_i]$
  • Set $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$.
  • Update $D_{t+1}(i) = D_t(i) \frac{e^{-\alpha_t y_i h_t(\mathbf{x}_i)}}{Z_t}$, where Z_t is a normalization factor.
end for
$H(\mathbf{x}) = sign(\sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}))$

---

measures the confidence or importance that it assigns to h_t. For binary h_t, α_t is set as

$$\alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t}) \tag{2}$$

in (Freund and Schapire, 1997). The distribution D_t is then updated as

$$D_{t+1}(i) = \frac{D_t(i)e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \tag{3}$$

where Z_t is a normalization factor.

The final classifier H produced by AdaBoost is a weighted majority vote of the T base classifiers

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \tag{4}$$

where α_t is the weight assigned to h_t.

3.3 Multi-View AdaBoost for Multilingual Subjectivity Analysis

In the multilingual setting, each instance (sentence) is described using feature sets from multiple languages, where each feature set from one language can be treated as one view of the instance. To address multilingual subjectivity analysis in an AdaBoost framework, we propose to integrate multi-view learning into the AdaBoost. In particular, we develop two approaches using different multi-view learning strategies: a multi-view majority voting AdaBoost (MVAB1) and a multi-view weighted voting AdaBoost (MVAB2).

3.3.1 Multi-View Majority Voting AdaBoost

Each view of the multi-view data is expect to be able to learn a classifier independently. Combining different views to achieve a better classification model is the key idea of multi-view learning. Due to the noise and strength of different views, a majority voting scheme has been shown to be effective in multi-view learning (Amini et al., 2009). Amini et al. (2009) proposed to use multi-view majority voting to perform multilingual text classification on parallel data.

Algorithm 2 Multi-view Weighted Voting AdaBoost

Input: A multi-view binary training set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, with $y_i \in \{+1, -1\}$.

Output: The final classifier H.

Initialize: $D_1(i) = \frac{1}{n}$.

for $t = 1$ **to** T **do**

- Separately train a set of single view classifiers $\{h_{tv}\}$ using distribution D_t.
- Compute the weights $\{\beta_v\}$ by minimizing the weighted least square loss in Eq. (7).
- Compute the base classifier h_t using the weighted voting scheme in Eq. (6).
- Set $\epsilon_t = Pr_{i \sim D_t}[h_t(\mathbf{x}_i)) \neq y_i]$
- Set $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$.
- Update $D_{t+1}(i) = D_t(i) \frac{e^{-\alpha_t y_i h_t(\mathbf{x}_i)}}{Z_t}$, where Z_t is a normalization factor.

end for

$H(\mathbf{x}) = sign(\sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}))$

They derived a generalization error bound for classifiers learned on examples with multiple artificially views created using machine translation. They empirically evaluated their approach on a comparable multilingual corpus, Reuters RCV1/RCV2, showing that additional views obtained using a machine translation system can significantly increase classification performance, especially when few labeled data are available for training.

We propose to use multi-view majority voting to produce a base learner within the AdaBoost framework, aiming to improve subjectivity classification performance. Given a weighted training set with multiple views, a multi-view instance can be expressed as $\mathbf{x} = (x^1, x^2, \ldots, x^V)$, where each sub-vector x^v provides a representation of the same object in one feature space $\mathcal{X}^v$. In multilingual subjectivity analysis, each view is the textual representation of instances in a given language (e.g. Arabic, English, French, German, Romania, and Spanish). The Multi-view Majority Voting AdaBoost approach is then carried out in the following way. At each round t, a set of view-specific binary classifiers $\{h_{tv}(x^v)\}$ can be separately trained using specific views on the weighted training data with a distribution D_t. The base classifier is then obtained using a majority voting scheme:

$$h_t(\mathbf{x}) = sign(\sum_{v=1}^{V} h_{tv}(x^v)) \tag{5}$$

where $h_{tv}(x^v) \in \{1, -1\}$. The remaining steps of the AdaBoost procedure for training instance reweighting and final combination parameter determination are same as the standard AdaBoost for binary classifications. The overall procedure of the algorithm is described in Algorithm 1. With this algorithm, we expect to integrate the strengths of the subjectivity analysis data in different languages and boost the subjectivity classification performance.

3.3.2 Multi-View Weighted Voting AdaBoost

The multi-view majority voting scheme assumes the set of view-specific classifiers contribute equally for the final classifier. This assumption is too strong in many cases. We thus propose to pursue a more advanced multi-view combination scheme, where the combination classifier is a linear weighted combination of the view-specific classifiers. This leads to our next approach,

a Multi-view Weighted Voting AdaBoost approach. Similar to Multi-view Majority Voting AdaBoost, at each round t of Multi-view Weighted Voting AdaBoost, we separately train a set of single view classifiers $\{h_{tv}\}$ over the training instances with a weighted distribution D_t, one for each language (view) v. But instead of taking a majority vote, we set the combination base classifier h_t as a linear combination of the single view classifiers with a set of weight parameters $\{\beta_v\}$; i.e.,

$$h_t(\mathbf{x}) = sign(\sum_{v=1}^{V} \beta_v h_{tv}(x^v)) \tag{6}$$

where $0 \leq \beta_v \leq 1$ and $\sum_{v=1}^{V} \beta_v = 1$. In order to obtain the weight parameters, we train this linear combination model on the weighted training set, using the outputs of the single-view classifiers as features. Specifically, we minimize the following weighted least square loss on the training instances to to obtain the β values:

$$L = \sum_{i=1}^{n} D(i)(\sum_{v=1}^{V} \beta_v h_{tv}(x_i^v) - y_i)^2 \tag{7}$$

With the obtained base classifier h_t in Eq. (6), we can then find its importance weight α_t and update the distribution D_t. We proceed with this procedure for T rounds and output the final hypothesis H by combining the importance weighted multilingual base classifiers. The overall algorithm is presented in Algorithm 2.

4   Experiments

In this section, we empirically evaluate our proposed approaches for the task of subjectivity analysis on Multilingual Multi-Perspective Question Answering corpus. We first introduce the dataset and describe implementations, and then present the experimental results.

4.1   Multilingual Dataset

We use the same dataset as studied in (Banea et al., 2010). This is a multilingual subjectivity analysis dataset constructed from the Multi-Perspective Question Answering (MPQA) corpus. The original MPQA corpus contains 535 English newswire articles, which are collected from various sources. Each article is manually annotated for subjectivity labels (Wiebe and Cardie, 2005). Although the original corpus is labeled at the phrase and clause levels, we adopt the sentence-level annotations as suggested by (Banea et al., 2008, 2010). If the sentence contains at least one private state, whose opinion strength is higher or medium, we regard it as *subjective*. Otherwise, we see it as *objective*. Thus, we collected 9732 sentences. Among the 9732 sentences in this corpus, 5380 of them are annotated as subjective, while the rest 4352 sentences are labeled as objective.

In order to obtain comparable corpora to MPQA in other languages, Banea et al. (2010) translated the original English (En) newswire articles into five other languages, namely Arabic (Ar), French (Fr), German (De), Romanian (Ro) and Spanish (Es), using machine translation. To generate subjectivity labeling for target languages, they projected the original sentence-level subjectivity annotation from the source English data onto the target language data. Thus, we get a multilingual subjectivity analysis dataset with six languages. [1]

---

[1] The original English MPQA corpus can be downloaded from http://www.cs.pitt.edu/mpqa.

Based on this multilingual corpus, we first performed the following preprocessing steps as (Banea et al., 2010) employed. We removed all the numbers, diacritics, all punctuation marks except ' and -. We kept ' and - because they may mark contractions, such as in "they're" (for English) and "s-ar" (for Romanian). Although Arabic has a different encoding, we treated it in a similar way as the Roman language. We then used the library (Lingua::AR::Word PERL Library) to map Arabic script to Roman-alphabet letters. In the multilingual corpus, there are six languages. Then for each language, we removed the rare words and selected the top 20% of (unigram) features to use as suggested by (Banea et al., 2010). We used term presence as weighting scheme for all features, which means if the sentence contains one specific feature, then its corresponding value for that feature is 1, otherwise, the value is 0. We did this for two reasons. First, (Pang et al., 2002) explored different feature weighting schemes for sentiment classification, showing that term presence is better than term frequency in sentiment classification tasks. Second, (Banea et al., 2010) adopted term presence as a weighting scheme in their experiments.

4.2 Implementation

In the experiments, we compared the empirical performance of the following approaches for subjectivity analysis, including the two proposed approaches.

- TDe, TEn, TEs, TFr, TRo: (Banea et al., 2008) proposed to train a subjectivity classifier for a new language on the translated data in the target language alone. We use *TDe* to denote the method that trains a subjectivity classifier on the target German language alone. Similarly, *TEn, TEs, TFr, TRo* denote that we use English, Spanish, French and Romania as the target language respectively.

- MLS: (Banea et al., 2010) proposed to train a multilingual subjectivity classifier by combining all different languages together to expand the feature space. We denote it as a MultiLingual Space method (MLS).

- MVMV: The multi-view majority voting method developed in (Amini et al., 2009).

- MVAB1: The multi-view majority voting AdaBoost approach.

- MVAB2: The multi-view weighted voting AdaBoost approach.

The last four approaches are multilingual approaches and the others are single language based approaches. For all these approaches, we experimented with two different classifiers: Naïve Bayes (NB) and Support Vector Machines (SVM), which are also used in previous studies on multilingual subjectivity analysis (Banea et al., 2008, 2010). We chose them due to their robust performances and the diversity of their learning methodologies. For Naïve Bayes, we used the multinomial model (McCallum and Nigam, 1998). For Support Vector Machines, we used the LIBLINEAR package (Fan et al., 2008). The LIBLINEAR is an open source library and works very efficiently on large sparse data sets. For LIBLINEAR, we set the tradeoff parameter c with the default value, $c = 1$. For the two boosting approaches, MVAB1 and MVAB2, we set the maximum iteration number T as 50.

4.3 Experimental Results

We take all the subjective sentences as positive instances and all the objective sentences as negative instances. The six single view approaches are experimented on each of the six target lan-

Table 1: Average results for the comparison approaches based on SVM classifiers.

Method	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	AllP	AllR	AllF	MAcc
TEn	75.92	73.78	74.82	68.57	70.96	69.73	72.25	72.37	72.28	72.53
TRo	75.01	73.76	74.37	67.80	69.24	68.51	71.41	71.50	71.44	71.75
TEs	74.04	73.43	73.71	68.26	68.95	68.57	71.15	71.19	71.14	71.39
TFr	75.04	73.00	73.99	67.21	69.48	68.31	71.13	71.24	71.15	71.47
TDe	72.97	71.93	72.44	65.91	67.05	66.46	69.44	69.49	69.45	69.75
TAr	72.70	72.06	72.35	65.76	66.47	66.08	69.23	69.26	69.22	69.55
MLS	76.72	76.00	76.34	70.45	71.29	70.84	73.59	73.65	73.59	73.89
MVMV	76.78	77.99	77.37	72.79	71.36	72.06	74.79	74.68	74.72	75.01
MVAB1	77.95	79.15	78.53	74.68	73.29	73.95	76.32	76.22	76.24	76.47
MVAB2	78.62	79.39	78.98	75.03	74.12	74.54	76.83	76.75	76.76	76.97

guages alone. The four multilingual approaches use all the parallel texts in the six languages. We performed ten-fold cross validation on the multilingual dataset as did in (Banea et al., 2010). Two sets of such experiments are conducted, with Support Vector machines and Naïve Bayes as base classifiers respectively. The average test results of different approaches are reported in Table 1 for SVM and in Table 2 for Naïve Bayes.

Each test result is evaluated with 10 measurements denoted with the following abbreviation style: *Obj* represents objective, *Subj* represents subjective, and *All* stands for overall macro measures, computed over the objective and subjective classes; *P, R, F,* and *MAcc* correspond to precision, recall, F1-measure and macro-accuracy, where

$$Precision = \frac{\# \text{ of correct positive predictions}}{\# \text{ of positive predictions}},$$

$$Recall = \frac{\# \text{ of correct positive predictions}}{\# \text{ of positive instances}},$$

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

From Table 1, we can see that all the four multilingual methods consistently outperform the single-view methods across all languages in terms of all 10 measurements. This verifies the hypothesis that training on translated target language alone is not enough and multilingual subjectivity analysis is useful. Among the four multilingual methods, *MVMV* performs slightly better than *MLS* in terms of subjective precision and objective recall, while *MVAB1* and *MVAB2* significantly outperform both *MLS* and *MVMV* in terms of all ten measurements. The *MVAB2* performs slightly better than *MVAB1* in terms of all ten measurements. Comparing to *MLS*, *MVAB1* improves the accuracy by 2.58%, and improves the macro F1-measure by 2.65%; *MVAB2* improves the accuracy by 3.08%, and improves the macro F1-measure by 3.17%. Comparing to *MVMV*, *MVAB1* improves the accuracy by 1.46% and improves the macro F1-measure by 1.52%; *MVAB2* improves the accuracy by 1.96%, and improves the macro F1-measure by

Table 2: Average results for the comparison approaches based on Naïve Bayes classifiers.

Method	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	AllP	AllR	AllF	MAcc
TEn	74.01	83.64	78.53	75.89	63.68	69.25	74.95	73.66	73.89	74.72
TRo	73.50	82.06	77.54	74.08	63.40	68.33	73.79	72.73	72.94	73.72
TEs	74.02	82.84	78.19	75.11	63.05	69.14	74.57	73.44	73.66	74.44
TFr	73.83	83.03	78.16	75.19	63.61	68.92	74.51	73.32	73.54	74.35
TDe	73.26	83.49	78.04	75.32	62.30	68.19	74.29	72.90	73.12	74.02
TAr	71.98	81.47	76.43	72.62	60.78	66.17	72.30	71.13	71.30	72.22
MLS	75.43	83.66	79.33	76.64	66.30	71.10	76.04	74.98	75.21	75.89
MVMV	75.91	84.56	79.98	77.47	66.38	71.46	76.69	75.47	75.72	76.49
MVAB1	76.95	85.49	80.98	78.92	67.91	72.98	77.93	76.70	76.98	77.68
MVAB2	77.74	85.73	81.53	78.96	68.52	73.34	78.35	77.13	77.44	78.19

2.04%. It is worth mentioning that both *MVAB1* and *MVAB2* increase the overall precision and recall levels to above 76%. Using a paired t-test, all these improvements were found to be significant at p=0.001. Those improvements demonstrate that our Multi-view AdaBoost approaches are more effective than simply expanding the feature space with multiple languages, or only using a simple multi-view majority voting strategy.

From Table 2 we can see that again the four multilingual methods outperform the single language methods in terms of macro precision, recall and F1-measure as well as macro accuracy. However, *MLS* achieves almost the same performance in terms of subjective recall measurement as the single language methods, *TEn* and *TDe*, which implied that more advanced multilingual models are needed. The two proposed multi-view AdaBoost approaches, *MVAB1* and *MVAB2*, outperform all the other comparison methods in terms of macro precision, and F1-measure as well as macro accuracy. Even in term of recall for subjective and objective classes, both of *MVAB1* and *MVAB2* outperform all other methods. Comparing to *MLS*, *MVAB1* improves the accuracy by 1.79%, and improves the macro F1-measure by 1.77%; *MVAB2* improves the accuracy by 2.30%, and improves the macro F1-measure by 2.23%. Comparing to *MVMV*, *MVAB1* improves the accuracy by 1.19%, and improves the macro F1-measure by 1.26%; *MVAB2* improves the accuracy by 1.70%, and improves the macro F1-measure by 1.72%. All the improvements were found to be significant at p=0.001 by using a paired t-test. These results demonstrate that our proposed multi-view AdaBoost approaches are robust to different types of base classifiers, and their advantages can be maintained. This again suggests the proposed multi-view AdaBoost approaches provide a more effective framework to exploit multilingual information for multilingual subjectivity analysis than the existing methods.

4.4 Impact of Training Size

Next we studied the impact of the training size for different approaches. Among the 9732 sentences, we randomly selected 2732 (about 1500 subjective sentences and 1232 objective sentences) as test data. From the rest 7000 sentences, we randomly selected training instances with a range of different sizes $m \in \{500, 1000, 3000, 7000\}$. For each training size, we repeated
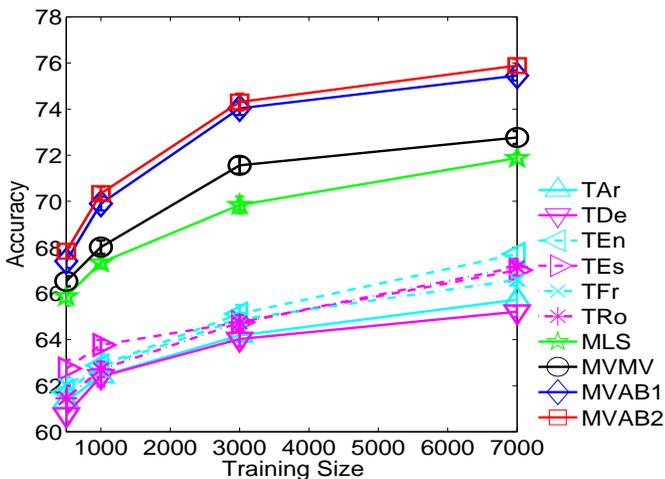
(a) Results based on SVM classifiers.
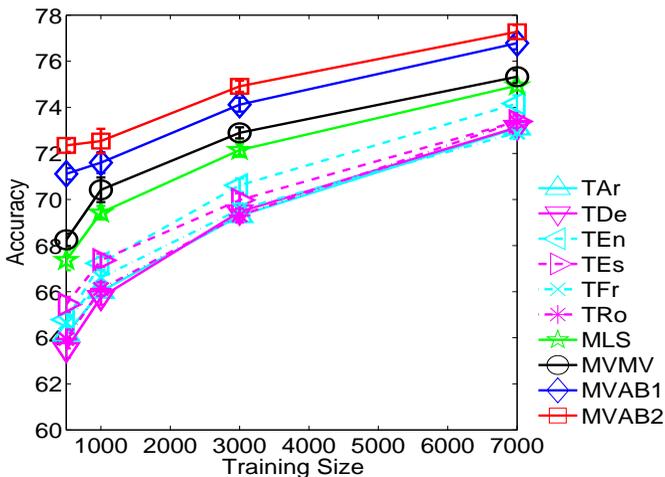
(b) Results based on Naïve Bayes classifiers.

Figure 1: Average classification accuracies ± standard deviations across a range of different training sizes. The top figure presents results based on Support Vector Machine classifiers and the bottom figure presents results based on Naïve Bayes classifiers.

the experiments over 10 runs by randomly selecting the training instances from the 7000 sentences. We again experimented with the same two base classifiers: SVM and Naïve Bayes (NB). We compared the four multilingual methods with the six single language methods. We reported the average test accuracies and standard deviations in Figure 1. We can see that for both SVM-based classifiers and Naïve Bayes-based classifiers, the four multilingual methods consistently outperform the six single language methods across the range of different training sizes. The improvements of multilingual methods over the single language methods are clearly significant across all range of training sizes for both SVM-based classifiers and Naïve Bayes-based classifiers, which justified that using multilingual information enables every single

Table 3: Results for Statistical Significance (McNemar's) test.

Null Hypothesis	p-value (SVM, Trainsize=500)	p-value (NB, Trainsize=7000)
MVAB1 vs. TEn	6.3×10^{-10}	4.1×10^{-9}
MVAB1 vs. MLS	7.9×10^{-7}	8.5×10^{-8}
MVAB1 vs. MVMV	3.5×10^{-4}	2.7×10^{-5}
MVAB2 vs. TEn	1.3×10^{-10}	1.8×10^{-9}
MVAB2 vs. MLS	2.6×10^{-7}	3.1×10^{-8}
MVAB2 vs. MVMV	1.2×10^{-4}	1.9×10^{-5}

language to reach a high accuracy that is not individually obtainable. Among the four multilingual methods, our proposed two multi-view AdaBoost methods, *MVAB1* and *MVAB2*, significantly outperform the other two simple multi-view methods, *MLS* and *MVMV*. With the more sophisticated view weighted training, *MVAB2* outperforms *MVAB1*. More specifically, for the SVM-based classifiers, when the training size is 3000, *MVAB1* increases the accuracy by 4.19% over *MLS* and by 2.47% over *MVMV*; *MVAB2* increases the accuracy by 4.50% over *MLS* and by 2.75% over *MVMV*. Their advantages over the single language methods are even much larger. *MVAB1* increases the accuracy of the best single language method (over English) by 8.91%. *MVAB2* increases the accuracy of the best single language method (over English) by 9.19%. For NB-based classifiers, our proposed approaches achieve their best improvements when the training size is small (500). In this case, *MVAB1* increases the accuracy by 6.33% comparing to the best single language method over English, increases the accuracy by 3.74% comparing to *MLS*, and increases the accuracy by 2.86% comparing to *MVMV*. *MVAB2* performs even better, and it increases the accuracy by 7.56% comparing to the best single language method over English, increases the accuracy by 4.98% comparing to *MLS*, and increases the accuracy by 4.10% comparing to *MVMV*.

To justify the significance of those improvements across ranges, we used a McNemar paired test for labeling disagreements (Gillick and Cox, 1989) with $p < 0.05$ being significant. We focus on the improvements the two proposed approaches obtained over other methods. Since *TEn* works the best among models trained with monolingual corpus, we select it as the representative for all six monolingual methods. From Figure 1, we can see that the two proposed approaches achieve the smallest improvement when the training size is very small (500) for SVM-based classifiers, and when the training size is very big (7000) for NB-based classifiers. We thus select those results to conduct significance test. We report the p values in Table 3. From Table 3, we can see that all the improvements made by *MVAB1* and *MVAB2* over other methods are statistically significant. These results again demonstrate the efficacy of the proposed multi-view AdaBoost framework on multilingual subjectivity analysis.

Conclusion

In this paper, we proposed to integrate multi-view learning into the AdaBoost framework to perform multilingual subjectivity analysis, with the aim of developing more effective subjectivity analysis tools for both English and languages beyond English. Our experimental results on multilingual MPQA corpus show that the approaches developed within our proposed framework can significantly outperform other existing methods.

References

Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., and King, J. (2011). How can you say such things?!?: recognizing disagreement in informal political argument. In *Proc. of the Workshop on Languages in Social Media*.

Abdul-Mageed, M. and Diab, M. T. (2011). Subjectivity and sentiment annotation of modern standard arabic newswire. In *Proc. of the 5th Linguistic Annotation Workshop*.

Abdul-Mageed, M., Diab, M. T., and Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard arabic. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Alm, C. O. (2011). Subjective natural language problems: motivations, applications, characterizations, and implications. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Amini, M., Usunier, N., and Goutte, C. (2009). Learning from multiple partially observed views - an application to multilingual text categorization. In *Advances in Neural Information Processing Systems (NIPS)*.

Banea, C., Mihalcea, R., and Wiebe, J. (2010). Multilingual subjectivity: are more languages better? In *Proc. of the International Conference on Computational Linguistics (COLING)*.

Banea, C., Mihalcea, R., Wiebe, J., and Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Banfield, A. (1982). *Unspeakable sentences: Narration and representation in the language of fiction*. Routledge and Kegan Paul.

Carenini, G., Ng, R., and Zhou, X. (2008). Summarizing emails with conversational cohesion and subjectivity. In *Proc. of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*.

Fan, R., Chang, K., Hsieh, C., Wang, X., and Lin, C. (2008). Liblinear: A library for large linear classification. *Journal of Machine Learning Research (JMLR)*, 9:1871–1874.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Gillick, L. and Cox, S. (1989). Some statistical issues in the comparison of speech recognition algorithms. In *Proc. of the Inter. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proc. of the ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*.

Li, B., Liu, Y., Ram, A., Garcia, E. V., and Agichtein, E. (2008). Exploring question subjectivity prediction in community qa. In *Proc. of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Lu, B., Tan, C., Cardie, C., and Tsou, B. (2011). Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

McCallum, A. and Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *Proc. of AAAI-98 Workshop on Learning for Text Categorization*.

Mihalcea, R. and Banea, C. (2007). Learning multilingual subjective language via cross-lingual projections. In *Proc. of the Annual Meeting of the Association of Computational Linguistics (ACL)*.

Miller, G. A., Leacock, C., Tengi, R., and Bunker, R. T. (1993). A semantic concordance. In *Proc. of the DARPA workshop on Human Language Technology (HLT)*.

Pang, B., Lee, L., Rd, H., and Jose, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Riloff, E., Wiebe, J., and Phillips, W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proc. of the National Conference on Artificial intelligence (AAAI)*.

Saffari, A., Leistner, C., Godec, M., and Bischof, H. (2010). Robust multi-view boosting with priors. In *Proc. of the European conference on computer vision (ECCV)*.

Seki, Y., Eguchi, K., Kando, N., and Aono, M. (2005). Multi-document summarization with subjectivity analysis at DUC 2005. In *Proc. of 2005 Document Understanding Conference*.

Somasundaran, S. and Wiebe, J. (2009). Recognizing stances in online debates. In *Proc. 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*.

Somasundaran, S. and Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proc. of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.

Wan, X. (2009). Co-training for cross-lingual sentiment classification. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Wiebe, J. and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. In *Language Resources and Evaluation*.

Wiebe, J. and Mihalcea, R. (2006). Word sense and subjectivity. In *Proc. of the Joint Conference of the International Committee on Computational Linguistics on Computational Linguistics and the Association for Computational Linguistics (COLING-ACL)*.

Wiebe, J. M. (1994). Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

Xu, Z. and Sun, S. (2010). An algorithm on multi-view adaboost. In *Proc. of the international conference on Neural information processing: theory and algorithms*.

Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proc. of the 2003 conference on Empirical methods in natural language processing (EMNLP)*.