International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:11, No:6, 2017

Sentiment Analysis: Comparative Analysis of Multilingual Sentiment and Opinion Classification Techniques

Sannikumar Patel, Brian Nolan, Markus Hofmann, Philip Owende, Kunjan Patel

***Abstract*—**Sentiment analysis and opinion mining have become emerging topics of research in recent years but most of the work is focused on data in the English language. A comprehensive research and analysis are essential which considers multiple languages, machine translation techniques, and different classifiers. This paper presents, a comparative analysis of different approaches for multilingual sentiment analysis. These approaches are divided into two parts: one using classification of text without language translation and second using the translation of testing data to a target language, such as English, before classification. The presented research and results are useful for understanding whether machine translation should be used for multilingual sentiment analysis or building language specific sentiment classification systems is a better approach. The effects of language translation techniques, features, and accuracy of various classifiers for multilingual sentiment analysis is also discussed in this study.

***Keywords*—**Cross-language analysis, machine learning, machine translation, sentiment analysis.

I. Introduction

SOCIAL networks, blogs, and reviews sites have become popular platforms for people to express their opinions. Social network and micro-blogging platforms like Facebook and Twitter have millions of users who generates millions or billions of lines of textual information per day. This data contains opinions, sentiments, attitudes and emotions toward entities and aspects such as products, organizations, individuals, places, social events, global problems etc. Many companies extract opinions for different purposes, e.g to know about product demand, influencing factors on the product, people choice *etc*. This process of extracting and classifying opinion on the different subject is known as sentiment analysis or opinion mining.

Broadly, there are two types of methods for sentiment analysis, machine learning based and lexical base. The machine learning method relies on two approaches, supervised learning and unsupervised learning. Supervised learning requires labeled data to train algorithms [1], while unsupervised learning, does not require labeled data [2]. The combination of labeled and unlabeled data yields semi-supervised learning [3]. Lexical based methods use a dictionary of words, where each word is associated with specific sentiment score [4] such as positive(+1), negative(-1) and neutral(0).

Sannikumar Patel, Brian Nolan, Markus Hofmann, Philip Owende and Kunjan Patel are with the Department of Computing, Institute of Technology Blanchardstown, Dublin, Ireland (e-mail: B00093164@student.itb.ie, brian.nolan@itb.ie, markus.hofmann@itb.ie, philip.owende@itb.ie, kunjan.patel@itb.ie).

The majority of supervised and unsupervised approaches for sentiment analysis uses words found as features. Other features such as syntactic features and frequency of words can also be used in the process of class labeling. The labels of a class can range from positive, negative and neutral to actual emotions like sad, happy or angry [5]. A classifier learns patterns from given features and then predicts sentiment of new instances based on their features. The data used for training directly affects the performance of the classifier. Therefore, data used to test classifiers is usually from the same domain as the data used for training. This characteristic is referred as domain specificity. Similar to domain specificity, the classifier can only be tested on the language on which it has been trained.

The possible ways of performing multilingual sentiment analysis are machine translation (MT) and building language specific classification system. For example, in MT classifier is trained using the dataset in the English language and for testing, the data instances are translated into English from another language. Whereas in language specific classification systems classifiers are trained and tested on the same language, this approach is called native classification here.

The purpose of this study is to determine the better approach to sentiment classification in multiple languages, as well as to find whether the performance of machine translation models does affect the classification. A series of experiments are performed in order to find possibilities. The two main approaches are implemented and compared: 1) native classification and, 2) machine translated classification. Different neural network-based machine translation models are used to translate movie review data from one language to another, and comparative experiments on supervised learning techniques are performed to classify the movie reviews as a positive or negative class. The two different neural network-based translation models are used for translation of data from English to Hindi and vice-versa.

The remainder of this paper is organized as follows. Section II presents literature review. Section III presents the methodology used. Section IV presents experiments performed. Section V presents results of machine translation and classification techniques. Finally, Section VI presents conclusion.

II. Literature Review

Sentiment analysis has received significant attention since it can be used to provide insights like opinion, choice, and habit of people on a product, places, person, *etc*. In sentiment classification, opinion and sentiment expressed in words such as *good, bad, worst, amazing, terrible, magnificent* are important, since they express polarity of text. Since the sentiment analysis is a text classification problem, it could be handled by machine learning algorithms like Naïve and support vector machine [1], [6].

Starting from being a document level classification task [1], sentiment analysis have been handled as sentence level [7] to aspect level by researchers [8], [9]. Furthermore, researchers have also explored text classification problems like sarcasm detection [10], [11], conditional and comparative sentence classification [12], [13], negation detection and classification [14], topic modeling and cross-domain sentiment classification [15], [9], [16]. Several sentiment classification techniques have also been used for detection of spam in reviews [17], product reviews [18], election results prediction [19], to score the aspect of product on e-commerce web sites [20], for event detection [21], and for graph based sentiment analysis on Twitter [9].

A multilingual sentiment analysis means to perform sentiment classification of opinionated text in multiple languages. Main motivation for multilingual sentiment classification is by building sentiment analysis systems for different languages [22]. However, most of the research is done in English. There is limited resource for other languages but another possible way of performing multilingual sentiment analysis is transfer learning or machine translation [23]. In [24] author exploited sentiment resources in English to perform classification of Chinese reviews. In [25] resources from the English language is adopted for sentiment analysis in Spanish. In [26] lexicon based methods are used for classification of Arabic tweets.

An attempt to use machine translation techniques in sentiment analysis have not been widely used due to the poor quality of translated text, but recent advancement in machine translation systems using the artificial neural network and deep learning has motivated such attempts [27].

III. Methodology

This section discusses methods and approaches used for multilingual sentiment classification. The two main approaches taken for comparative experiments and results are native classification and machine translated data classification.

(A) A native classification is a primary approach in which classifiers are trained and tested using the same dataset and the same language. For example, a classifier trained using Hindi dataset and tested using instances from the same dataset only. It uses original form of data, none of the training and testing instances are translated using machine translated techniques. Fig. 1 describes the native classification approach.

(B) The machine translated data classification approach employees different classifiers and machine translation techniques. The classifiers are trained using source language dataset and tested using target language dataset instances. The source language here refers to the language of dataset, using which classifier is trained and the target language refers to dataset language from which testing instances are translated. For example, in Hindi to English classification, Hindi is the source language and English is the target language because testing instances are translated to Hindi from English. Similarly, in English to Hindi translation, English is the source language and Hindi is the target language. Machine translated classification approach is described in Fig. 2, in which the training instances are selected from the source language dataset and the testing instances are selected from the target language dataset. After that, testing instances are translated using two different translation models. These models are discussed in subsection B.

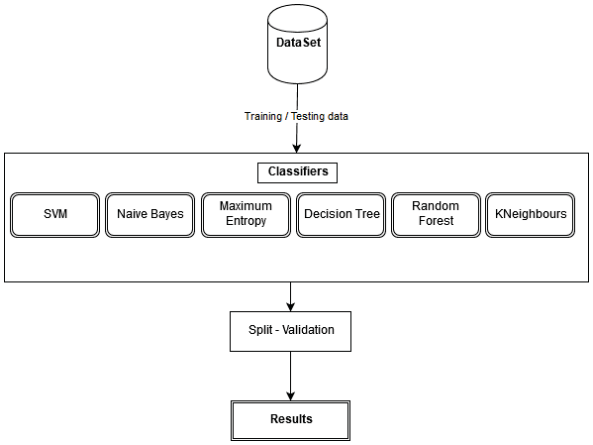

Fig. 1 Native sentiment classification using different classifiers and features extraction techniques such as n-gram

The following sub sections describe datasets, features, classifiers and machine translation models used.

A. Description of Datasets Used

Different kind of data combinations are used in this study.

(i) HindEnCorp: a parallel corpus of English and Hindi languages introduced in [28] is used. The HindEnCorp consists of 2,74,000 parallel sentences collected from various sources such as news articles, blogs, and Wikipedia. This dataset is used for training and testing the machine translation models.

(ii) English movie reviews dataset used in this study contains more than 10,000 positive and negative movie reviews [29]. Out of them all, random 5,000 reviews are selected. This selected reviews are used for various purposes, such as training the classifiers, testing the classifiers and machine translation.

(iii) Hindi movie reviews dataset used in classification and machine translation contains 5,000 reviews. Half of them reviews are extracted automatically from movie reviews website and labeled by detecting contained rating and emoticons. The remaining reviews are manually collected and labeled.

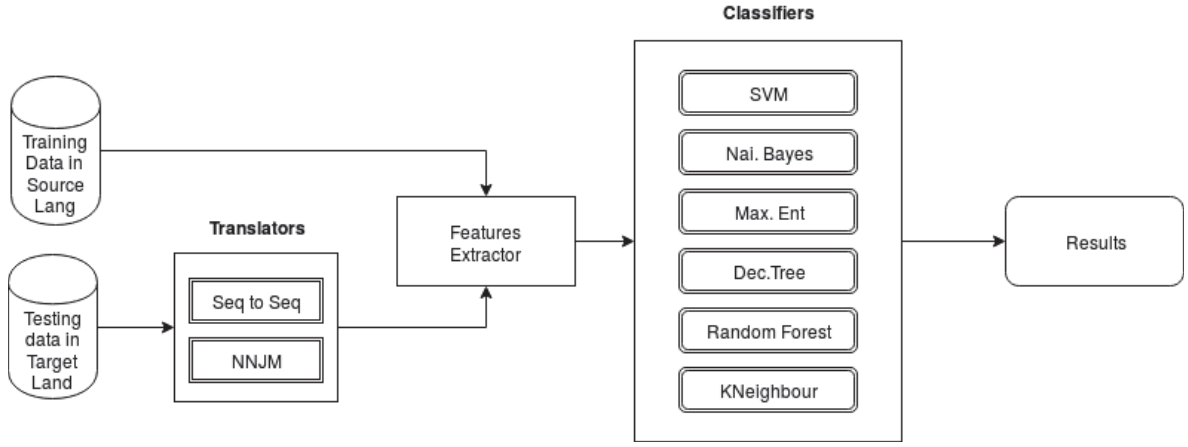

Fig. 2 Machine translated sentiment classification using different translator and classifiers

B. Machine Translation (MT)

A MT is a process of translating text from one language to another. Plenty of techniques exists, widely used for machine translation are rule-based, statistical, example-based, hybrid and neural machine translation. The statistical machine translation techniques such as phrase-based and word-based translation have been widely used for machine translation tasks. After the evolution in the neural network area, the statistical neural machine translation has got a wide amount of attention by the research community [27].

A neural machine translation models such as sequence-to-sequence [30] and Neural Network Joint Language Model (NNJM) is used for comparative analysis in this paper. Both models are trained for machine translation from Hindi to English (HN-EN) and English to Hindi (EN-HN) language.

A Bilingual Evaluation Understudy (BLEU) algorithm is used for evaluating the quality of text which is machine translated [31]. The central idea behind this algorithm is the closer a MT to a professional human translation, the better it is. The score in this algorithm is calculated by comparing an individual sentence with human translated reference. After that, individual scores are averaged over the whole corpus to get the final score.

1) Sequence to Sequence Neural Model: A basic sequence-to-sequence model, as introduced in [32] consists of two Recurrent Neural Networks (RNNs). This RNNs contains an *encoder* that encodes a sequence of symbols into a fixed-length vector representation, and the *decoder* that decodes the representation into another sequence of symbols. The encoder and decoder of the proposed model are jointly trained to maximize the conditional probability of a target sequence on given source sequence. This model does not create a pair of phrases to align its occurrence frequency, rather it is focused toward learning linguistic regularities.

The adopted sequence to sequence model was empirically evaluated on the task of translation from English to French. It is configured to use for Hindi to English and English to Hindi translation in this paper.

2) NNJM: The NNJM is a basic neural network architecture and lexicalized probability model to create a powerful machine translation decoding technique [33]. It works on the concept of word source windows using an n-gram model, which is also known as neural network language models (NNLM). It augments an n-gram target language model with m-word source window. This model consists of an input and output vector, where the input vector is a 14-word context vector where each word is mapped to a target word.

The adopted both models were empirically evaluated on different language translation tasks, but for this study both models are configured to used for Hindi to English and English to Hindi translation. The parameter changes applied on both models are separately discussed in experiments section.

C. Sentiment Classification

1) Feature Selection: The feature selection techniques like n-gram and tf-idf are used for extracting the feature from documents of text. N-gram is continuous sequence of *n* items from a given text. The n-gram creates a pair of words, the pair size of 1 is referred as a unigram, size 2 as bigram, size 3 as trigram and so on. For the given sentence *'The movie is good'* unigram would be like ('The', 'movie', 'is', 'good'), and bi-gram would ('The movie', 'movie is', 'is good'). Another technique used for feature selection is tf (Term frequency) and Tfidf (Tern frequency - inverse document frequency).

A tf-idf is statistic weighting factor calculation method used to determine the importance of the word in text documenst [34]. The number of time term/word T appears in document D is called its *term frequency*. Sometime the terms like ('This','the','is') are more common in text documents. In such cases term frequency incorrectly gives more importance to such words. To minimize this diverse weighting problem the *inverse document frequency* is used. tf-idf proportionally diminishes the weight of term which occurs very frequently and increases the weighting of the term which is occurring

rarely. That way it gives the dictionary of words associated weightings.

2) Classification: The purpose of this study is to find a better approach for multilingual sentiment analysis from native and machine translated classification, as well as to check whether machine translation can be employed to perform sentiment analysis for different languages. MT models described in Section III. B are used for translating the set of sentences. The sentences translated using both translation models are used as a testing dataset for classifiers.

A supervised machine learning algorithm such as SVM, Naïve bayes, maximum entropy, decision tree, random forest, and k-nearest neighbors (K-NN) are used for comparative analysis. All classifiers are trained with different features set such as Unigram + tf, Bigram + tf-idf *etc*. As discussed earlier in this section, native and machine translation approaches are implemented. Later, both approaches are compared with to find a best possible approach for multilingual sentiment classification. The same set of classifiers and features are used in both, native and machine translated classification approaches. The different combination of features used for classification are listed in Table I.

TABLE I
COMBINATION OF DIFFERENT FEATURES FOR CLASSIFICATION

Feature set No.	Feature set
F1	Unigram + tf
F2	Bigram + tf
F3	Unigram + tf-idf
F4	Bigram + tf-idf
F5	Unigram + Bigram + tf-idf
F6	Unigram + Trigram + tf-idf

TABLE II
BLEU SCORE OF MACHINE TRANSLATION TECHNIQUES FOR ENGLISH TO HINDI AND HINDI TO ENGLISH

Language Combinations	BLEU Scores	
	SeqtoSeq	NNJM
English - Hindi	29.0	27.0
Hindi - English	30.1	29.2

IV. EXPERIMENTS

In order to test and compare the performance of different classifiers on translated and non-translated dataset sentences. A series of experiments are performed on MT and sentiment classification techniques with 6 different feature sets.

A. Machine Translation

This subsection contains details related to configuration and parameters changes applied on translation models in order to achieve translation tasks on both, English and Hindi languages.

1) Sequence to Sequence Model: The process of training this model includes generating a vocabulary of 40,000 words for both Hindi and English language. Then after all the sentences in the dataset are separated into four buckets of varying length, (5,10), (10,15), (15,25) and (25,50). A bucket means, if the input in an English sentence has 3 words, and the corresponding output in a Hindi has 6 words, then they are put into the bucket (5,10) and padded to length 2 for encoder and length 4 for the decoder. Same as, an English sentence has 8 tokens and the corresponding Hindi sentence has 18 tokens, then they will not fit into the (10, 15) bucket, and so the (15, 25) bucket will be used. Further, sentences with size more than 50 are excluded from the training and testing dataset. Finally, the model is trained separately for Hindi to English and English to Hindi translation. The 200K sentences including development set and 2 seq to seq neural layer of 256 unit are used for the training process. Both Hn-En and En-Hn models are trained till perplexity of less than 5 is achieved with an initial learning rate of 0.5.

2) NNJM Model: The training approach in NNJM model is similar to neural network language model, except the parallel corpus is used. The same training procedure implemented as in [33] is adopted. First, the weight is randomly initialized in the range of [-0.05,0.05] with an initial learning rate of 10^-3 and mini batch size of 128^3. At every epoch, the likelihood of the validation set, which is defined as 20,000 mini batches is calculated. If the likelihood is not better than a previous epoch, the learning rate is multiplied by 0.5. The training is continued for 40 epochs. The training data contains 200K sentences including training and development sets, which are further converted to the vocabulary of 17,520 source words and 17,520 target words.

A total of 1,000 sentences are used for testing the all given models and the BLEU score is calculated for each translated output.

B. Sentiment Classification

The experiments related to sentiment classification are performed in four sets. The main two approaches, introduced in Section III, are further divided into four sets and experimented with each language separately. Such as native Hindi, native English, machine translated Hindi and machine translated English classification.

1) Native Hindi Classification: The native Hindi classification uses traditional approach of sentiment classification in which classifiers are trained and tested using the same dataset. It uses the original dataset in which dataset language is Hindi only. This set of experiments uses same Hindi movie reviews dataset introduced in Section III, randomly 4,000 sentences are selected for training and the remaining 1,000 sentences are used for testing.

2) Machine Translated Hindi Classification: The machine translated Hindi classification approach uses Hindi and English movie reviews dataset. The Hindi movie reviews are used for training the classifiers, and English movie reviews are used for testing the classifiers. A random selected sentences, from testing dataset are translated to Hindi separately using the Sequence to Sequence and NNJM model. Finally, the separate experiments are made on classifiers to evaluate the effect of both translation techniques. The amount of data used for whole process is, 4,000 Hindi reviews for training and 1,000 English reviews for testing.

TABLE III
ACCURACY OF NATIVE HINDI CLASSIFICATION WITH DIFFERENT FEATURES AND CLASSIFIERS

Features	SVM	Naïve Bayes	Maximum Entropy	Decision Tree	Random Forest	K-NN
F1	73%	75%	76%	70%	66%	59%
F2	75%	72%	76%	69%	72%	62%
F3	78%	75%	77%	64%	65%	71%
F4	74%	72%	74%	73%	71%	59%
F5	75%	72%	74%	70%	70%	75%
F6	75%	73%	72%	68%	72%	70%

TABLE IV
ACCURACY OF NATIVE ENGLISH CLASSIFICATION WITH DIFFERENT FEATURES AND CLASSIFIERS

Features	SVM	Naïve Bayes	Maximum Entropy	Decision Tree	Random Forest	K-NN
F1	81%	80%	80%	76%	78%	57%
F2	80%	80%	80%	75%	77%	57%
F3	80%	80%	79%	74%	79%	70%
F4	82%	81%	79%	75%	78%	70%
F5	79%	79%	79%	74%	78%	70%
F6	78%	78%	78%	74%	77%	70%

3) Native English Classification: This approach repeats same steps as the native Hindi classification. The classifiers are trained using random selected English movie reviews from the dataset and tested using random instances from the same dataset. This approach is not using any translated training and testing sentences. The randomly selected 4,000 and 1,000 reviews are used for training and testing purpose.

4) Machine Translated English Classification: The machine translated English classification approach uses English and Hindi movie reviews dataset. The English movie reviews are used for training the classifiers, and Hindi movie reviews are used for testing the classifiers. A randomly selected sentences, from testing dataset are translated to English separately using the Sequence to Sequence and NNJM model. Finally, the separate experiments are made on classifiers to evaluate the effect of both translation techniques. The amount of data used for the whole process is 4,000 English reviews for training and 1,000 Hindi reviews for testing.

The results of above four steps are compared to find, a better approach, a translation model, and a classifier. The detailed result is discussed in next section.

V. RESULTS AND DISCUSSIONS

Evaluation results are listed in Tables II-VI, which illustrates the performance of different machine translation and sentiment classification techniques in a combination of multiple feature sets. Following observations are drawn based on the presented results.

A. Machine Translation

The two different neural networks model implemented to determine the variation in performance of classification due to the variation in performance of translation. The following results are obtained after series of experiments, performed on translation models. A Sequence to Sequence model is found to be performing better for machine translation. It has dominated with highest BLEU score for both English to Hindi and Hindi to English translation. The results related to machine translation is given in Table II.

B. Sentiment Classification

This section covers results achieved after a series of experiments performed on sentiment classification approaches. As described in Section IV, results are divided into two subsections, given below.

1) Hindi Classification: Among the classifiers used for native Hindi classification, the highest score was 78% using features set F3. In contrast, the classification result of machine translated data using NNJM model is 67% highest, which is even 2% lower compared to the highest score of the sequence to sequence translated data classification. The most useful classifier and features found in both cases are SVM and Naïve bayes, with features F3, F4, and F5. The major difference can not be only seen in the case of SVM and Naïve bayes, but other classifiers such as maximum entropy, decision tree, and the random forest have also been performed better with native approach compared to translation. It can be also seen that lower translation score of NNJM model in comparison of the sequence to sequence model has also affected the results of classification. The classification results of NNJM translated data was almost lower compared to sequence to sequence model in all the cases. The overall result shows the native approach of sentiment classification for multilingual sentiment analysis is more prominent compared to machine translation, in the case of Hindi language. The additional results are given in Tables III and V.

2) English Classification: The native English and machine translated English classification also repeats the same stories. The highest score was 82% in case of native English classification. While the results achieved through the sequence to sequence and NNJM model were 72% and 68% using Naïve bayes and SVM. This difference was not only seen in the case of Naïve bayes and SVM, again all native classifiers have performed better. Here as well, the lower results of NNJM has also affected the classification results, which can be clearly seen from different between classification results of NNJM and sequence to sequence models. The additional results are given in Tables IV and VI. At last, It could be said that native approach of classification is also better in the case of English

TABLE V
ACCURACY OF HINDI CLASSIFICATION PERFORMED ON MACHINE TRANSLATED DATA INSTANCES, USING SEQ TO SEQ AND NNJM MODELS

Features	SVM		Naïve Bayes		Maximum Entropy		Decision Tree		Random Forest		K-NN	
	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM
F1	63%	62%	67%	61%	65%	63%	52%	56%	55%	57%	54%	54%
F2	61%	63%	68%	65%	67%	66%	53%	50%	60%	55%	52%	55%
F3	65%	62%	66%	61%	67%	62%	51%	52%	56%	57%	55%	58%
F4	69%	65%	67%	65%	67%	65%	54%	53%	56%	59%	51%	50%
F5	68%	67%	67%	65%	68%	66%	53%	53%	58%	60%	59%	61%
F6	68%	64%	67%	67%	67%	65%	56%	49%	57%	53%	60%	61%

TABLE VI
ACCURACY OF ENGLISH CLASSIFICATION PERFORMED ON MACHINE TRANSLATED DATA INSTANCES, USING SEQ TO SEQ AND NNJM MODELS

Features	SVM		Naïve Bayes		Maximum Entropy		Decision Tree		Random Forest		K-NN	
	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM	SeqtoSeq	NNJM
F1	69%	65%	71%	68%	71%	67%	58%	56%	64%	58%	52%	56%
F2	70%	66%	72%	67%	71%	66%	56%	58%	63%	62%	57%	55%
F3	70%	68%	71%	68%	70%	66%	57%	53%	65%	56 %	63%	60%
F4	70%	68%	72%	67%	69%	65%	57%	53%	60%	63%	60%	53%
F5	71%	67%	72%	66%	69%	65%	57%	54%	62%	55%	65%	60%
F6	70%	66%	70%	66%	68%	65%	54%	54%	60%	57%	64%	59%

language.

Overall from the results of all approaches, the conclusion can be made that, the native approach of classification is more prominent compared to machine translation and classification. The another conclusion can also be made that, the performance of translator do also affect the performance of classification.

VI. CONCLUSION

The two approaches are implemented and compared to find one of the best possible approaches for multilingual sentiment analysis. First, different six classifiers are trained and tested in the native language. Secondly, the testing data sentences are translated using different neural machine translation models and then classified. At the end, all experiments results are compared with each other. The compared results suggest, building language specific sentiment classification systems is better than language translation.

The separate experiments are also performed on machine translation techniques to determine the possible effects of translators performance on sentiment classification. The results show the difference in performance of machine translation techniques also affects the performance of classification systems.

Despite having some limitations like availability of additional resources, such as POS taggers and Stemmers, native classification is found a prominent option for multilingual sentiment analysis. This characteristic shows, future work should be concentrated on developing such resources for individual languages unless machine translation is not gaining state of the art performance.

REFERENCES

[1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP*, pages 79–86, 2002.

[2] Peter D Turney. Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, (July):417–424, 2002.

[3] Andrew B Xiaojin. Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, pages 1–130, 2009.

[4] Xiaowen Ding, Xiaowen Ding, Bing Liu, Bing Liu, Philip S. Yu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. *Proceedings of the international conference on Web search and web data mining - WSDM*, page 231, 2008.

[5] Kevin Hsin Yih Lin, Changhua Yang, and Hsin Hsi Chen. Emotion classification of online news articles from the reader's perspective. *Proceedings - 2008 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2008*, pages 220–226, 2008.

[6] Jalel Akaichi. Social networks' Facebook' statutes updates mining for sentiment classification. *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom*, pages 886–891, 2013.

[7] Hong Yu and Vasileios Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136, 2003.

[8] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent Twitter Sentiment Classification. *Computational Linguistics*, pages 151–160, 2011.

[9] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1031–1040, 2011.

[10] Mondher Bouazizi, Tomoaki Otsuki Ohtsuki, and Senior Member. A Pattern-Based Approach for Sarcasm Detection on Twitter. 2016.

[11] Mondher Bouazizi and Tomoaki Ohtsuki. Sarcasm detection in twitter: all your products are incredibly amazing - are they really? *2015 IEEE Global Communications Conference, GLOBECOM*, pages 1–6, 2016.

[12] Ramanathan Narayanan, Bing Liu, and Alok Choudhary. Sentiment analysis of conditional sentences. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 EMNLP 09*, (August):180, 2009.

[13] Nitin Jindal and Bing Liu. Identifying comparative sentences in text documents. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '06*, page 244, 2006.

[14] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, pages 267–307, 2011.

[15] Zeynep Zengin Alp and Sule Gunduz Oduducu. Extracting Topical Information of Tweets Using Hashtags. *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 644–648, 2015.

[16] Brian Heredia, Taghi M. Khoshgoftaar, Joseph Prusa, and Michael Crawford. Cross-Domain Sentiment Analysis: An Empirical

Investigation. *IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 160–165, 2016.

[17] Qingxi Peng and Ming Zhong. Detecting Spam Review through Sentiment Analysis. *Journal of Software*, pages 2065–2072, 2014.

[18] Hang Cui, Vibhu Mittal, and Mayur Datar. Comparative Experiments on Sentiment Classi cation for Online Product Reviews. *Entropy*, pages 1265–1270, 2003.

[19] Mochamad Ibrahim, Omar Abdillah, Alfan F. Wicaksono, and Mirna Adriani. Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW*, pages 1348–1353, 2016.

[20] S. A. A. A. Alrababah, K. H. Gan, and T. P. Tan. Product aspect ranking using sentiment analysis and topsis. *Third International Conference on Information Retrieval and Knowledge Management (CAMP)*, pages 13–19, Aug 2016.

[21] Mário Cordeiro. Twitter event detection: combining wavelet analysis and topic inference summarization. *Proceedings of Doctoral Symposium on Informatics Engineering*, 2012.

[22] Shravan Vishwanathan. Sentiment Analysis of French Movie Reviews. *Proceedings of 3rd IRF International Conference*, pages 80–82, 2014.

[23] Alexandra Balahur and Marco Turchi. Multilingual sentiment analysis using machine translation. *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pages 5260, Jeju, Republic of Korea.*, (July):52–60, 2012.

[24] Xiaojun Wan. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. *Proceedings of the Conference on Empirical Methods*, pages 553–561, 2008.

[25] Julian Brooke, Milan Tofiloski, and Maite Taboada. Cross-linguistic sentiment analysis: From english to spanish. *International Conference RANLP*, pages 50–54, 2009.

[26] Mahmoud Al-ayyoub, S. B. Essa, and Izzat Alsmadi. Lexicon-based sentiment analysis of Arabic tweets. *International Journal of Social Network Mining*, 2(July 2016):101–114, 2015.

[27] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *ArXiv*, pages 1–23, 2016.

[28] Ondej Bojar, Vojtěch Diatka, Pavel Rychlý, Pavel Straák, Vít Suchomel, Aleš Tamchyna, and Daniel Zeman. Corpus for Machine Translation. pages 3550–3555, 2002.

[29] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*, 2005.

[30] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to Sequence Learning with Neural Networks. *Nips*, pages 3104–3112, 2014.

[31] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. pages 311–318, 2002.

[32] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, 2014.

[33] Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul. Fast and Robust Neural Network Joint Models for Statistical Machine Translation. *Acl*, pages 1370–1380, 2014.

[34] Delta TFIDF: An Improved Feature Space for Sentiment Analysis. *Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM*, (May):490–497, 2008.