

# SentiTurkNet: a Turkish polarity lexicon for sentiment analysis

Rahim Dehkharghani<sup>1</sup> · Yucel Saygin<sup>1</sup> ·  
Berrin Yanikoglu<sup>1</sup> · Kemal Oflazer<sup>2</sup>

**Abstract** Sentiment analysis aims to extract the sentiment polarity of given segment of text. Polarity resources that indicate the sentiment polarity of words are commonly used in different approaches. While English is the richest language in regard to having such resources, the majority of other languages, including Turkish, lack polarity resources. In this work we present the first comprehensive Turkish polarity resource, *SentiTurkNet*, where three polarity scores are assigned to each synset in the Turkish WordNet, indicating its positivity, negativity, and objectivity (neutrality) levels. Our method is general and applicable to other languages. Evaluation results for Turkish show that the polarity scores obtained through this method are more accurate compared to those obtained through direct translation (mapping) from SentiWordNet.

**Keywords** Sentiment analysis · Polarity lexicon · Polarity extraction · Turkish WordNet

---

✉ Rahim Dehkharghani  
rdehkharghani@sabanciuniv.edu

Yucel Saygin  
ysaygin@sabanciuniv.edu

Berrin Yanikoglu  
berrin@sabanciuniv.edu

Kemal Oflazer  
ko@cs.cmu.edu

<sup>1</sup> Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey

<sup>2</sup> Computer Science Program, Carnegie Mellon University-Qatar, Education City, PO Box 24866, Doha, Qatar

# 1 Introduction

Sentiment analysis and opinion mining on natural language text have received much attention in the last two decades. The goal of sentiment analysis is to extract the aggregate opinion that is embedded in given data. Sentiment analysis is often performed on textual data (e.g., a review or comment), while the input may also be a speech or video recording whose sentimental polarity is to be estimated. Popular applications of sentiment analysis include automatic extraction of the sentiment from the social media, about a particular product, service, or even political events.

Polarity lexicons are commonly used in estimating the sentiment polarity of a review based on the polarity of its constituent words obtained from the lexicon. While simpler approaches only use a polarity lexicon (Vural et al. 2012), more advanced approaches also benefit from labelled data in the given domain (a corpus of reviews with known polarities) (Pang et al. 2002; Ye et al. 2009; Agarwal et al. 2011; Gezici et al. 2012). The latter approaches consisting of statistical and learning-based methods are effective when there is sufficient data to learn from. They also work well in document level where there is sufficient information, but not necessarily in finer grained levels, such as phrase level (Havasi et al. 2013).

General-purpose polarity lexicons such as SentiWordNet (Baccianella et al. 2010) are domain-independent and have the shortcomings that they do not capture sentiment variations across different domains or cultures, nor can they handle the changing aspects of the language; however, these lexicons do provide a fast and scalable approach to sentiment analysis.

A typical example for the shortcomings of domain-independent polarity lexicons is the term “big” that is positive for *room size* in the *hotel* domain but negative when referring to *battery size* in *camera* domain. As for cultural-dependence, one can give the example of the noun “Atatürk” (a former Turkish leader) which is mostly positive in Turkish culture, while it may be neutral in others. In order to solve these issues, domain-dependent and language-dependent (or culturally-dependent) lexicons are required. Another issue is that while languages are changing, polarity resources also need to be updated to reflect the changes. However doing so manually is time consuming, costly and open for bias. Finally, the polarity of an idiomatic phrase may differ from the polarity of its parts. For example, “costing an arm and a leg” has a negative sentiment while no single word has negative polarity in the phrase. Hence, covering idioms is also necessary for sentiment analysis.

The need for domain-specific lexicons is approached by some researchers as an adaptation problem where a general purpose polarity lexicon is adapted to a specific domain using some domain-specific data (Choi and Cardie 2009; Demiröz et al. 2012). Others have worked on constructing a lexicon in a given domain starting from a seed word set (Hatzivassiloglou and Mckeown 1997).

Yet another issue in sentiment analysis is the need for common sense knowledge to properly understand the meaning or sentiment in a given text. Common sense knowledge can be defined as the collection of facts and information that people build up during their lives. While there are common sense knowledge resources such as Cyc (Lenat and Guha 1989) and ConceptNet (Havasi et al. 2007) for English, there

are only few works to build (Cambria et al. 2012) or incorporate common sense knowledge in different recognition/understanding tasks, including sentiment analysis or language understanding (Wang et al. 2013; Sureka et al. 2009).

Numerous polarity resources already exist for English, e.g., SentiWordNet (SWN), SenticNet (SN) (Cambria et al. 2014), and NRC Emotion Lexicon (Mohammad and Turney 2013). On the other hand, the absence of polarity resources in many other languages such as Turkish, hampers the development of sentiment analysis tools and applications in these languages. In order to close this gap in Turkish, we have undertaken the development of the first polarity resource for Turkish.

A simple approach for building polarity resources for non-English languages has been to translate available polarity resources from English. The reason why we did not take the same approach and translate SentiWordNet to Turkish is twofold. Firstly, meaning between languages is often lost in translation. Translating a Turkish word into an English word only implies that this English word is the closest term in English for the given Turkish word, rather than their meaning being equivalent. Indeed, the meaning of many words only exist within a native context: The Turkish word “*gönül*” which is translated to English as “heart/soul/feelings” lacks a single equivalent term in English. Secondly, translation of meaning does not necessarily correspond to translation of the polarity strength in language dependent terms. For example, “Tanrı” [God] is a positive term in Turkish although the term may be objective in another language. Indeed, polarity scores given in SentiWordNet for the synset “supreme-being, God” are (*pos, neg, obj*) = (0, 0, 1), supporting this observation.

In this paper, we propose a semi-automatic method for assigning polarity strengths to the synsets in the Turkish WordNet, by starting from a manually labelled polarity lexicon indicating only the polarity class (positive, negative, or objective/neutral) of these synsets. The method uses the correspondence information obtained from Turkish WordNet and the polarity strength of the equivalent synset from SentiWordNet to derive the polarity strength of a particular synset. Although we applied the proposed methodology on Turkish, our method is language independent and can be applied on other languages.

We evaluated the assigned polarity scores using three different evaluation methods, as explained in Sect. 6.3. Experimental results show that the polarity scores obtained are reliable and our methodology outperforms the baseline method of directly translating SentiWordNet.

The contributions of this paper can then be summarized as (1) proposing a novel language-independent approach to build a polarity lexicon based on WordNet and (2) building the first such sentiment polarity resource for Turkish and evaluating its accuracy.

## 2 Turkish and its challenges in sentiment analysis

Turkish is a member of the Turkic family of Altaic languages. Particular characteristics of Turkish make natural language processing (NLP) and sentiment analysis tasks difficult for this language. Morphologically, Turkish is an

agglutinative language with morphemes attaching to a root word as “beads-on-a-string”. Words are formed by very productive affixations of multiple suffixes to root words, from a lexicon of about 30K root words (not counting proper names.) Nouns do not have any classes nor are there any markings of grammatical gender in morphology and syntax. When used in the context of a sentence, Turkish words can take many inflectional and derivational suffixes. It is quite common to construct words which correspond to almost a sentence in English: For example, the equivalent of the Turkish word: “*sağlamlaştırabileceksek*” in English can be expressed with the fragment *if we will be able to make [it] become strong (fortify it)* (Oflazier and Bozşahin 1994).

For Turkish, the morphological structure of a word is also necessary for sentiment analysis in addition to the root word, as suffixes may change the polarity of a word. For instance, the word *iştahsız* (*having no appetite*), is negative (due to suffix *-sız*), while its antonym, *iştahlı*, is positive (due to suffix *-lı*). Note that the root word itself, *iştah*, is also positive. This issue is handled in our system using morphological analysis to extract and analyze suffixes of synonym terms and gloss in a synset.

### 3 Related work

In this section, We report some works related to polarity lexicon generation and discuss them under three groups: “English”, “Turkish”, and “Other Languages”. Comprehensive surveys on sentiment analysis can be found in Havasi et al. (2013) and Liu (2012).

There exist a few well-known polarity lexicons in English, such as SentiWordNet and SenticNet, as explained in Sect. 4. Some research in the area of polarity lexicon generation aim to improve these resources by modifying the polarity scores so as to be more accurate. For instance, Hung and Lin (2013) re-evaluate the polarities of objective words that make up more than 93 % of all the words in SentiWordNet by assessing their associated sentences. The proposed revision improves the sentiment classification accuracies significantly (by around 4 %). Poria et al. (2013) enrich SenticNet with affective information, by assigning an emotion label to each term, using WordNet Affect (Strapparava et al. 2004). Other work contribute to polarity lexicon generation by taking special cases into account. For instance, Bosco et al. (2013) developed a corpora with irony detection which is a difficult problem for sentiment analysis systems. In Tsai et al. (2013), a two-step method is employed to build a concept-level sentiment dictionary using common-sense knowledge. At the first step, a sentiment value is assigned to each concept in ConceptNet (Havasi et al. 2007). Then a random-walk method is used to improve those sentiment values.

For Turkish, there are no previous efforts for developing sentiment lexicons, but there have been a few attempts for sentiment analysis on Turkish texts. Kaya et al. (2012) have investigated Turkish political news in online media. In this work, unigrams and bigrams together with polar Turkish terms are used as classification features, which in turn are used to train a classifier to estimate the label of unseen documents. The authors have used four different classifiers: Naive Bayes, Maximum

Entropy, SVM, and a character based n-gram language model, and compared their effectiveness. They conclude that the Maximum Entropy classifier and the n-gram language model are more effective than SVM and Naive Bayes classifiers in classifying Turkish political news. The classification accuracy in different cases ranges from 65 to 77 %.

Eroğul (2009) investigated those linguistic information that affect sentiment analysis such as POS tags and negation markers, along with unigrams and bigrams. An NLP tool for Turkish [*Zemberek* (Akin and Akin 2007)] is used to analyse the words. The obtained accuracy on classifying Turkish movie reviews into positive and negative is reported to be 85 %.

Aytekin (2013) designed a model which assigns positive and negative polarities to opinionated texts in Turkish blogs to present a general view on products and services. The model uses semi-supervised learning based on the Naive Bayes approach. The training set consists of Turkish words stating sentiments. Polar words in this work have been translated from English. The obtained accuracy in this work ranges from 65 to 84 % in different cases.

Vural et al. (2012) present a framework for unsupervised sentiment analysis in Turkish text documents. They customized SentiStrength—a sentiment analysis library for English—for Turkish by translating its polarity lexicon. SentiStrength (Thelwall et al. 2012) is a sentiment analysis library which assigns a positive and a negative score to an input text in English. In this work, after segmenting a text into sentences and each sentence into terms, polarity scores are assigned to each sentence by translating English polarity lexicon to Turkish. *Zemberek* is used also for stemming, negation extraction, spell checking and *ASCII* to Turkish conversion. The authors evaluated their framework by applying it on Turkish movie reviews and report a classification accuracy of 76 % on classifying reviews as positive or negative.

In languages other than English and Turkish, we report only one work which is most relevant to ours. Das and Bandyopadhyay (2010) propose a method for building SentiWordNet(s) for three Indian languages: Hindi, Bengali and Telugu. The key focus in this work is translating English SentiWordNet and the Subjectivity Word List (list of polar English terms) (Wilson et al. 2005) to a target language so as to build a polarity resource. They also provide a game which lets a player assign polarity values to each term. The main difference between this work and ours is that in this work, two English polarity resources are translated to the target language, while we use a more complex approach. In fact, we use a direct translation approach as baseline and show that the proposed method outperforms that baseline.

## **4 Polarity resources used in building SentiTurkNet**

### **4.1 English resources**

We have used the following three English resources during the construction of SentiTurkNet.

- *English WordNet* (Miller 1995): This lexical resource groups synonym terms in a set called *synset* that includes a *gloss* (natural language explanation) for each synset. There are about 117,000 synsets in English WordNet.
- *SentiWordNet* (Baccianella et al. 2010): This resource is built with the purpose of supporting sentiment analysis tasks in English. Three polarity scores summing to one are assigned, indicating the positivity, negativity, and objectivity of each English Wordnet synset.
- *SenticNet* (Cambria et al. 2014): This resource assigns numerical values to each term according to its *pleasantness*, *attention*, *sensitivity*, *aptitude* and also the *overall polarity strength*. We have translated this resource to Turkish by a bilingual dictionary<sup>1</sup> and used the *overall polarity strength* as features in our algorithm.

## 4.2 Turkish resources

We have used only one Turkish resource in this work: Turkish WordNet. This resource consists of about 15,000 synsets along with the gloss, equivalent English synset, POS tag and so on (Bilgin et al. 2004). Each synset includes these fields:

- *Synonyms* are the synonym terms in a synset.
- *Gloss* is the Turkish gloss for the synonym list. *Gloss* is not available for all synsets; therefore we added them some explanations from the TDK (Turkish Language Organization) monolingual dictionary<sup>2</sup>.
- *Synset ID* is a unique identifier for each synset.
- *ILI ID* is the Interlingual Index used for mapping the Turkish synset to its equivalent English synset in English WordNet.
- *POS tag* is the part of speech tag of the terms in the synset –noun, verb, adverb, or adjective.
- *Hypernym synset ID* is the synset *ID* of the hypernym synset (denoting a more general concept). This ID is not available for all synsets; therefore we used only those available.
- *Near-antonym synset ID* is the synset *ID* of the near-antonym synset. This ID is not available for all synsets; therefore we used only those available.

A sample entry from Turkish WordNet is provided in the top part of Table 1. The bottom part shows information derived from the manual labelling (Sect. 5.2) and WordNet mapping (Sect. 4.2.1).

In the original version of Turkish WordNet, some of the synsets do not have Turkish gloss. As our approach requires this gloss, we extracted Turkish explanations for synsets from a Turkish dictionary (TDK). This mono-lingual dictionary consists of over 80,000 entries.

---

<sup>1</sup> <http://www.seslisozluk.net>.

<sup>2</sup> <http://www.tdk.gov.tr>.

**Table 1** A synset from the Turkish Wordnet extended with manually assigned sentiment class and English WordNet mapping

Field	Value
<i>Synonyms</i>	<i>güzelleştirmek, süslemek</i>
<i>Gloss</i>	<i>daha güzel hale getirmek</i>
<i>POS tag</i>	<i>Verb</i>
<i>Synset label</i>	<i>Pos</i>
<i>Hypernym synset label</i>	<i>Pos</i>
<i>Near-antonym synset label</i>	<i>Neg</i>
<i>Equivalent English synset</i>	<i>Ameliorate, improve, better, amend...</i>

#### 4.2.1 WordNet mapping

Turkish Wordnet has been already mapped (one to one) to English WordNet by using the *ILIs*. In this mapping, some Turkish synsets have a mapping to English WordNet v2.0 and some others to WordNet v2.1. Since all synsets among different versions of English WordNet have been mapped to each other, we used the existing mappings between Turkish to English synsets, to map the Turkish WordNet to English WordNet 3.0.

As SentiWordNet 3.0 is based on WordNet 3.0, we could then extract the polarity scores of the equivalent English synset, for each Turkish synset from SentiWordNet. These polarity scores are used as two features in Sect. 5.3 and to establish a baseline to which the proposed method is compared against Sect. 6.2.

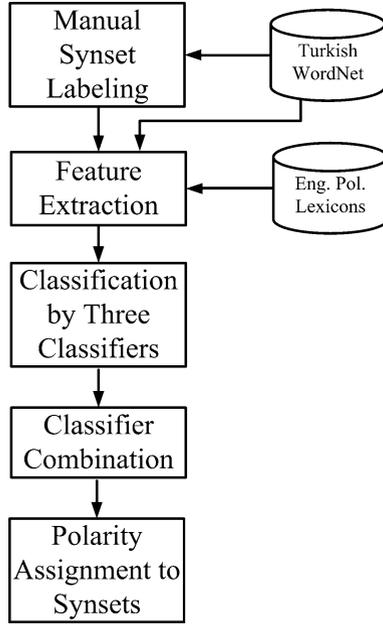
## 5 Building SentiTurkNet

The problem addressed in this paper is to build a polarity lexicon for Turkish, indicating the polarity scores for all the synsets in the Turkish WordNet (14,795 of them). The assigned polarity scores are triplets indicating the positivity, negativity, and objectivity strength of each synset, summing to 1 as in SentiWordNet.

The proposed methodology starts manually assigning one of the three polarity classes (positive, objective/neutral, or negative) to each one of the synsets. Note that this is a relatively easy step compared to the ultimate goal of assigning sentiment polarities to each synset, not just class labels.

After the manual labelling, we extract various features about the synsets from the resources indicated in Sect. 4. The extracted features include some characteristics of the synonyms and gloss of the synset, as indicated by different resources. We then build a classifier to learn this classification given the features extracted from the synsets. In other words, the classifier learns the mapping from extracted features to polarity classes and once it is trained, the confidence scores returned by the classifier for a given synset  $s_i$  are used as the polarity strength values  $pos(s_i)$ ,  $obj(s_i)$ ,  $neg(s_i)$ .

The process is illustrated in Fig. 1 and can be summarized in four steps that are explained in the following subsections:



**Fig. 1** Flow diagram of the proposed methodology for building SentiTurkNet

- *Step 1* Manually labelling all synsets in Turkish WordNet as positive, negative, or objective (Sect. 5.2).
- *Step 2* Extracting features related to each synset (Sect. 5.3).
- *Step 3* Learning the mapping between synsets described by the extracted features and the three class labels (positive, negative, objective/neutral) through machine learning techniques (Sect. 5.4).
- *Step 4* Combining output of the classifiers to obtain more accurate results. (Sect. 5.5)

## 5.1 Resource generation

In addition to the resources mentioned in Sect. 4, we developed and used two small polarity lexicons in extracting features for the classification.

*Polar Word Set (PWS)* We have semi-automatically generated a list of polar Turkish terms including 1000 positive and 1000 negative terms using the method proposed by Hu and Liu (2004). This method uses the synonymy and antonymy relations between terms to generate a large polar word set starting from a small seed set.

*Polar words with PMI scores* We have assigned polarity scores to each word in PWS using Pairwise Mutual Information (PMI) score between that word and pure positive or negative Turkish words listed in Table 2.

The *PMI* concept was first introduced by Turney (2002). Our *PMI* scores are calculated according to co-occurrence of two terms in a database of 10,000 Turkish

**Table 2** Pure positive and pure negative Turkish words used in the *PMI* formula

Pos.	<i>harika (excellent), güzel (beautiful/fine), mükemmel (perfect), sevgi (love), inanılmaz (unbelievable), mühteşem (gorgeous), iyi (good), şahane (fantastic), hayırlı(good), olumlu (positive)</i>
Neg.	<i>berbat (terrible), korkunç (terrible), iğrenç (disgusting), rezil (abject), felaket (disaster), kötü (bad), yetersiz (inadequate), üzgün (sad), fena (bad), olumsuz (negative)</i>

sentences that have been manually labelled as positive, negative, or objective (neutral). The *PMI* score of two words  $w_i$  and  $w_j$  is given in Eq. 1.

$$PMI(w_i, w_j) = \frac{P(w_i, w_j)}{P(w_i) * P(w_j)} \quad (1)$$

where  $P(w_i)$  is the probability of seeing  $w_i$  in the above mentioned 10,000 labelled Turkish sentences. Similarly  $P(w_i, w_j)$  is the probability of seeing  $w_i w_j$  sequence in a sentence in the same database.

In our case,  $w_i$  is each one of the polar words in PWS and  $w_j$  is a pure positive or negative word in Table 2. Note that a higher *PMI* score between the term  $w_i$  and positive (or negative) terms indicates a higher positive (or negative) polarity for  $w_i$ .

We calculate the *PMI* score of each word,  $w_i$ , in PWS with ten pure positive words and assign the average of these scores to  $w_i$  as its positivity score (Eq. 2). The negativity score (*NegPMI*) is computed in similar way by using the ten pure negative word list.

$$PosPMI(w_i) = \frac{\sum_{w_j \in PurePos} PMI(w_i, w_j)}{10} \quad (2)$$

where *PurePos* is the above mentioned ten pure positive word list in Table 2.

The word  $w_i$  is then assumed to be positive according to the *PMI* scores, if  $PosPMI(w_i)$  is greater than its  $NegPMI(w_i)$ .

## 5.2 Manual labelling of the polarity lexicon

As the first step, all 14,795 synsets in the Turkish WordNet are manually labelled to indicate only their polarity *class* as positive, negative, or objective. The manual labelling is done by native Turkish speakers. Labelling the synsets in this simple manner, *without* assigning polarity strengths, is needed to train the classifier, whose output scores are then used as polarity values.

## 5.3 Feature extraction

We extract 23 features shown in Table 3 for each synset. The extracted features include some characteristics (e.g. average polarity) of the synonyms and gloss of the synset, as indicated by different resources.

Before feature extraction, the gloss of each synsets are tokenized, then each token is stemmed to extract its root word and suffixes.

**Table 3** Features are extracted for each synset using SenticNet (SN), PolarWordSet (PWS) and SentiWordNet (SWN)

---

Feature name
$f_1$ : Avg. polarity of pos. synonyms based on PMI
$f_2$ : Avg. polarity of neg. synonyms based on PMI
$f_3$ : Avg. polarity of pos. synonyms based on SN
$f_4$ : Avg. polarity of neg. synonyms based on SN
$f_5$ : Number of pos synonyms based on PWS
$f_6$ : Number of neg. synonyms based on PWS
$f_7$ : Number of synonyms that are adjectives
$f_8$ : POS tag of the synset
$f_9$ : Number of capitalized synonyms
$f_{10}$ : Number of pos. synonyms in gloss according to PWS
$f_{11}$ : Number of neg. synonyms in gloss according to PWS
$f_{12}$ : Avg. polarity of pos. terms in gloss based on PMI
$f_{13}$ : Avg. polarity of neg. terms in gloss based on PMI
$f_{14}$ : Avg. polarity of pos. terms in gloss based on SN
$f_{15}$ : Avg. polarity of neg. terms in gloss based on SN
$f_{16}$ : Number of pos. terms in gloss based on PWS
$f_{17}$ : Number of neg. terms in gloss based on PWS
$f_{18}$ : Number of adjectives in gloss
$f_{19}$ : Number of capitalized terms in gloss
$f_{20}$ : Pos. score of equivalent synset in SWN
$f_{21}$ : Neg. score of equivalent synset in SWN
$f_{22}$ : Label of hypernym synset
$f_{23}$ : Label of near-antonym synset

---

- $f_1$ – $f_4$ : The first four features compute the average polarity scores of synonyms in a synset using different resources. The first two features are the average PMI score of positive and negative terms, as classified according to their PosPMI and NegPMI scores. The next pair of features uses the polarity scores of SenticNet. In SenticNet, we assume a term (or phrase) is positive if its polarity score is greater than or equal to zero or as negative otherwise. Note that simply using the average polarity of all synonyms would require also using the purity measure. We take a different and more symmetric approach and use the average polarity of positive and negative synonyms separately.
- $f_5$ – $f_6$ : These features capture the frequency of positive and negative polar terms in each synset according to PWS.
- $f_7$ – $f_9$ : These features cover certain characteristics of synonyms.  $f_7$  captures the number of synonyms in a synset that are adjective. Generally, those synsets with higher number of adjectives are more subjective. Adverbs are not considered in  $f_7$  because less than 1 % of the synsets are tagged as adverbs.  $f_8$  captures the part of speech tag of the synset. The rationale behind  $f_8$  is that adjective and adverb

synsets have a tendency to be more subjective than do noun or verb synsets.  $f_8$  is different from  $f_7$  in that, some synsets tagged as adjective have non-adjective synonyms.  $f_9$  is the number of synonyms that start with a capital letter. These synonyms (generally proper nouns) are most probably objective e.g. “Milli Güvenlik Kurulu” (National Security Corporation).

- $f_{10}$ – $f_{11}$ : Similar to  $f_5$ – $f_6$ , this pair represents the frequency of positive and negative polar terms in a gloss.
- $f_{12}$ – $f_{15}$ : Similar to  $f_1$ – $f_2$ , this set computes the average polarity scores of the terms (unigrams and bigrams) in a gloss.
- $f_{16}$ – $f_{17}$ : Similar to  $f_5$ – $f_6$ , this pair represents the frequency of polar terms in a gloss.
- $f_{18}$ – $f_{19}$ : Similar to  $f_7$  and  $f_9$ , these features represent the number of adjectives and (first letter) capitalized terms in gloss.
- $f_{20}$ – $f_{21}$ : This pair indicates the positivity and negativity scores of equivalent English synset (in SentiWordNet), as found by the mapping via WordNet and SentiWordNet.
- $f_{22}$ – $f_{23}$ : This pair indicates the polarity class of hypernym and near-antonym synsets of a given synset. Most of the synsets in Turkish WordNet have hypernymy and near-antonymy relations with other synsets which can be used to estimate the polarity of the given synset. Some synsets in Turkish WordNet lack the hypernymy or near-antonymy relations; if these relations are not available, a default value (e.g.  $-1$ ) is assigned to  $f_{22}$  and  $f_{23}$ .

## 5.4 Synset classification

We trained three different classifiers to learn the mapping between features and polarity classes: logistic regression (*LR*) (Hosmer Jr and Lemeshow 2004), feed-forward neural networks (*NN*) (Haykin 1994), and support vector machine with sequential minimal optimization algorithm (*SMO*) (Burges 1998). These three classifiers are some of the most commonly used classifiers for various reasons, such as good generalization accuracy (SVM, NN) and simplicity and computing posterior probabilities (LR). We used *Weka 3.6* (Holmes et al. 1994) for implementing these classifiers.

## 5.5 Classifier combination

After training the base classifiers, we used a classifier combination method called *stacking*, to learn how to combine the individual classifier results. Classifier combination is a commonly used technique for improving generalization accuracy (Mitchell 1997). In this approach, the output of these three base classifiers are given as input to a final classifier which learns to map them to the desired polarity classes.

In our case, the training set of the new classifier receives input samples that consist of confidence scores obtained from each classifier as features ( $3 \times 3 = 9$  features), along with the label (the known polarity class of the corresponding synset). During testing, given a synset, the classifier assigns different confidence

values to each of the three classes; we then interpret the output  $o_i$  as the polarity strength of the synset for the corresponding class  $i$  (positive, negative, and objective). Classifier combination brought an increase of 8 % points in classification accuracy, over the base classifiers.

## 5.6 Example

In Table 4, we provide a real example for the proposed methodology. The top part of the table shows the information obtained from the extended Turkish WordNet, while the bottom parts shows the scores assigned by mapping from SentiWordNet and the proposed method. For the latter, we give the results of the three base classifiers and the combination (indicated as SentiTurkNet score). As can be seen with this language/culture dependent synset, the result of the proposed method is in accordance with the term that is accepted as mostly positive in Turkish. On the other hand, polarities obtained from translations from SentiWordNet indicate it as objective.

## 6 Experimental results

### 6.1 Data set

In the evaluations, we either used a small test set that was sequestered for this purpose or all of the data using cross-validation.

The test set is a small subset of the synsets (3 %) that is kept sequestered for testing purposes. For this subset, called the gold standard set, we manually assigned a quantized polarity strength value to each synset in one of eight possible polarity levels ranging from 0 to 7. The reason for using this categorization was so that we could compare our resource with SentiWordNet where the same quantization is used (multiples of 0.125, between 0 and 1) and because assigning a value in a finer resolution would have been difficult.

### 6.2 Methodology

We evaluated the proposed approach in three different tests:

- *Test 1* Mean absolute error (MAE) between manually assigned ground-truth polarities on a small test set and the polarities estimated by the proposed method;
- *Test 2* Misclassification error of the proposed method as calculated by comparing the estimated class labels with those assigned manually, using fivefold cross-validation on all data;
- *Test 3* Sentiment analysis improvements when using *SentiTurkNet* instead of the mapped *SentiWordNet* to Turkish for classifying Turkish movie reviews.

As baseline, we use the MAE and misclassification error rates obtained by using a direct mapping from Turkish to English synsets. Specifically, since Turkish WordNet has been mapped (one to one) to English WordNet, the polarity scores of

**Table 4** An entry from SentiTurkNet, together with assigned polarities

Field	Value
<i>Synonyms</i>	<i>Cuma namazi</i> [Friday Prayers]
<i>Gloss</i>	<i>Müslümanların Cuma günleri yaptığı ibadet</i> [Worship muslims perform on Friday]
<i>POS tag</i>	<i>Noun</i>
<i>Synset label</i>	<i>Pos</i>
<i>Hypernym synset label</i>	<i>Pos</i>
<i>Near-antonym synset label</i>	<i>Not specified</i>
<i>Equivalent English synset</i>	<i>salat, salah, salaat...</i>
<b><i>SentiWordNet scores</i></b>	<b><i>(P, O, N) = (0, 1, 0)</i></b>
<i>score by NN</i>	<i>(P, O, N) = (0.52, 0.45, 0.02)</i>
<i>score by LR</i>	<i>(P, O, N) = (0.54, 0.45, 0.01)</i>
<i>score by SMO</i>	<i>(P, O, N) = (0.33, 0.66, 0.01)</i>
<b><i>SentiTurkNet scores</i></b>	<b><i>(P, O, N) = (0.49, 0.44, 0.06)</i></b>
<b><i>SentiTurkNet label</i></b>	<b><i>Pos</i></b>

The bold rows are for highlighting the final polarity scores assigned by two polarity resources (SentiWordNet and SentiTurkNet) for the given synset in SentiTurkNet and its equivalent synset in SentiWordNet, plus its label assigned by SentiTurkNe

an English synset are used as polarity scores of its equivalent Turkish synset, from which class labels are also deduced.

## 6.3 Results

### 6.3.1 Test 1

In the first evaluation, we used the small and sequestered test set and compared the (*MAE*) between the manually assigned ground-truth polarities on this set and the ones obtained with proposed methodology. The *MAE* values presented in Table 5 are computed using Eq. 3.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (3)$$

where  $f_i$  is the estimated polarity level (0–7) and  $y_i$  is the ground-truth polarity level of the  $i_{th}$  synset, computed separately for positive and negative cases and  $n$  is the number of evaluated synsets.

As seen in this table, the *MAE* computed over all the synsets by the final system is 2.45 and 1.95 (with a weighted average of 2.28) for positive and negative synsets, respectively. Note that the ground-truth scores are eight polarity levels ranging from 0 to 7, and e.g. 2.28 means that the error rate in estimating the polarity level of a synset by proposed methodology is 0.31 and 0.24 over 1. These results support the assumption that translating (mapping) *SentiWordNet* to another language can be improved as proposed.

**Table 5** MAE on test data

Classifier	Pos	Neg	Avg
<i>SentiWordNet</i> mapped to Turkish	3.73	3.01	3.48
<i>SentiTurkNet</i> with <i>SMO</i>	2.95	2.21	2.70
<i>SentiTurkNet</i> with <i>LR</i>	2.81	2.25	2.62
<i>SentiTurkNet</i> with <i>NN</i>	2.99	2.14	2.70
<i>SentiTurkNet</i> with classifier combination	<b>2.45</b>	<b>1.95</b>	<b>2.28</b>

The bold values show the lowest error rate we obtained by classifier combination

### 6.3.2 Test 2

In the second test, we evaluated the classification of synsets into three polarity classes, using the trained classifier. Note that here we are evaluating the outcome of the trained classifier in comparison with the manually assigned labels. If the manually assigned label differs from the label of maximum polarity score (out of three scores for (*pos*, *obj*, *neg*), this is counted as a misclassification error.

We used fivefold cross-validation where the mapping between features and three polarity classes is learned using 80 % of the data (training set) and the system is tested with the remaining 20 % of the data, for an unbiased testing. This process is repeated five times with different 80–20 % splits of the data and the results are averaged.

In this test we also evaluated the relative importance of different feature groups. Features  $f_1$ – $f_9$  extract polarity from synonym list;  $f_{10}$ – $f_{19}$  are related to the gloss;  $f_{20}$ – $f_{21}$  correspond to polarities that would be obtained if a direct mapping was used; and  $f_{22}$ – $f_{23}$  uses the manually assigned class label of hypernym and near-antonym synsets.

As seen in Table 6, the best accuracy of 91.11 % is achieved using all features and classifier combination of three classifiers. The feature pairs  $f_{20}$ – $f_{21}$  (polarity

**Table 6** Classification accuracy by the individual classifiers using fivefold cross validation on all data(%)

Feature subset	Accuracy (%)			
	SMO	NN	Logistic	Classifier combination
$f_1$ – $f_9$	79.03	79.71	79.42	86.72
$f_{10}$ – $f_{19}$	79.02	78.74	78.97	85.26
$f_{20}$ – $f_{21}$	79.03	79.16	79.22	86.11
$f_{22}$ – $f_{23}$	81.63	81.99	81.93	87.32
$f_1$ – $f_{19}$	79.05	79.79	79.56	85.07
$f_1$ – $f_{21}$	79.05	79.85	80.14	87.99
$f_1$ – $f_{23}$	81.90	82.44	82.01	88.82
<b>All features</b>	<b>82.89</b>	<b>83.32</b>	<b>83.13</b>	<b>91.11</b>

The bold values show the highest accuracy we obtained by using all features

scores of corresponding English synsets) and  $f_{22}$ – $f_{23}$  (manually assigned class labels of hypernym and near-antonym synsets) are good indicators for the polarity of a synset, as expected. However, we see that by adding the other features, we are able to obtain higher accuracies (up to 4 % points).

### 6.3.3 Test 3

The last evaluation studies sentiment analysis improvements when using *STN* instead of the mapped *SWN*, for classifying Turkish movie reviews. More specifically, we use polarity scores obtained from *STN* or from the mapped *SWN*, to classify 300 reviews from Turkish movie dataset.<sup>3</sup>

The method simply tokenizes the reviews and extracts the average polarity of terms in each review, to feed to a simple sentiment analysis classifier [by logistic regression (Hosmer Jr and Lemeshow 2004)] we had developed previously. The accuracy of ternary classification (positive, negative, objective) by logistic regression and fivefold cross-validation method using *STN* is 66.7 % while it is 61.3 % by using the mapped *SWN* to Turkish.

The low accuracy may be caused by the lack of language features such as negations, conjunctions, and intensifiers; our goal was to show the difference between polarity scores in two polarity resources by using them in a sentiment classification task. An example review that was correctly classified as positive using *STN* but incorrectly classified as negative using *SWN* is “*Sadece müziği için bile izlenir*” [It can be watched just for of its soundtrack].

We did not do word-sense disambiguation (WSD) within the sentiment analysis system, as WSD is an ongoing problem in Turkish and is out of the scope of this work. Instead, for a given term with a given POS tag, we simply used the average polarity of all of its synsets with a matching POS tag. No *NLP* technique except extracting the root of words and their POS tag is used for this purpose.

The misclassified reviews by our system generally are those that include words which are absent in *STN* or those that are subjective but need background knowledge to distinguish this subjectivity such as “*izlerken bana ordaymış hissi veren nadir filmlerden.*” [“of those rare movies that gives the feeling of being there (in movie) while watching”].

## 6.4 Discussion

Results presented in Sect. 6.3 indicate that the proposed methodology is quite successful in predicting the label of synsets. Using only the mapping approach (baseline) would correspond to  $f_{20}$ – $f_{21}$  that get us 86.11 % accuracy in classifying synsets as positive, negative, or objective, but by using all features, we obtain 91.11 % accuracy.

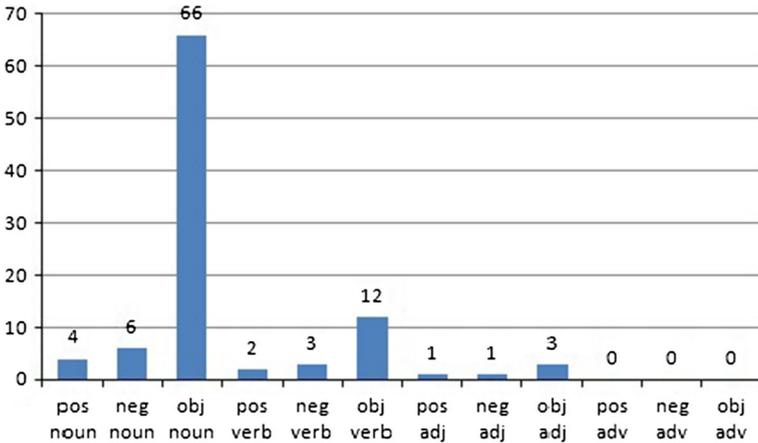
The errors are mostly caused by features related to glosses. It is common for a positive (or negative) synset to be explained by a non-positive (or non-negative)

---

<sup>3</sup> This dataset is collected from <http://www.beyazperde.com>.

**Table 7** A negative synset misclassified as neutral (objective)

Fields	Content
Synonym	iştahsız
Gloss	Yemek yeme isteği olmayan, boğazsız [no desire to eat]
Actual label	Neg.
Estimated label	Obj.



**Fig. 2** Distribution (%) of pos/neg/obj parts of speech in SentiTurkNet

sentence. In most of the synsets, this deficiency is compensated by other features. An example for this statement is given in Table 7.

The distribution (in percent) of positive, objective, and negative synsets in each part of speech is illustrated in Fig. 2. As can be seen, the majority of synsets are objective in all parts of speech. Note that the situation is similar for SentiWordNet as well, where the overwhelming majority of all words are marked as dominantly objective. Also among four parts of speech, nouns constitute the majority. Note that because of the low percentage of adverbs (less than 1 %), they do not appear in this chart.

## 7 Summary and contributions

The two contributions of this work are building the first comprehensive polarity lexicon for Turkish (SentiTurkNet) and proposing a semi-automatic approach to do this for other languages as well. The developed lexicon contains polarity score triplets for all synsets in the Turkish WordNet, containing almost 15,000 synsets.

SentiTurkNet is thus based on Turkish WordNet and is mapped (one to one) to English WordNet and consequently to SentiWordNet.

The quality of the lexicon is established using different approaches, including low MAE between the estimated and the manually assigned polarities for a small portion of the lexicon for which groundtruth exists. Furthermore, we showed that the use of the new lexicon result in higher classification accuracy in sentiment classification, compared to using translated resources.

The shortcoming of the developed lexicon is its relatively small coverage size. As for the proposed methodology, it is applicable to any language for which a WordNet exists, but it is time consuming to manually label the polarity classes of the synsets.

Here we compare SentiTurkNet with SentiWordNet because it is the most similar resource to SentiTurkNet and the main idea for building SentiTurkNet has been derived from SentiWordNet. The similarities and differences are as follows:

- Both resources benefit from the polarity of the gloss of a synset as a feature to estimate the polarity scores for the synset.
- Both resources assign polarity scores to each synset in WordNets of different languages such that the sum of these scores equals to one.
- English WordNet (and consequently SentiWordNet) has around 117,000 synsets while Turkish WordNet (and SentiTurkNet) has 15,000 synsets.
- In SentiWordNet, the label of a synset is estimated as one of eight categories; hence, polarity scores in SentiWordNet are multiples of 0.125, while the polarity scores in SentiTurkNet are continuous values in the range [0, 1].

## 8 Conclusion and future work

Sentiment analysis especially in non-English languages suffers from the shortage of polarity resources. We built the first polarity resource for Turkish by a novel methodology. This methodology is inspired by the idea used in building the SentiWordNet; however, our overall methodology is novel. This methodology can be employed to build such polarity resources for other non-English languages.

We conclude that translating polarity resources from another language (English) to Turkish (or any other language) is not the best approach because (1) not all terms in a language have equivalent terms in other languages and (2) language/cultural dependent terms generally possess different polarities in different languages. Despite these difficulties, polarity resources from other languages can be used to extract features; for example, we used positivity and negativity scores of each synset in SentiWordNet and SenticNet as features for classification of Turkish synsets.

We have made a subset of this resource public which can be downloaded from (<http://myweb.sabanciuniv.edu/rdehkharghani/sentiwordnet/>). We will make the entire resource publicly available in near future.

The constructed polarity resource is the first version (v1.0) of SentiTurkNet. We will extend this resource in the near future by (1) covering negation in Turkish gloss which may increase the accuracy of estimated results, and (2) benefiting from dependency parse trees, for analysing the gloss.

## References

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In: *Proceedings of the workshop on languages in social media* (pp. 30–38). Association for Computational Linguistics.
- Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turkic languages. *Structure, 10*, 1–5.
- Aytekin, C. (2013). An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities. *Journalism and Mass Communication, 3*(3), 179–198.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *LREC, 10*, 2200–2204.
- Bilgin, O., Çetinoğlu, Ö., & Oflazer, K. (2004). Building a wordnet for Turkish. *Romanian Journal of Information Science and Technology, 7*(1–2), 163–172.
- Bosco, C., Patti, V., & Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems, 2*, 55–63.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121–167.
- Cambria, E., Hussain, A., Durrani, T., & Zhang, J. (2012). Towards a chinese common and common sense knowledge base for sentiment analysis. In H. Jiang, W. Ding, M. Ali, X. Wu (Eds.), *Advanced research in applied artificial intelligence* (pp. 437–446). Berlin, Springer.
- Cambria, E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In: *Twenty-eighth AAAI conference on artificial intelligence* (pp. 1515–1521).
- Choi, Y., & Cardie, C. (2009). Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: *Proceedings of the 2009 conference on empirical methods in natural language processing: Volume 2-Volume 2* (pp. 590–598). Association for Computational Linguistics.
- Das, A., & Bandyopadhyay, S. (2010). *SentiWordNet for Indian languages* (pp. 56–63). China: Asian Federation for Natural Language Processing.
- Demiröz, G., Yanikoglu, B., Tapucu, D., & Saygin, Y. (2012). Learning domain-specific polarity lexicons. In: *2012 IEEE 12th international conference on data mining workshops (ICDMW)* (pp. 674–679). IEEE.
- Eroğul, U. (2009). *Sentiment analysis in Turkish*. MSc thesis, Middle East University, Turkey.
- Gezici, G., Yanikoglu, B., Tapucu, D., & Saygin, Y. (2012). New features for sentiment analysis: Do sentences matter? In: *SDAD 2012 The 1st international workshop on sentiment discovery from affective data* (p. 5).
- Hatzivassiloglou, V., & Mckeown, K. R. (1997). Predicting the semantic orientation of adjectives. In: *Proceedings of ACL-97, 35th annual meeting* (pp. 174–181). Association for Computational Linguistics.
- Havasi, C., Cambria, E., Schuller, B., Liu, B., & Wang, H. (2013). Knowledge-based approaches to concept-level sentiment analysis. *IEEE Intelligent Systems, 28*(2), 12–14.
- Havasi, C., Speer, R., & Alonso, J. (2007). ConceptNet 3: A flexible, multilingual semantic network for common sense knowledge. In N. Nicolov, G. Angelova and R. Mitkov (Eds.), *Recent advances in natural language processing (RANLP)* (pp. 27–29). Borovets.
- Haykin, S. (1994). *Neural networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.
- Holmes, G., Donkin, A., & Witten, I. H. (1994). Weka: A machine learning workbench. In: *Proceedings of the 1994 second Australian and New Zealand conference on intelligent information systems, 1994* (pp. 357–361). IEEE.
- Hosmer, D. W, Jr, & Lemeshow, S. (2004). *Applied logistic regression*. New Jersey: Wiley.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177). ACM.
- Hung, C., & Lin, H.-K. (2013). Using objective words in SentiWordNet to improve word-of-mouth sentiment classification. *IEEE Intelligent Systems, 28*(2), 0047–54.
- Kaya, M., Fidan, G., & Toroslu, I. H. (2012). Sentiment analysis of Turkish political news. In: *Proceedings of the the 2012 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technology-Volume 01* (pp. 174–180). IEEE Computer Society.

- Lenat, D. B., & Guha, R. V. (1989). *Building large knowledge-based systems; Representation and inference in the Cyc project*. Menlo Park: Addison-Wesley Longman Publishing Co., Inc.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. USA: Morgan and Claypool Publishers.
- Miller, G. A. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Mitchell, T. M. (1997). *Machine learning*., McGraw-Hill series in computer science New York: McGraw Hill.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- Oflazer, K., & Bozşahin, H. C. (1994). Turkish natural language processing initiative. In: *Proceedings of the third Turkish symposium on artificial intelligence and artificial neural networks, Ankara, Turkey*.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing-Volume 10* (pp. 79–86). Association for Computational Linguistics.
- Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., & Bandyopadhyay, S. (2013). Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2), 31–38.
- Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: An affective extension of wordnet. *LREC*, 4, 1083–1086.
- Sureka, A., Goyal, V., Correa, D., & Mondal, A. (2009). Polarity classification of subjective words using common-sense knowledge-base. In H. Sakai, M. Chakraborty, A.-E. Hassanien, D. Slezak, W. Zhu (Eds.), *Rough sets, fuzzy sets, data mining and granular computing* (pp. 486–493). Berlin: Springer.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163–173.
- Tsai, A. C.-R., Wu, C.-E., Tsai, R. T.-H., Hsu, J. Y.-J., et al. (2013). Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intelligent Systems*, 28(2), 22–30.
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424). Association for Computational Linguistics.
- Vural, A. G., Cambazoğlu, B. B., Şenkul, P., & Tokgöz, Z. Ö. (2012). A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. In: E. Gelenbe, & R. Lent (Eds.), *ISCIS* (pp. 437–445). London: Springer.
- Wang, Q.-F., Cambria, E., Liu, C.-L., & Hussain, A. (2013). Common sense knowledge for handwritten chinese text recognition. *Cognitive Computation*, 5(2), 234–242.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347–354). Association for Computational Linguistics.
- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 6527–6535.