Cross-lingual Twitter Polarity Detection via Projection across Word-Aligned Corpora

Mohamed Abdel-Hady, Riham Mansour, and Ahmed Ashour
{Mohamed.Abdel-Hady,rihamma,v-aash}@microsoft.com

Microsoft Research, Advanced Technology Lab in Cairo, Egypt

Abstract. In this paper, we propose an unsupervised framework that leverages the sentiment resources and tools available in English language to automatically generate stand-alone polarity lexicons and classifiers for languages with scarce subjectivity resources and thus avoids the need for labor intensive manual annotation. Starting with a list of English sentiment-bearing words, we expand this lexicon using WordNet synsets. For each sentence pair in a given bilingual parallel corpus, the high-precision English polarity lexicon is applied to the English side then the output sentiment label is projected onto the target language side via statistically derived word alignments. The resulting lexicon is applied to a large pool of unlabeled tweets in the target language, in order to automatically label tweets as training data to train polarity classifier. Our experiments using Spanish and Portuguese as target ones have shown that the resulting classifiers help to improve polarity classification performance compared to lexicon-based classification for under-resourced languages in social media.

1 Introduction

Twitter is a popular micro-blogging service where users post status messages (called tweets). The users use tweets to share their personal feelings and sometimes express opinions in the form of user-defined hashtags, emoticons or normal words about different topics such as events, movies, products or celebrities. There has been a large amount of NLP research on this user-generated content in the area of sentiment classification [1, 2]. Traditionally most of the research work has focused on large pieces of text, such as product and movie reviews that represent summarized thoughts of authors. Although tweets became publicly available for the research community, they are different from reviews primarily because they are limited to 140 characters and have a more colloquial linguistic style. The frequency of misspellings and slang in tweets is much higher than in reviews.

The sentiment classification can be formulated either as a lexicon-based task that requires an extensive set of manually supplied sentiment words or a supervised machine learning task that requires a large amount of hand-labeled training tweets. While English is still the top language in Twitter, it is no longer the majority. The manual annotation of training tweets for each language is expensive, error-prone and time consuming. The sentiment analysis research has shown that a two-stage approach is more effective [1]: the first stage is subjectivity classification in which subjective instances are distinguished from objective ones, then whether the subjective instances has "positive" or "negative" polarity is detected.

Cross-language sentiment classification aims to leverage the existing sentiment analysis resources and tools, such as lexicons and classifiers, available in a source language such as English in order to automate building these resources and tools for under-resourced languages. For instance, the automatic generation of lexicons and training data. The motivation is that people all over the globe use extremely diverse languages to express their opinions. Consumers can use sentiment analysis to research products or services before making a purchase decision. On the other side, companies can automatically gather customer feedback and opinions about their products and services in different countries [2].

In this paper, we propose an *unsupervised method* to automatically generate resources for polarity analysis for a new target language by leveraging the resources and tools available for English. Specifically, we use cross-lingual projection of polarity labels using bi-lingual parallel corpora to create a lexicon in the target language. The lexicon is then used to auto-label some tweets to bootstrap a classifier that could auto-label more tweets. A combination of the lexicon-based classifier and the corpus-based classifier is used then to auto-label tweets. To our knowledge, the proposed approach is the first to examine the multilingual setting with the combination of lexicon-based and corpus-based classifiers on tweets. Further, we examine a self-training approach of creating a corpus-based classifier in the target language based on the output of a translated lexicon-based classifier. This differentiates our proposed work significantly from the work presented by [3] as their approach embraces a corpus-based classifier trained on a labeled corpus translated from the source language. Our results section depicts a comparison with the machine translation approaches.

2 Related Work

There is a lot of research work investigating how to automatically extract sentiment from text. While traditional work [4] focused on movie reviews, more recent research has explored social networks for sentiment analysis. The methods involved differ somewhat since texts like tweets have a different purpose and a more colloquial linguistic style [5]. Go et al. [6] have trained a sentiment classifier to label tweets sentiment polarities as positive or negative. Pak et al. [7] trained classifiers to also detect neutral tweets that do not contain sentiment. More recent work has focused on concept-level sentiment analysis [8] to bridge the cognitive and affective gap between the words in their bare meanings and the sentiments of the concepts conveyed by the words.

Sentiment classifier training requires a large amount of labeled training data, but the manual annotation of tweets is expensive and time-consuming. To gather training data, Go, Pak and others used a heuristic method introduced by Read [9] to assign sentiment labels to tweets based on emoticons instead. To train a sentiment classifier, the source texts first have to be converted into some type of features. The most prominent features are n-grams, polarity lexicons, part-of-speech tags and special micro-blogging features [1] that include among other emoticons, hashtags, punctuation and character repetitions and words in capital letters. Because the language used on Twitter is often informal and differs from traditional text types [5], most approaches include a preprocessing step. Usually emoticons are detected, URLs removed, abbreviations expanded and twitter markup is removed or replaced by markers.

Go et al. [6] compared different machine learning techniques such as naive Bayes, maximum entropy and Support Vector Machines (SVM), with unigram, bigram and part-of-speech features. Their training data consisted of 1.6M tweets equally split between positive and negative classes. The evaluation was performed on 359 hand-annotated tweets (182 positive and 177 negative tweets). Pak and Paroubek [7] collected 300k training tweets and performed evaluation on 216 hand-annotated tweets. They conducted experiments with multinomial naive Bayes, SVM and Conditional Random Field classifiers using features such as unigrams, bigrams and trigrams; multinomial naive Bayes using bigrams was the best.

Barbosa and Feng [1] followed the two-stage approach and instead of n-gram features they used part-of-speech tags, lexical and microblogging features to build SVM classifiers. Zhang et al. [10] combined a lexicon-based classifier with an SVM to increase the recall of the classification. The most representative system is introduced by [11], which is the state-of-the-art system that performed best in SemEval 2013 Twitter Sentiment Classification Track. The system implemented a number of effective hand-crafted features.

Multilingual sentiment classification approaches often classify sentiment by using cross-lingual training [12, 3, 13] with classification approaches for English texts. This however requires resources, like parallel corpora, that bridge between English and every target language. Other approaches rely on machine-translation to first translate texts into English, and applying English classification techniques [14] to the translated texts. Narr et al. [15] presented a semi-supervised approach that use emoticons as noisy labels to generate training data from a large set of unlabeled tweets. They use a two-class naive Bayes classifier with n-gram features for language-independent sentiment classification. Davido et al. [16] presented an unsupervised sentiment classification framework which utilizes 50 Twitter hashtags and 15 smileys to automatically annotate training tweets to identify various sentiment types without any labor-intensive manually labeled data or pre-provided polarity lexicons. They used a k-nearest neighbors (kNN) classifier.

3 Baselines

3.1 Lexicon-based Polarity Classifier

The first approach to generate a target-language polarity classifier is to create a polarity lexicon by translating an existing source language lexicon either using machine translation system or word-aligned parallel corpus. Starting with the target-language resulting lexicon, we trained a lexical classifier similar to the one introduced by [17]. At the core of this method is a high-precision subjectivity lexicon that can label large amounts of raw text. Their method is further improved with a bootstrapping process that learns extraction patterns. In our experiments, however, we apply only the rule-based classification step, since the extraction step requires tools for syntactic parsing and information extraction in the target languages. The classifier relies on two main heuristics to label tweets as positive or negative : (1) if at least one strong positive term occurs in the tweet, it is labeled positive; (2) if at least one strong negative term occurs in the tweet, it is labeled negative.

MPQA Lexicon (L_{MPQA}) We adopted the subjectivity lexicon[1] from OpinionFinder [18], an English subjectivity analysis system which classifies sentences as subjective or objective. The lexicon was compiled from manually developed resources augmented with entries learned from corpora. It contains 6,856 entries including 990 multi-word expressions. The entries in the lexicon have been labeled (1) for polarity either positive, negative, or neutral, (2) for part of speech, and (3) for reliability those that appear most often in subjective contexts are strong clues of subjectivity, while those appearing less often are labeled weak.

Expanding a Polarity Lexicon (L_{M+W}) In order to increase the coverage of the lexicon before we start the translation or cross-lingual projection, we adopt a approach similar to [19]. We used synonym relations from English WordNet[2] to expand the initial seed English polarity lexicon defined in Section 3.1. The assumption is that synonym carries same sentiment/polarity as compared to the root word. We make a hypothesis of traversing WordNet like a graph where words are connected to each other based on synonym or antonym relations. Consider each word in this list as a node of the graph. Each node has many in-links and many out-links. This is an undirected graph which is not fully connected i.e. not all the nodes are connected to every other node. For every word in the seed lexicon, we identify its synonyms and append with appropriate polarity label in the seed lexicon. Unlike [19], we performed one iteration of traversal, we did not identify the synonyms of the seed lexicon words synonyms. As a post-processing step, we exclude any term that appears more than once with different polarity.

Machine Translation of a Polarity Lexicon (L_{M+W+MT}) Terms that are strongly subjective are translated from English to the new language using *Microsoft Translator*[3], with term polarity projected from the English to the term translation. There were several challenges encountered in the translation process. Each English word has multiple translations and machine translation systems

[1] http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
[2] http://wordnet.princeton.edu/
[3] http://www.bing.com/translator

rely on the context of a word in order to select the most relevant translation. Unfortunately, for lexicon translation, words are translated without any context where the translation has to rely on the most probable sense in the target language. Some words may lose their subjective meaning once translated. That is, an English word may have positive polarity but its context-independent translation in the target language is objective or negative, e.g., although the term "ego" bears negative polarity in English, its Portuguese translation is "eu" which is a neutral word means "I". Moreover, the lexicon sometimes includes identical entries expressed through different parts of speech, e.g., "grudge" has two separate entries, for its noun and verb roles, respectively. On the other hand, the machine translation system does not make this distinction, and therefore we have again to rely on the most frequent heuristic.

3.2 A Corpus-Based Polarity Classifier

Some tweets are detrimental to the lexicon-based approach. For instance, emoticons, colloquial expressions, abbreviations, hashtags are frequently used in tweets. Although these expressions may have sentiment orientation, they do not exist in the polarity lexicon. Consider the tweet, *"I have watched Messi yesterday, just lovvee him :-)"*. It clearly expresses a positive opinion on Messi by the word *lovvee* and the emoticon *:-)*. But the lexicon-based method would regard the tweet as expressing no/neutral opinion on Messi, since there is not a polarity word in the tweet. This leads to the low recall problem for the lexicon-based approach despite of its high precision, which depends entirely on the presence of polarity words to determine the sentiment orientation. In this section we describe two baseline approaches that label tweets automatically to deal with the low recall problem of the lexicon-based approach. These approaches then train a classifier on the auto-labeled tweets.

Emoticon Heuristic ($C_{emoticons}$) To generate the training set of subjective tweets and automatically annotate them by polarity, we used a method similar to that of Pak and Paroubek [7]. We assign noisy polarity labels to unlabeled tweets based on the existence of positive or negative emoticons using the manually-labeled emoticon lexicon in [20]. If a tweet contains at least one happy emoticon, and no negative ones, we assign it to the positive class, and vice versa. Because of the shortness of tweets, we assume here that the sentiment of a smiley applies to the whole text of the tweet. Although this may be wrong in some cases, it simplifies the analysis significantly. We used a range of emoticons that are very distinctly positive or negative as noisy labels.

Machine Translation of Polarity Tweets ($C_{manual+MT}$) We explore the possibility of using machine translation of English polarity tweets to generate the training data required to build polarity classifier in a target language, similar to the method of Banea et al. [13] used to create an annotated corpus in the Romanian language through the translation of English-language news articles. In our experiments, we assume the existence of a set of tweets manually annotated for polarity in the English language, We obtain a set of training examples in the target language through machine translation using *Microsoft Translator* and project the polarity labels from each English tweet to its target translation. Finally, we train a polarity classifier in the target language.

4 Cross-lingual Projection via Word-Aligned Corpus

In this section, we present our proposed cross-lingual polarity projection approach that depends on the existence of word-aligned parallel corpus. The architecture of our training framework is presented in Figure 1.

Word Alignment of Parallel Corpus Sentence alignment and word alignment is performed on a given bilingual parallel corpus. First, sentence level alignment is performed then we applied word dependent transition model based HMM (WDHMM) for word alignment [21]. Compared to the baseline HMM, the WDHMM can reduce alignment error rate by more than 13%. It even outperforms IBM model 4 after two direction word alignment combination. WDHMM is run on the bilingual corpus in both directions for forward and backward alignment then we merge the results.

Polarity Labels Projection ($L_{M+W+align}$) We project the polarity labels to the target side of the parallel corpus to automatically annotate target language sentences, according to the result of word alignment. Table 1 presents a sample of parallel sentences where both source and target sentiment-bearing are bold faced. Each source word/phrase is aligned with multiple target word/phrase. We count the number of times a source word/phrase is aligned to a target word/phrase. The output is a polarity lexicon in the target language including the most frequent target word/phrase for each source word/phrase.

4.1 Combining Lexicon-based and Corpus-based Approaches ($C_{M+W+align}$)

In our proposed approach, we combine the lexicon-based method that can give high precision, but low recall with the corpus-based approaches. First, the polarity lexicon is applied to annotate thousands of unlabeled tweets in English, Spanish and Portuguese, described in Table 2. The tweets automatically annotated as positive and negative are then used as training data to build corpus-based polarity classifier. Instead of being labeled manually, the training examples are given by the lexicon-based approach defined in Section 3.1.

4.2 Decision Fusion of Lexicon-based and Corpus-based Approaches ($E_{M+W+align}$)

Figure 2 presents our hierarchical approach to combine the decisions of lexicon-based and corpus-based polarity classifiers. A tweet is first classified by the

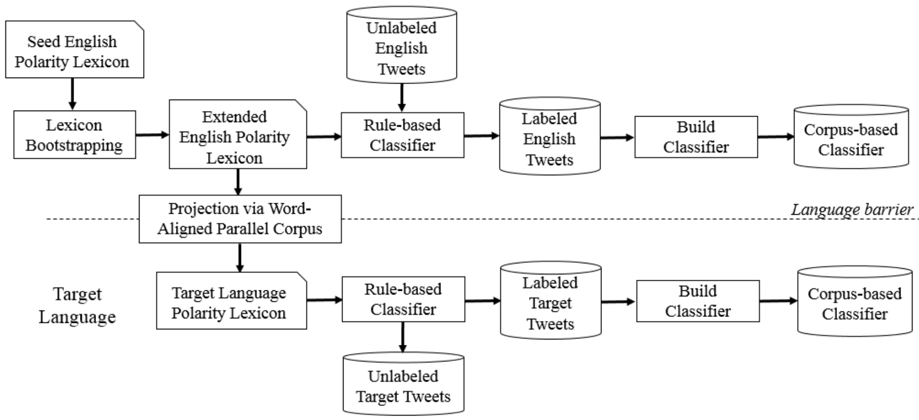

Fig. 1: Training Phase: Combining Lexicon-based and Corpus-based Approaches

polarity lexicon, as defined in Section 3.1. If the tweet is labeled objective, it is classified by the polarity classifier trained as defined in Section 4.1. The classifier relies on three heuristics to label input tweet: (1) if the confidence of the classifier is greater than a predefined threshold θ_1, the tweet is labeled positive; (2) if the confidence is less than a predefined threshold θ_2, the tweet is labeled negative; (3) if none of the previous rules apply, the tweet is labeled unknown. Our experiments show that choosing θ_1 to be 0.8 and θ_2 to be 0.2 retains the best results.

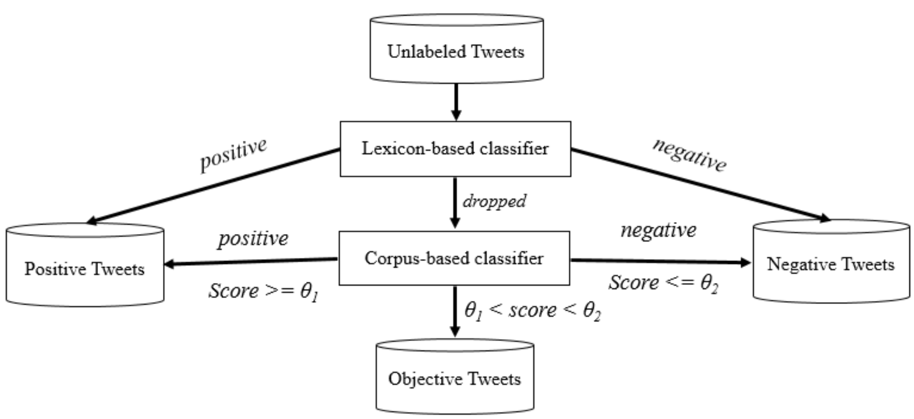

Fig. 2: Decision Fusion of Lexicon-based and Corpus-based Approaches

English Sentence	Spanish Sentence
I **love** *the local cuisine*	*Me* **encanta** *la comida tpica*
That was **delicious***!*	*Estaba* **delicioso***!*
I'm quite **disappointed***.*	*Estoy bastante* **decepcionado***.*
I thought it was **crazy**	*Pienso que fue* **loco**
Congratulations on **excellent** *project!*	*Felicitaciones por* **excelente** *proyecto!*

(a) from English to Spanish

English Sentence	Portuguese Sentence
Who's your **favourite** *sportsperson?*	*Qual o seu atleta* **favorito***?*
Who's your **favourite** *team?*	*Qual a sua equipa* **favorita***?*
That was a **boring** *game!*	*Foi um jogo* **chato***!*
meet the standard of **good** *practice*	*Respeitar as regras de* **boas** *prticas*

(b) from English to Portuguese

Table 1: Alignment-based projection of sentiment labels via Word-Aligned Parallel Corpora

5 Empirical Evaluation

5.1 Setup

For evaluation, we compare between our proposed framework and the baselines described in Section 3. C_{M+W} is the classifier trained on tweets that are auto-labeled by lexicon L_{M+W}. E_{M+W} represents the decision fusion of lexicon L_{M+W} and classifier C_{M+W}. C_{manual} represents the classifier trained on manually-annotated tweets. This fully-supervised method is considered the upper bound for comparison with the other unsupervised approaches since the effectiveness of any unsupervised method depends on the gap between its performance and the supervised method. We employ two performance evaluation metrics: namely the F-measure and the subjectivity coverage, which reflects the number of instances in the test set that the approach is able to provide a polarity decision for.

5.2 Data

Table 2 describes the unlabeled, auto-labeled and manually-labeled training tweets and test tweets that we used for both supervised and unsupervised polarity classification in different languages. We evaluate the results against three gold-standard sets of English, Spanish[4] and Portuguese tweets manually annotated for subjectivity. The distribution of the two classes (positive and negative) in the test sets is illustrated in Table 2. We used two bilingual parallel corpora for cross-lingual projection of polarity labels. The English-Spanish parallel corpus is 20 million parallel sentences and the English-Portuguese corpus contains

[4] http://www.daedalus.es/TASS2013/corpus.php

15 million parallel sentences. The corpora involve a variety of publicly available data sets namely the United Nations proceedings[5], proceedings of the European Parliament[6], Canadian Hansards[7] and web crawled data.

Table 2: Data description for the three languages

Language	Unlabeled	Auto-labeled		Manually labeled		Test Set	
		Positive	Negative	Positive	Negative	Positive	Negative
English	700,378	77,736	43,428	1,603	546	166	62
Spanish	148,261	7,432	2,123	14,819	8,434	327	68
Portuguese	172,406	8,915	2,715	1,634	738	721	293

5.3 Training

We train a maximum entropy binary classifier to assign polarity label to subjective tweets. We used a maximum entropy classifier as our supervised learning algorithm. Our basic features are unigrams and bigrams. We apply feature reduction using Log Liklihood Ratio (LLR) to select the top 20K features that highly co-relate with the training data. We also added emoticons and hashtags as features which are specific to the Twitter data. All feature types are combined into a single feature vector. Pang et al. [4] have shown that feature presence (binary value) is more useful than feature frequency for the SVM classifier. Therefore, we use binary feature presence instead of feature frequency. Training data are the tweets labeled by the below techniques.

5.4 Results

The results obtained by running the experiments on English, Spanish and Portuguese are shown in Table 3, Table 4 and Table 5, respectively.

For English Our baselines for English in the lexicon-based approaches is L_{MPQA} and in the corpus-based approaches is the $C_{emoticons}$ auto-labeling approach. Our proposed WordNet expansion of lexicon L_{M+W} increases the coverage from 40% to 52%. The unsupervised classifiers $C_{M+W(0.8,0.2)}$ and $E_{M+W(0.8,0.2)}$ trained based on lexicon L_{M+W} achieves a coverage increase of 35% over L_{MPQA}. When comparing the unsupervised classifiers $C_{M+W(0.8,0.2)}$ and $E_{M+W(0.8,0.2)}$ with the auto-labeling baseline $C_{emoticons}$, their negative F1 showed significant gain of 52% and 50% over the emoticons baseline respectively. We return the low F1 of the emoticons baseline in the negative class to the fact that negative emoticons are very few, noisy and inaccurate and the number of training instances collected with these emoticons are few.

[5] http://catalog.ldc.upenn.edu/LDC94T4A
[6] http://www.statmt.org/europarl/
[7] http://www.isi.edu/natural-language/download/hansard/

Table 3: Performance of polarity classification in English language

Experiment	Coverage	Positive			Negative		
		Precision	Recall	F1	Precision	Recall	F1
C_{manual} (upper bound)	100	87.17	98.19	92.35	92.68	61.29	73.79
$C_{emoticons}$	100	72.44	98.19	83.37	71.43	8.06	14.49
L_{MPQA}	40	91.30	88.73	90	63.63	70	66.66
L_{M+W}	52	91.95	91.95	91.95	78.13	78.13	78.13
$C_{M+W(0.8,0.2)}$	75	86.29	84.43	85.35	63.46	68.75	66
$E_{M+W(0.8,0.2)}$	75	86.55	84.43	85.48	62.75	66.67	64.65

Table 4: Performance of polarity classification in Spanish language

Experiment	Coverage	Positive			Negative		
		Precision	Recall	F1	Precision	Recall	F1
C_{manual} (upper bound)	100	94.51	94.80	94.65	74.62	73.52	74.07
$C_{emoticons}$	100	83.64	96.94	89.80	37.50	8.82	14.28
$C_{manual+MT}$	100	86.73	84.71	85.71	43.18	55	48.38
L_{MPQA}	33	93.54	77.67	84.87	30.55	64.70	41.50
L_{M+W+MT}	41	87.60	80.92	84.13	40.48	53.13	45.95
$L_{M+W+align}$	44	96.95	84.67	90.39	46.51	83.33	59.70
$C_{M+W+align(0.8,0.2)}$	71	91.63	81.94	86.51	46.75	67.92	55.38
$E_{M+W+align(0.8,0.2)}$	73	91.30	81.12	85.91	45.68	67.27	54.41

For Spanish The coverage of $L_{M+W+align}$ is more than that of L_{M+W+MT} by 3% and than that of L_{MPQA} by 11%. $C_{M+W+align(0.8,0.2)}$ has an increase over L_{MPQA} by 38% in the coverage. The same unsupervised classifier has an increase over the L_{M+W+MT} baseline by 30% in coverage. The other unsupervised classifier $E_{M+W+align(0.8,0.2)}$ outperforms both L_{MPQA} and L_{M+W+MT} baselines by 9% in coverage. The two unsupervised classifiers $C_{M+W+align(0.8,0.2)}$ and $E_{M+W+align(0.8,0.2)}$ have an F1 that is less than the $C_{emoticons}$ baseline by 3% and 4% in the positive class respectively. However, the both classifiers significantly outperforms the $C_{emoticons}$baseline by 41% and 40% in the negative class F1 respectively. $C_{M+W+align(0.8,0.2)}$ and $E_{M+W+align(0.8,0.2)}$ have almost the same positive F1 as the $C_{manual+MT}$ baseline while they showed gain of 7% and 6% respectively for the negative F1.

For Portuguese The coverage of $L_{M+W+align}$ is more than that of L_{M+W+MT} by 15% and than that of L_{MPQA} by 20%. $C_{M+W+align(0.8,0.2)}$ has an increase over L_{MPQA} by 25% in coverage , and a decrease of 3% in the negative class F1. The same unsupervised classifier has an increase over L_{M+W+MT} by 22% in coverage. The other unsupervised classifier $E_{M+W+align(0.8,0.2)}$ outperforms both L_{MPQA} and L_{M+W+MT} baselines by 40% and 33% in the coverage respectively. $C_{M+W+align(0.8,0.2)}$ and $E_{M+W+align(0.8,0.2)}$ have an F1 that is more than the $C_{emoticons}$ baseline by 24% and 28% in the negative class F1 respectively.

Table 5: Performance of polarity classification on Portuguese language

Experiment	Coverage	Positive			Negative		
		Precision	Recall	F1	Precision	Recall	F1
C_{manual} (upper bound)	100	91.41	90.01	90.70	76.31	79.18	77.72
$C_{emoticons}$	100	64.83	93.52	76.58	71.42	24.21	36.16
$C_{manual+MT}$	100	74.72	92.23	82.56	54.84	23.21	32.62
L_{MPQA}	39	90.74	71.28	79.84	52.02	81.08	63.38
L_{M+W+MT}	44	83.71	69.94	76.21	48.92	67.91	56.87
$L_{M+W+align}$	59	90.98	83.06	86.84	65.38	79.53	71.77
$C_{M+W+align(0.8,0.2)}$	64	83.69	85.13	84.40	61.67	59.04	60.33
$E_{M+W+align(0.8,0.2)}$	69	84.07	83.20	83.63	63.84	65.30	64.56

$C_{M+W+align(0.8,0.2)}$ and $E_{M+W+align(0.8,0.2)}$ showed significant gain of 28% and 32% respectively for the negative F1.

5.5 Discussion

Based on our experiments, we can conclude that cross-lingual projection via word-aligned parallel corpora offers a viable approach to generating resources for subjectivity annotation in an under-resourced target language. The lower F1 for the positive and negative classes of $C_{emoticons}$ classifiers in the three languages indicate that the existence of emoticons in a tweet doesn't imply the tweet polarity directly. It is a fact in the literature that an ensemble of diverse and accurate classifiers outperforms its ensemble members. Error diversity means that the members have different misclassified objects. For the three languages, the decision fusion of polarity lexicon and Maximum Entropy classifier doesn't show gains. The reason is that the classifier and the lexicon are not diverse since the former is trained on the tweets labeled by the latter.

6 Conclusions

In this paper, we explored the use of word-aligned parallel corpora for creating resources and tools for subjectivity analysis in other languages, by leveraging on the resources available in English. We introduced and evaluated different approaches to generate subjectivity lexicons and subjectivity annotated corpora in a target language, and exemplified the technique on Spanish and Portuguese. The experiments show promising results, as they are comparable to those obtained using manually annotated tweets.

References

1. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). (2010) 3644

2. Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T., Haruechaiyasak, C.: Discovering consumer insight from twitter via sentiment analysis. Journal of Universal Computer Science **18(8)** (2012) 973–992
3. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: Annual Meeting Association for Computational Linguistics (ACL '07). (2007)
4. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. (2002) 79–86
5. Han, B., Baldwin, T.: Lexical normalisation of short text messages: Makn sens a# twitter. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. (2011)
6. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Volume 150., Ainsworth Press Ltd (2009) 1–6
7. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC10). (2010)
8. Cambria, E., Olsher, D., Rajagopal, D.: Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI, Quebec City (2014)
9. Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL Student Research Workshop, ACLstudent 05. (2005) 4348
10. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining lexiconbased and learning-based methods for twitter sentiment analysis. Volume 89. (2011)
11. Mohammad, S.M., Kiritchenko, S., Zhu, X.: Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:1308.6242 (2013)
12. Wan, X.: Co-training for cross-lingual sentiment classification. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing. (2009) 235243
13. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual subjectivity analysis using machine translation. In: EMNLP. (2008)
14. Denecke, K.: Using sentiwordnet for multilingual sentiment analysis. In: IEEE 24th International Conference on Data Engineering Workshop (ICDEW 2008). (2008) 507512
15. Narr, S., Hülfenhaus, M., Albayrak, S.: Language-independent twitter sentiment analysis. (2012)
16. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Proceedings of the 23rd international conference on Computational Linguistics (COLING). (2010)
17. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of EMNLP. (2003)
18. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP 2005. (2005)
19. Bakliwal, A., Arora, P., Varma, V.: Hindi subjective lexicon: A lexical resource for hindi adjective polarity classification. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). (2012)
20. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media. (2011) 30–38
21. He, X.: Using word-dependent transition models in HMM based word alignment for statistical machine translation. In: Proceedings of the second workshop on SMT (WMT), Association for Computational Linguistics. (2007)