2016 12th International Conference on Signal-Image Technology & Internet-Based Systems

# SentiWordNet for New Language: Automatic Translation Approach

Alaettin Ucan, Behzad Naderalvojoud, Ebru Akcapinar Sezer and Hayri Sever
*Computer Engineering Department*
*Hacettepe University*
*Ankara, Turkey*
{*aucan, n.behzad, ebru, sever*}@*hacettepe.edu.tr*

***Abstract*—This paper proposes an automatic translation approach to create a sentiment lexicon for a new language from available English resources. In this approach, an automatic mapping is generated from a sense-level resource to a word-level by applying a triple unification process. This process produces a single polarity score for each term by incorporating all sense polarities. The major idea is to deal with the sense ambiguity during the lexicon transfer and provide a general sentiment lexicon for languages like Turkish which do not have a freely available machine-readable dictionary. On the other hand, the translation quality is critical in the lexicon transfer due to the ambiguity problem. Thus, this paper also proposes a multiple bilingual translation approach to find the most appropriate equivalents for the source language terms. In this approach, three *parallel*, *series* and *hybrid* algorithms are used to integrate the translation results. Finally, three lexicons are achieved for the target language with different sizes. The performance of three lexicons is evaluated in the lexicon-based sentiment classification task and compared with the results achieved by the supervised approach. According to experimental results, the proposed approach can produce reliable sentiment lexicons for the target language.**

***Keywords*-Sentiment lexicon; Translation approach; Machine learning; Sentiment analysis**

## I. Introduction

Sentiment Analysis, also known as Opinion Mining, is to determine the contextual polarity of a document that indicates the author's judgement, evaluation or emotional state. In this context, the sentiment orientation of a text is taken into account rather than its topic. This issue is associated with the other text mining applications like irony detection, subjectivity detection, opinion extraction and so on. In general, the sentiment of a text which conveys the author's opinion, is induced from its constituent sentences, and consequently from the subjective terms used in the sentences. Therefore, the general polarity of a text depends on the polarity of terms, however, the effect of each term is not the same in the determination of the text polarity. It means that the general polarity of a text is likely induced from strongly subjective terms.

Existing approaches to sentiment analysis can be grouped into two major categories: (1) keywords and lexicon-based methods (2) statistical and learning-based methods. The major problem contained in the sentiment analysis via machine learning techniques arises from the classical subject-based text classification problem which requires a particular labeled data associated with the subject matter. Although for many non-English languages like Turkish much research has been carried out based on machine learning techniques [1], [2], [3], [4], sentiment analysis via lexicon-based methods has received less attention due to the lack of sentiment lexicon resources.

To generate a sentiment lexicon for a new language that expresses the sentiment orientation of terms, we need to access to all terms used in the language and manually tag all of them. This work is very time-consuming and approximately impossible in a plausible time, even if all the language terms are available. In this condition, the only available solution is to generate such resources automatically. However, for most languages, there is no comprehensive dictionary like WordNet to express the language terms and senses (synsets) as well as their meaning relationships. This is important because the term polarity is directly associated with the term meanings rather than its lexical form. In such cases, the approaches of creating a sentiment dictionary, which indicates the polarity of terms, is more limited than the methods used in the English language. For instance, SentiWordNet is a sense-level lexicon that has been automatically generated to determine the Positive-Negative (PN) polarity of WordNet synsets [5]. In this resource, three numerical scores *Obj(s)*, *Pos(s)* and *Neg(s)* are assigned to synsets instead of terms such that they indicate how much objective, positive and negative a synset is respectively. As the summation of three scores equals to 1, three fundamental subtasks of opinion mining are implicitly realized upon term synsets, i.e. (1) determining the subjectivity, (2) determining the orientation (PN-polarity) and (3) determining the strength of orientation.

In this paper, we take into account an automatic translation approach along with a unification technique to generate a sentiment lexicon dictionary for a new language from SentiWordNet. Our hypothesis is that sentimental expressions may be the same in different languages. This means that the emotion conveyed by an English text is almost the same while it is translated into another language. Despite the availability of all synset polarities in SentiWordNet, they cannot be used directly in sentiment analysis systems due to the ambiguity problem. According to this fact, we aim to propose an automatic mapping approach for transferring a

978-1-5090-5698-9/16 $31.00 © 2016 IEEE
DOI 10.1109/SITIS.2016.57

sense-level resource to a word-level in the target language. Because of the low number of subjective terms that their synsets include both positive and negative polarities (less than 6%), the ambiguity problem may have less influence on the general performance of the lexicon transfer approach. Although the proposed approach is applied to Turkish, it is applicable to any other language, as we do not use any linguistic knowledge of the target language in the methods described in the paper.

This study generates a sentiment lexicon for the Turkish language by using multiple bilingual translation dictionaries. To integrate the translation results gained by each dictionary, three algorithms are proposed. While the first algorithm combines the translation results in a series manner, they are compared in a parallel form in the second algorithm. The third algorithm is performed by incorporating the two previous algorithms and finds suitable Turkish terms for those the two former algorithms have not achieved any equivalent for them. This algorithm indeed uses synonym relationships contained in the WordNet. By applying a triple unification process, a sense to word mapping is realized on the SentiWordNet terms to compute a single subjective score (called as *SubScore*). As a result of the proposed approach, three sentiment lexicons called as TSDs, TSDp, TSDh are produced for the Turkish language. The performance of translation methods which have been reflected on the corresponding TSD dictionaries is evaluated in two Turkish sentiment benchmarks. These benchmarks have been collected from Turkish hotel and movie reviews. Actually, this study attempts to increase the experience of using lexical resources in sentiment analysis and makes a comparison with statistical and learning-based methods.

## II. Related Work

In order to extract opinions from English texts, many studies have attempted to automatically determine the PN-polarity of subjective terms as well as determining the subjectivity of terms. Thus, the primary research has been carried out over the three baseline tasks: (1) determination of text subjectivity, which specifies whether a given text presents an opinion about its subject matter. In many studies, this task is considered as a binary text categorization under "objective" and "subjective" categories [6], [7] ; (2) determination of the text orientation, which expresses the PN-polarity of a subjective text [6], [8] ; (3) determination of the strength of text orientation, which expresses the degree of PN-polarity associated with the text matter [5], [9].

In an earlier study, [10] has focused on the properties of terms within the sentiment analysis. It demonstrated that the conjunctions (such as *and, or, but, either-or, neither-nor*) between adjectives provide indirect information about their semantic polarities. In another approach, [8] has employed pointwise mutual information (PMI) to compute the semantic similarity of any term with two small sets of Positive-Negative subjective terms (*seed set*). This approach has been also employed in [11] to determine the term subjectivity instead of the term orientation. In this approach, a bootstrapping algorithm is realized by PMI metric to select positive and negative terms based on their similarities to strongly positive and negative sets of subjective terms. The bootstrapping technique has been also used in [12] to learn many subjective patterns for determining the term subjectivity.

For the English language, [7] has proposed a gloss classification based method to determine whether a given term has a positive, negative or objective connotation. Their hypothesis is that terms with similar orientation have similar glosses and terms without orientation have non-oriented glosses. In fact, they took into account the both problems of determining subjectivity and determining term's orientation, simultaneously. Since different senses of the same term may have different sentiment polarities, they have used the WordNet synsets instead of terms in their approach. In languages which do not have access to any machine-readable dictionary, creating a sentiment dictionary will become more difficult than the other rich languages. In such cases, most of the previous studies have tried to translate the English lexical resources to any other language [13], [14], [15].

In case that non-commercial WordNet is not available, a word-level translation process has been used to transfer any lexical resource to the target language. In [16], a word-level lexical transfer technique has been applied to each entry of SentiWordNet and Subjectivity Word List of Opinion-Finder [17] to generate a sentiment lexicon dictionary for Bengali language. They have also applied a control procedure for inflected words because inflected words may be stemmed through the translation process. The bilingual translation approach has been also used in [13] to create corpora and lexical resources for a new language (e.g. Romanian). [15] has proposed several computational techniques to develop sentiment lexicons for three Indian languages, Bengali, Hindi and Telugu from the English resource SentiWordNet. In their approach, a word-level synset transfer technique has been applied to each English synset, and also words having ambiguity potential were eliminated during the translation. For German language, two sentiment lexicons with different sizes have been proposed in [18] using a semi-automatic translation approach from the Subjectivity Word List and SentiSpin [19]. This work indicates that the size of dictionaries does not take advantage in the classification accuracy in spite of better coverage of the polarity-based features.

In a recent work [20], a cross-lingual learning approach has been proposed to learn sentiment lexicons for a new language from available resources in English. This approach indeed uses two *intra* and *inter* language subgraphs to model (1) the semantic relations among the words in the source language, and (2) the words relations between the source and target languages. In this approach, a Bilingual word graph

Label Propagation (BLP) is also introduced to make a better coverage of the sentiment lexicon.

For the Turkish language, a sentiment lexicon has been recently generated from Turkish WordNet using a semi-automatic method [21]. Turkish WordNet [22] contains only 15000 synset along with some linguistic features such as *Turkish gloss, equivalent English synset, synonyms, POS tag* and so on. However, some linguistic features are not available for all synsets, for instance many synsets do not have any Turkish gloss or synonyms. Hence, the authors have tried to revise Turkish WordNet before using it in their methodology. In their approach, the sentiment polarities (*positive, negative and objective*) are manually assigned to each synset of Turkish WordNet and then polarity scores are calculated based on the supervised approach. This lexicon is not accessible for research purposes.

## III. Sentiment Lexicon Translation

As mentioned earlier, many sentiment lexicons have been generated for a new language based on manually or semi-automatically translation approach from annotated data available in English resources. In sentiment analysis, these lexicons either are used directly in the subjectivity detection process [23] or are used to extract features in the supervised approach [24], [21]. Regarding the success of the translation approach and development of bilingual translation applications, we automatically create a sentiment lexicon for Turkish language by translating the annotated lexicon in the English language. We use SentiWordNet (SWN) as a source lexicon because it has been generated by spending a big effort in many years[1]. However, SWN is a sense-level lexicon and polarity scores are assigned to the synsets instead of words. Therefore, this study focuses on making an automatic mapping between the SWN synsets and Turkish words or phrases. To this end, we first need to unify any term's synsets of SWN for mapping them to the Turkish terms. Actually, we want to automatically generate a word-level sentiment lexicon for any new language like Turkish from a sense-level lexicon in the source language.

### A. Synset Unification

The SentiWordNet consists of 6 different features as follows (1) POS tag (2) Synset ID (3) PosScore (4) NegScore and (5) SynsetTerms (6) Gloss. The objectivity score is calculated from Positive and Negative scores separately. In total, this lexicon contains 117,659 synsets which indicate individual English concepts. After splitting all synset terms, 206,941 terms are obtained by their different meanings. As each term can be placed in several synsets (due to the ambiguity), more than one PN-scores may be present for each term. In order to unify these scores, we first unify the Positive and Negative scores of each synset by subtracting the negative one from the positive. The achieved new score (named *SubScore*) indicates the *subjectivity level* of each synset by a single value. This technique can implicitly resolve the ambiguity problem while a synset possesses both non-zero Positive and Negative scores at the same time, e.g. the synset "unaccommodating#1" has identical Positive and Negative scores of 0.25. Thus, this synset is considered as an objective synset in spite of having the 50% subjectivity in SWN. In the second step of unification, a weighted average is taken by Eq. 1 over all the polarity scores (*SubScores*) of the synsets that the given term encompasses them.

$$S_t = \frac{\sum_{i=1}^{n} \frac{1}{i} SubScore_i}{\sum_{i=1}^{n} \frac{1}{i}} \tag{1}$$

In Eq. 1, *n* denotes the number of all possible synsets (senses) of the given term *t* such that they have been sorted based on their usage probabilities in the language. This manner causes that more popular synsets make more impact on the final polarity score. In this step, $S_t$ is considered as a coverage of all synset polarities and yields a single score associated with the given term. Actually, it provides a sense to word mapping and implicitly eliminates the word-level ambiguity.

As a term may appear in several POS tags, more than one score may achieve by the previous step since for each POS tag a separate score has been calculated. We take the arithmetic average of the scores of the different tags because we suppose that all tags have the same usage probabilities. In the end of the unification process, a Subjective List of 38,188 unique terms (with non-zero *SubScores*) are obtained from SWN including 17,770 positive and 20,418 negative. In the next subsection, three methods are proposed to translate this Subjective List to Turkish.

### B. Word-Level Translation Approach

The Turkish equivalent of any entry of Subjective List can be found using internet translation services. We develop a framework for automatic translation by using 3 online English-Turkish dictionaries, *Tureng*[2] (Dict1), *Zargan*[3] (Dict2) and *Bab.la*[4] (Dict3). We also use Google Translate API (GTA) along with the other dictionaries because it achieves better results in the translation of multi-word expressions. Moreover, GTA returns more appropriate equivalents to the target language since it uses statistical analysis rather than the traditional rule-based analysis [25]. Overall, the Subjective List achieved by the unification process is automatically translated into the Turkish by using a multiple bilingual translation approach in which the translation results are integrated by three different algorithms.

[1] http://sentiwordnet.isti.cnr.it/

[2] http://tureng.com/

[3] http://zargan.com/

[4] http://tr.bab.la/

*1) Serial-Connected Translation Approach :* For each entry of the Subjective List, we achieve the translation results from 4 dictionaries in the following order: GTA, Dict1, Dict2 and Dict3. In this method, GTA is considered as the most reliable resource according to our empirical experiment which has been carried out over 100 English terms including single and multi-word expressions randomly selected from SWN. If GTA does not produce any translation result, tureng dictionary, Dict2, is considered as the second alternative. Tureng dictionary, abbreviated from the first syllables of the words "Turkish and English", has been developed by Ozgur Suyel, a Turkish-English translator, with more than 10 years experience in translation. This dictionary by having more than 2,000,000 English and Turkish words and phrases locates in the second rank in the mentioned experiment. Similarly, the two other dictionaries are considered in the translation process when the former dictionary does not yield any translation result. This translation method has encompassed 94% of the Subjective List and translated 36,077 terms to Turkish. However, 8,826 terms have been repeated in the Turkish Subjective List since the Turkish equivalent of several English terms is the same. For instance, the Turkish equivalent of 18 English terms like "appalling", "awful", "dreaded", "horrible", "terrible" and so on, is "korkunç" in the translated list. In order to assign an appropriate *SubScore* for such Turkish terms, two methods are further proposed in section III-C. It is worth noticing that the final subjective polarity score calculated is called *SubScore* in this paper.

*2) Parallel-Connected Translation Approach:* Despite the high translation ratio of Subjective List in the serial approach, the reliability of translation cannot be measured. In order to make a reliable automatic translation, the parallel approach is proposed to augment the quality of translation. In the parallel approach, translation results of all dictionaries are considered simultaneously. The Turkish equivalent of a given term is selected from those which are gained by more than one dictionary. The hypothesis is that a Turkish term gained by the majority of dictionaries is more appropriate equivalent for the given English term. Although the number of translations reduced to 11,369 in the Turkish Subjective List, we believe that the translation results are more reliable. Like the previous approach, there exist 1,957 repeated terms with several *SubScores* in the translated list achieved by the parallel method.

*3) Hybrid Translation Approach:* In the parallel approach, although the translation results may be better than the serial one, a low number of terms are translated into the Turkish (almost 30%). To leverage the both quality and density in translation, the third approach incorporates the two previous methods to augment the reliability of translation as well as scalability. It also proposes a method to find the Turkish equivalent of unknown terms which have not been found any Turkish equivalent via translation algorithms. In this approach, we first apply the parallel translation algorithm to all terms; in cases that this algorithm does not produce any result, the serial algorithm is used. In the end of this process, total 36,077 terms of the Subjective List were translated to Turkish. However, there are 2,111 English terms which have not been attained any Turkish equivalent from the serial and parallel approaches. To find their equivalents, we take into account their WordNet synonyms and translate their synonym terms via the third translation approach. Finally, we could catch 1,186 (56%) of the unknown terms by this approach and totally found Turkish equivalents of 37,263 (97.5%) terms of the Subjective List. However, this method increases the repeated terms to 8,936 in the Turkish Subject List.

Table I
THE TERM STATISTICS OF THE SENTIMENT LEXICONS

| | SWN | TSDs | TSDp | TSDh |
|---|---|---|---|---|
| Total terms | 38,188 | 36,077 | 11,369 | 36,077 |
| Repeated after translation | - | 8,826 | 1,957 | 8,936 |
| Unique terms | 38,188 | 27,251 | 9,412 | 27,141 |

### C. Subjective Score Alignment for Repeated Terms

In unification process, the subjective scores of English terms (*SubScores*) calculated were assigned to the Turkish terms in the translated list. However, some terms are repeated in the Turkish lexicons since the Turkish equivalent of a set of English terms is the same. Therefore, these Turkish terms will have more than one *SubScore* in the translated list. In order to assign appropriate scores to these terms, two methods are used. In the first method, it is assumed that all English terms have the same chance in the sentiment alignment process and we take the arithmetic average of their *SubScores*. On the other hand, the second method selects the most subjective case (it can be positive or negative) as the *SubScore* of the Turkish equivalent since it supposes that all English terms may convey almost similar sentiments. Thus, the Turkish lexicon achieved by this method includes unique terms that carry the maximum sentiment polarities.

By applying the proposed methods to the translation process, three Turkish sentiment lexicons are generated by two sets of subjective scores. We call the lexicon achieved by the serial approach as TSDs (abbreviated from Turkish Sentiment Dictionary via serial approach) and two other dictionaries as TSDp and TSDh obtained from the parallel and hybrid approaches, respectively. The statistics of the obtained lexicons are presented in Table I. Furthermore, Figure 1 illustrates the distribution of positive and negative terms in the English and Turkish lexicons. From this figure, although the different number of terms exist in each lexicon, the same negative-positive ratios (25:21) are observed among the terms of all the lexicons. This may show that the score alignment process likely allocates appropriate polarity scores to Turkish terms.

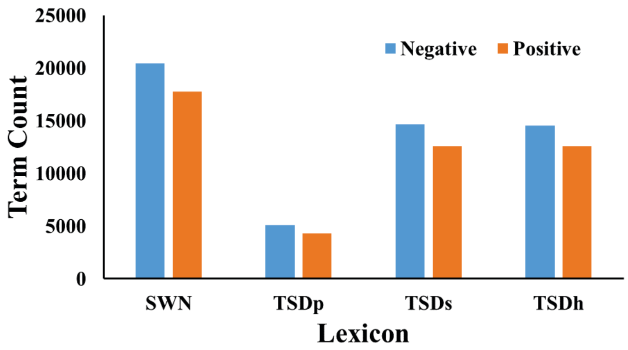


Figure 1. The distribution of positive and negative terms in SWD and Turkish Sentiment Dictionaries

## IV. Turkish Sentiment Data

In most studies, the sentiment labeled data is often extracted from the movie and hotel reviews [26], [27], [28], [29], [18]. These such data are either used in the supervised approach for training or used in the lexicon-based methods for sentiment evaluation. Because of the shortage of the freely available labeled data for Turkish, this study proposes a sentiment labeled data from a massive movie and hotel reviews. To this end, we have selected two most popular movie and hotel recommendation websites from those which attain a high rate in the Alexa[5] website. We selected "beyazperde.com" and "otelpuan.com" for movie and hotel reviews, respectively. By using Html Agility Pack[6], which is considered as an agile HTML parser, the reviews of 5,660 movies were investigated. The all 220,000 extracted reviews had been already rated by own authors using stars 1 to 5. As most of the reviews were positive, we selected the positive reviews as much as the negative ones to provide a balanced situation. The total of negative reviews rated by 1 or 2 stars were 26,700, thus, we randomly selected 26,700 out of 130,210 positive reviews rated by 4 or 5 stars. Overall, 53,400 movie reviews by the average length of 33 words were selected.

The similar manner was used to hotel reviews with the difference that the hotel reviews had been rated by the numbers between 0 and 100 instead of stars. From 18,478 reviews extracted from 550 hotels, a balanced set of positive and negative reviews was selected. As there were only 5,802 negative hotel reviews using 0 to 40 rating, we selected 5800 out of 6499 positive reviews rated from 80 to 100. The average length of all 11,600 selected positive and negative hotel reviews were 74 which is more than two times of the movie reviews. These two datasets[7] will be freely available for research purposes in the near future. After removing Turkish StopWords and applying normalization and lemmatization processes by using Zemberek [30], Turkish NLP tool, both datasets are represented in the vector space model.

[5]Alexa provides traffic data, global rankings and other information on 30 million websites (http://www.alexa.com)

[6]https://htmlagilitypack.codeplex.com/

## V. Lexicon Evaluation

Manually evaluating the proposed three Turkish lexicons is impossible, because we need a full manual tagging of all terms as well as assigning subjectivity score to them. Therefore, the effectiveness of each lexicon is evaluated on the subjectivity classification task. To achieve this goal, a simple lexicon-based method is used to classify the movie and hotel reviews based on the occurrence of Turkish subjective terms. In order to observe the effects of the average *SubScore* calculated by the SWN scores (denoted by *AvgSubScore*), we classify each review based on the sum of all *SubScores* of terms observed in the given text as Eq. 2:

$$pol(s) = \begin{cases} Pos, & \text{if } \sum_{i=1}^{n} SubScore_i > 0 \\ Neg, & \text{if } \sum_{i=1}^{n} SubScore_i < 0 \end{cases} \quad (2)$$

where *n* is the number of all terms (not unique) observed in the given review. In fact, the frequency of terms is considered in the Eq. 2.

To judge the performance of classification in each category, Precision ($P$), Recall ($R$) and F-measure ($F$) values are used as well as Accuracy ($Acc$) via following formulas: (1) $P = TP/(TP + FP)$, (2) $R = TP/(TP + FN)$, (3) $F = (2PR)/(P + R)$ and (4) $Acc = (TP + TN)/(TP + TN + FP + FN)$ where TP, TN, FP and FN denote the true positive, true negative, false positive and false negative, respectively.

### A. SubScore Alignment Evaluation

All three Turkish lexicons consist of two different *SubScores* as "average" and "maximum" scores because of the repeated terms in the Turkish Subjective List. To observe the effectiveness of each one, all reviews of the two benchmarks are classified by using the both scores. Table II presents the classification accuracy over all the movie and hotel reviews.

According to the results shown in Table II, the average score yields a better accuracy in all lexicons over the both benchmarks. As the number of movie reviews are increasingly higher than the hotel reviews (almost 4.6 times), the better results achieve by the hotel reviews. As can be seen, in two different scales the average score works better than the maximum one. Therefore, the average *SubScore* is considered as a more reliable sentiment score in all lexicons and used in the subsequent experiments as a baseline score. This demonstrates that while a term is considered as a Turkish equivalent of multiple English terms, those English terms more likely convey different senses. This issue refers

[7]http://humir.cs.hacettepe.edu.tr/

Table II
THE CLASSIFICATION ACCURACY ACHIEVED BY TURKISH LEXICONS

| Benchmark | Lexicon | AvgSubScore | MaxSubScore |
|---|---|---|---|
| Hotel Reviews | TSDp | 80.68% | 70.73% |
| | TSDs | 76.33% | 71.23% |
| | TSDh | 76.13% | 71.35% |
| Movie Reviews | TSDp | 70.35% | 67.51% |
| | TSDs | 66.38% | 63.90% |
| | TSDh | 67.49% | 63.24% |

to the existence of many homonyms in the Turkish language in which considering the maximum sentiment may lead to an unexpected score.

### B. Turkish Lexicon Evaluation

As shown in Table II, the parallel-based translation approach (TSDp) achieves more successful accuracy than the two others in terms of average and maximum *SubScores*. To better analyze the performance of each lexicon, Table III presents the achieved precision, recall and F-measure values for each Positive and Negative sentiment classes. According to these results, TSDp lexicon outperforms two other lexicons by the average F-measure values of 0.800 and 0.700 on hotel and movie reviews, respectively. The major point is that, despite achieving higher precision values by all lexicons in the Positive class, recall values are low. In contrast to Positive class, recall values are detected higher than precision ones in the Negative class. This perhaps shows the better quality of Positive *SubScores* than the Negative ones in all three Turkish lexicons. As the majority of terms are negative in all lexicons, it may cause ambiguity through the sentiment classification by which the low precision values are achieved in the Negative class. Nevertheless, it can make better coverage of all the reviews and leads to attain better recall values. This issue can be obviously seen in the precision and recall values of the TSDs and TSDh lexicons in the negative class since they have the most number of negative terms. Generally, TSDs and TSDh which cover the most number of subjective terms, probably suffer from the ambiguity problem in the negative class. On the other hand, the translation quality of the parallel-based approach can be observed among the others, though it cannot overwhelm the whole Subjective List of SWN.

Moreover, the achieved terms by the parallel approach can be considered as the most significant subjective terms since they are resulted from the contribution of more than one dictionary. It can be deemed that a feature selection process is implicitly applied to all subjective terms during the translation phase and more significant terms are selected.

To make a general evaluation and see the effectiveness of the proposed Turkish lexicons in the sentiment classification task, we perform a supervised learning approach to classify the hotel and movie reviews. Thus, we employ the *chi-square* ($X^2$) feature evaluation metric [31] to select an efficient number of terms as a feature set. According to [1] which has achieved the best sentiment classification accuracy by 375 features, we selected the same number of features by chi-square metric and represented all reviews by the selected features. For each benchmark, the half of reviews were used as a training data and the remaining ones were used for testing. As the support vector machines (SVMs) with linear kernel performs well in text classification, it was used to learn the sentiment classification model. The libSVM implemented in WEKA[8] was used in the experiment. Table III makes a comparison between the results obtained by the supervised approach and the results of the lexicon-based method achieved by the TSDp, TSDs and TSDh lexicons.

As mentioned earlier, the supervised approach usually performs well on the sentiment classification task. In this study, we have used one of the most well-known learning algorithms in text classification as a baseline method to compare with the proposed lexicon-based method. As can be seen, the lexicon-based method (TSDp) outperforms SVM in the hotel reviews by achieving the accuracy of 0.807. However, this superiority drops in the movie reviews and SVM yields the best results. This performance is expected since a larger training data is available in the movie reviews. Nevertheless, the lexicon-based method by using a simple strategy and without any dependency to context can yield a comparable result with the supervised approach. On the other hand, the best features are selected from a big context for the supervised approach which leads to gain more accurate results by F-measure values of 0.852 and 0.840 for Positive and Negative classes respectively. However, in a relatively small context (i.e. hotel reviews), the lexicon-based method (TSDp) performs better than SVM, especially in the Positive class by obtaining the F-measure value of 0.822 in comparison with 0.763. In this experiment, it can be again seen the successful performance (F-measure) of the three TSD lexicons in the Positive class over the both benchmarks. This success highlights the positive *SubScores* of the three lexicons which have been calculated from the subjective terms of SWN. Therefore, it can be concluded that the translation of positive terms is more successful than the negative ones and a low-level ambiguity may exist in the positive subjective terms. Nevertheless, the results obtained from lexicon-based and supervised approaches prove that the Turkish sentiment lexicon achieved by the parallel translation approach can be known as a reliable sentiment lexicon for Turkish language. It is also expected that the two other Turkish lexicons can perform well by a rule-based classifier. Because in the used classification method, there is no any control on the negation, intensification and other linguistic features.

[8]http://www.cs.waikato.ac.nz/ml/weka/

Table III
THE SUPERVISED SENTIMENT CLASSIFICATION RESULTS IN CONTRAST WITH THE LEXICON-BASED METHOD

| Benchmark | Method | Precision | | Recall | | F-measure | | Accuracy |
|---|---|---|---|---|---|---|---|---|
| | | Pos | Neg | Pos | Neg | Pos | Neg | |
| Hotel Reviews | TSDs | 0.902 | 0.625 | 0.706 | 0.865 | 0.792 | 0.725 | 0.763 |
| | TSDh | 0.909 | 0.614 | 0.701 | 0.871 | 0.792 | 0.720 | 0.761 |
| | TSDp | 0.893 | 0.721 | 0.761 | 0.871 | 0.822 | 0.789 | 0.807 |
| | SVM | 0.917 | 0.731 | 0.653 | 0.941 | 0.763 | 0.823 | 0.797 |
| Movie Reviews | TSDs | 0.857 | 0.471 | 0.618 | 0.767 | 0.718 | 0.584 | 0.664 |
| | TSDh | 0.829 | 0.521 | 0.633 | 0.753 | 0.718 | 0.616 | 0.675 |
| | TSDp | 0.808 | 0.600 | 0.668 | 0.758 | 0.731 | 0.669 | 0.703 |
| | SVM | 0.820 | 0.878 | 0.888 | 0.805 | 0.852 | 0.840 | 0.846 |

## VI. CONCLUSION

Although the supervised approach performs well on the sentiment classification task, the availability of sentiment annotated data is considered as a limitation for this approach. On the other hand, the term-based features like bag-of-words or n-grams cannot make more progress on the performance of this approach in cases that the sentiments of several texts are presented by more ambiguous words or phrases. This is important because natural language is ambiguous. In this condition, the sentiment lexicons play important role in sentiment analysis systems. This is considerable for supervised approach since these lexicons can be used in extracting more effective features along with term-based ones. However, despite the successful performance of using these lexicons in English sentiment analysis systems, they cannot be employed in a new language due to the lack of such lexical resources. Meanwhile, although the manually created lexicons can be more accurate than the automatic ones, this approach is time-consuming and cannot be employed in a plausible time.

Therefore, this study has proposed an automatic translation method to generate a Turkish sentiment lexicon independent from language and domain. Actually, the objective is to automatically create a sentiment lexicon in a short time, which can work as much as the supervised methods and also ones that are created manually. According to this, three sentiment lexicons, named TSDs, TSDp and TSDh, have been proposed for the Turkish language from SentiWordNet. These lexicons have been achieved by an automatic translation approach along with a unification process. As the linguistic knowledge has not been used in the proposed approach, it can be applied to any other language. This approach transfers the sense-level lexicon to word-level in the target language since the sense-level representation of sentence words is complex and requires much linguistic knowledge. Therefore, the sense-level polarity may not be useful for such languages, however, this can be explored in future.

According to the experimental results, the sentiment lexicon achieved by the parallel translation approach (TSDp) performs better than the others in the sentiment classification task along with *AvgSubScores*. The obtained results from the three Turkish lexicons demonstrate that the translation approach performs well over the positive terms and their *SubScores* are more reliable than the negative ones. This can arise from the fact that positive terms likely include less ambiguity than the negative ones in the Turkish language. However, it may change in the other non-Turkish languages since the ambiguity level may be different in such languages. As a result, we have evaluated the proposed approach in Turkish language, and have not subjectively evaluated this system as a sentiment analyzer yet. However, the achieved promising results encourage us to evaluate this approach in different languages in our future works.

## ACKNOWLEDGMENT

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under the Grant No: EEEAG-115E440

## REFERENCES

[1] F. Akba, A. Uçan, E. Sezer, and H. Sever, "Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews," *8th European Conference on Data Mining 2014*, vol. 191, pp. 180–184, 2014.

[2] Z. Boynukalin, "Emotion analysis of turkish texts by using machine learning methods," *Middle East Technical University*, 2012.

[3] M. Kaya, G. Fidan, and I. H. Toroslu, "Sentiment analysis of turkish political news," *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 174–180, 2012.

[4] M. A. Tocoglu and A. Alpkocak, "Emotion extraction from turkish text," *Proceedings - 2014 European Network Intelligence Conference, ENIC 2014*, pp. 130–133, 2014.

[5] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," *Proceedings of the 5th Conference on Language Resources and Evaluation*, pp. 417–422, 2006.

[6] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, p. 271, 2004.

[7] F. Sebastiani and A. Esuli, "Determining the semantic orientation of terms through gloss classification," *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pp. 617–624, 2005.

[8] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Transactions on Information Systems (TOIS)*, vol. 21, pp. 315–346, 2003.

[9] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? finding strong and weak opinion clauses," in *aaai*, vol. 4, 2004, pp. 761–769.

[10] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, 1997, pp. 174–181.

[11] M. Baroni and S. Vegnaduzzo, "Identifying subjective adjectives through web-based mutual information," *Proceedings of KONVENS*, vol. 4, pp. 17–24, 2004.

[12] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, 2003, pp. 105–112.

[13] R. Mihalcea, C. Banea, and J. Wiebe, "Learning multilingual subjective language via cross-lingual projections," *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 976–983, 2007.

[14] V. Perez-Rosas, C. Banea, and R. Mihalcea, "Learning sentiment lexicons in spanish," *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pp. 3077–3081, 2012.

[15] A. Das and S. Bandyopadhyay, "Sentiwordnet for indian languages," *Asian Federation for Natural Language Processing, China*, pp. 56–63, 2010.

[16] A. Das. and S. Bandyopadhyay, "Theme detection an exploration of opinion subjectivity," *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6, 2009.

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

[18] U. Waltinger, "Germanpolarityclues: A lexical resource for german sentiment analysis." *LREC*, 2010.

[19] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientations of words using spin model," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005, pp. 133–140.

[20] D. Gao, F. Wei, W. Li, X. Liu, and M. Zhou, "Cross-lingual sentiment lexicon learning with bilingual word graph label propagation," *Computational Linguistics*, 2015.

[21] R. Dehkharghani, Y. Saygin, B. Yanikoglu, and K. Oflazer, "Sentiturknet: a turkish polarity lexicon for sentiment analysis," *Language Resources and Evaluation*, pp. 1–19, 2015.

[22] O. Bilgin, Ö. Çetinolu, and K. Oflazer, "Building a wordnet for turkish," *Romanian Journal of Information Science and Technology*, vol. 7, pp. 163–172, 2004.

[23] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, pp. 267–307, 2011.

[24] A. Das and S. Bandyopadhyay, "Subjectivity detection in english and bengali: A crf-based approach," *Proceeding of ICON*, 2009.

[25] F. J. Och, "Statistical machine translation: Foundations and recent advances," *Tutorial at MT Summit*, 2005.

[26] R. Dehkharghani, B. Yanikoglu, D. Tapucu, and Y. Saygin, "Adaptation and use of subjectivity lexicons for domain dependent sentiment classification," *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, pp. 669–673, 2012.

[27] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02*, vol. 10, pp. 79–86, 2002.

[28] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417–424, 2002.

[29] A. G. Vural, B. B. Cambazoglu, P. Senkul, and Z. O. Tokgoz, "A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish," in *Computer and Information Sciences III*. Springer, 2013, pp. 437–445.

[30] A. A. Akın and M. D. Akın, "Zemberek, an open source nlp framework for turkic languages," *Structure*, vol. 10, pp. 1–5, 2007.

[31] B. Naderalvojoud, E. A. Sezer, and A. Ucan, "Imbalanced text categorization based on positive and negative term weighting approach," in *International Conference on Text, Speech, and Dialogue*. Springer, 2015, pp. 325–333.