

The Impact of Sentiment Features on the Sentiment Polarity Classification in Persian Reviews

Ehsan Asgarian¹ · Mohsen Kahani¹ · Shahla Sharifi²

Received: 22 December 2016 / Accepted: 26 September 2017 / Published online: 7 November 2017
© Springer Science+Business Media, LLC 2017

Abstract Natural language processing (NLP) techniques can prove relevant to a variety of specialties in the field of cognitive science, including sentiment analysis. This paper investigates the impact of NLP tools, various sentiment features, and sentiment lexicon generation approaches to sentiment polarity classification of internet reviews written in Persian language. For this purpose, a comprehensive Persian WordNet (FerdowsNet), with high recall and proper precision (based on Princeton WordNet), was developed. Using FerdowsNet and a generated corpus of reviews, a Persian sentiment lexicon was developed using (i) mapping to the SentiWordNet and (ii) a semi-supervised learning method, after which the results of both methods were compared. In addition to sentiment words, a set of various features were extracted and applied to the sentiment classification. Then, by employing various well-known feature selection approaches and state-of-the-art machine learning methods, a sentiment classification for Persian text reviews was carried out. The obtained results demonstrate the critical role of sentiment lexicon quality in improving the quality of sentiment classification in Persian language.

Keywords Opinion mining · Persian sentiment word miner · Feature engineering · Comprehensive Persian WordNet

Background

Well-grounded knowledge of popular opinion is essential in the decision-making process, both for consumers and executives of companies and organizations pertinent to the industry in question. Currently, with the advent of Web 2.0, a vast amount of content reflecting opinion is being generated. However, the process of analyzing opinions presents a challenge due to the large quantity of documents, opinion polls, and the conflicting viewpoints on any given subject. Therefore, there is evident demand for the retrieval and analysis of Web comments or reviews. In recent decades, AI researchers have sought to endow machines with cognitive capabilities to recognize, interpret, and express emotions and sentiments [1].

Supplying machines with cognitive capabilities to recognize, interpret, and express emotions and sentiments has been among the most important topics in artificial intelligence field of study.

In recent years, natural language processing studies have become more oriented toward opinion mining. An important function of opinion mining is the classification of documents according to an overall sentiment, whether it be positive or negative. Sentiment analysis is a major topic in Affective Computing research [1]. Initial studies on opinion mining frequently attempted to classify the opinions or overall sentiments of a document as either positive or negative feedback [2].

These classifications do not address all aspects of opinions containing subtle linguistic forms, simultaneous expression of positive and negative nuances, and implicit judgments based on explicit ones [3]. Consequently, NLP must be supplemented by cognitive and social perspective to resolve such issues.

✉ Mohsen Kahani
kahani@um.ac.ir

Ehsan Asgarian
ehsan.asgarian@mail.um.ac.ir

Shahla Sharifi
sh-sharifi@um.ac.ir

¹ Department of Computer Engineering, Ferdowsi University of Mashhad, Azadi Sq., Mashhad, Iran

² Department of Linguistics, Faculty of Letters and Humanities, Ferdowsi University of Mashhad, Mashhad, Iran

Researchers then tried to determine the degree of satisfaction or dissatisfaction with the document, instead of the previous two-state classification [4]. A considerable complication presents itself at this level with the erroneous assumption that the topic in question is the same throughout a text or document, while different parts of a document (different reviews) may deal with varying issues.

It is therefore, essential to identify the topics within different sections individually rather than analyzing the overall sentiment in reviews in a collective manner. Consequently, some researchers have conducted analyses on sentiment at the sentence level [5] or semantic phrase level [6]. At this level, a subjectivity analysis is carried out to distinguish between subjective and objective sentences (e.g., irrefutable facts, news reports). Neither document-level nor sentence-level analyses can reveal the target of the opinion. To obtain this level of fine-grained results, we need to go to the aspect level (aspect-based sentiment analysis) [7]. New generation opinion mining and NLP techniques [8–10] contain resources and the integration of a biologically inspired paradigm with statistical approaches, in order to understand and extract concepts from texts.

In recent years, many studies, conducted in this same fashion, have focused on non-English languages, specifically Spanish, Chinese, German, Czech, and Arabic [11–15]. Relatively new approaches to multilingual opinion mining are currently being developed [16]. Most research in this area is document-level sentiment analysis or the sentence-level sentiment analysis (subjectivity analysis) [17, 18]. However, few studies have addressed aspect-based opinion mining [19, 20]. These methods mostly employ machine translators based on relatively simple ideas in order to utilize a set of sentiment lexicons from other resources and some English text processing tools for the intended language [16, 21–23].

As the importance of using opinion mining, for the purpose of identifying public opinion increases, namely in the commercial sector, accuracy in classification becomes more vital. Therefore, in some new studies, the effects of several aspects of data representations and feature selection methods on the sentiment classification are investigated in various languages, such as English, Arabic, and Czech [15, 24–26]. In these, the feature vectors of the reviews were preprocessed using different methods and the resulting effects on the accuracy of different classifiers discussed.

The results obtained from these methods are, however, not accurate enough to be applied to other languages due to the differences in their syntactic rules and grammar, sentiment idioms or terms, and other intricacies of natural language. The Persian language is one of the Indo-European languages spoken by more than one hundred million people worldwide and it is the official language of three countries, namely Iran, Afghanistan (known as Dari language), and Tajikistan (known as Tajik language). From a computational standpoint, little attention has been paid to Persian language due to its complexities and limited resources [27, 28].

Determining feature sets is particularly challenging and yet highly significant when it comes to biologically inspired machine learning approaches. To the best of our knowledge, no study has been done to investigate the impact of sentiment lexicon and NLP tools for sentiment polarity classification in Persian language text. The main aim of this paper is to analyze the impact of several preprocessing tools and sentiment lexicon on the sentiment classification from different viewpoints. First, to achieve this, the required text-processing tools, a comprehensive Persian WordNet (FerdowsNet), a Persian SentiWordNet, and a Persian corpus for opinion mining, were developed. Next, an in-depth analysis of different methods of feature selection and sentiment classification of Persian review texts was developed.

In the following subsections, previous research on sentiment analysis as well as some sentiment lexicon generation methods are discussed in further detail.

Sentiment Classification Methods

In general, sentiment classification methods can be categorized into two groups: (i) methods using a sentiment lexicon or background knowledge (unsupervised or semi-supervised learning) and (ii) methods using supervised machine learning algorithms.

Recently, the first group of methods has attracted many researchers. The accuracy of this approach depends entirely on the accuracy of sentiment lexicon weights [29]. These methods are usually unsupervised and, therefore, independent of specific domains [30]. To classify the sentiments, techniques for assessing the semantic similarities between words should be applied. Here, the semantic similarities of expressions and a short list of initial sentiment words are used to classify the sentiments within reviews. In general, three methods can be used to assess semantic similarity:

1. Ontology-based methods (e.g., employing WordNet, ConceptNet, or other dictionaries and encyclopedias) [31, 32];
2. Extraction of syntactic dependency relations between sentiment candidate words and the words of the given lexicon [33, 34];
3. Extraction of the co-occurrence of sentiment candidate words and the words of the initial list (unsupervised learning methods) in sentences from various corpora (user reviews, blogs, web pages, or search engines) [35, 36].

In the second group, a set of opinion documents (reviews), labeled with negative or positive sentiments, are provided as the training data. Then, each document is represented as a vector of features. Several features have been used for this purpose. In previous publications, words along with their frequency, n-grams, POS¹ tags of words [37, 38], sentiment words and phrases, the place of each word in the document [39], negative words, and syntactic dependency [40] are given

¹ Parts of Speech

to the classifier algorithm as the input features of each sentence. After, the classifier algorithm is trained by applying the training set and a model is built. Finally, this model is used to determine the sentiments of other opinions (test set).

To improve the quality and efficiency of machine learning algorithms, superior features should be selected. Many feature selection methods (described in detail in [Feature Engineering](#) section) could be applied for this purpose.

Numerous studies use several classification algorithms such as Naive Bayes (NB), Maximum Entropy (ME), and particularly, Support Vector Machine (SVM) to classify the polarity of reviews [41, 42]. These methods are used in text classification applications as well. In this case, instead of considering keywords and concepts, the expressed sentiments are used for classification.

In most research, opinions are simply divided into two categories, namely positive and negative sentiments [43]. Some researchers have considered an additional category for neutral opinions. Some have even implemented a user-provided rating criterion for opinions (e.g., 1 to 5 stars). These ratings could be employed to build the training dataset. Therefore, in most cases, sentiment classification is considered at document or sentence level.

Sentiment Lexicon Generation Methods

Generating automatic sentiment lexicon is a major task in the field of sentiment analysis. Detecting the polarity of sentiment and its intensity without human assistance is complicated for a machine to perform. Therefore, in sentiment analysis methods, experts first provide a list of primary sentiment terms, along with the numeric values that determine the intensity. The existence of these sentiment terms in a sentence is an important feature for sentiment classifiers.

As mentioned above, most lexicon-based sentiment classifiers use knowledge bases, such as WordNet, to measure polarity. The semantic relations (e.g., synonym, antonym) between words are used to form a graph. At this point, by using the initial sentiment seed words, the polarity is propagated in the graph via various methods, such as shortest path, random walk, PageRank, boot-strapping, and classifiers [44–48].

The SentiWordNet lexicon [49] is one of the best available resources to identify sentiment words. It is generated by determining the sentiment weight of each synset in Princeton WordNet (PWN). SentiWordNet specifies the polarity (negativity, positivity or objectivity²) of each synset. SentiWordNet v1.0 [50] was created in four steps using a semi-supervised learning algorithm:

1. The positive and negative polarities of a limited number of initial synsets are manually indicated and are propagated for relevant synsets.
2. Some objective synsets are also selected and the Glosses of the synsets specified in the previous step are used for the learning phase of the classification method as training data.
3. Using a classification algorithm, other synsets are labeled as “neg,” “pos,” and “obj.”
4. In order to reduce the classification algorithm error, synsets provided in step 2 are used to train several classification algorithms (step 3). The results are then combined.

The initial version (SentiWordNet v1.0) was improved in SentiWordNet v3.0 using the iterative random walk algorithm and WordNet 3.0 graph [49]. SentiWordNet has been employed in many opinion mining applications as a sentiment lexicon, independent of the domain and subject [3, 23, 31]. Also, SentiWordNet and WordNet-Affect [51] have been employed to develop other sentiment lexicons such as SentiFul [52], SenticNet [53, 54], SentslangNet [55], and WSD-based SentiWordNet [29]. Moreover, utilizing the link between SentiWordNet and PWN combined with the link between PWN and non-English WordNets, SentiWordNet has been used in numerous opinion mining applications in other languages as well [56, 57].

Persian language opinion mining studies have predominantly used small and manually-created sentiment lexicons [58, 59]. However, some researchers simply translated English sentiment lexicons to Persian [60, 61]. Dehkharghani et al. proposed a new method to create a Persian sentiment lexicon (UTIIS) using a Persian WordNet (FarsNet³ v1.0) and English resources [47]. They manually created the primary sentiment words (seeds) using English resource translation (Micro-WNOp corpus [62]). They also used the random walk method to propagate the weights of the seed words to determine the weights of the remaining words in the semantic graph of FarsNet. As FarsNet v1.0 was sparse and incomplete, it did not fulfill the requirements of their study. Therefore, they extended their Persian sentiment lexicon (UTIIS) based on PWN synsets. The synsets not covered by FarsNet were included into UTIIS by translating the related synsets in PWN. UTIIS consists of 1815 positive sentiment words and 1856 negative sentiment words organized into three groups of Persian nouns, adjectives, and verbs [47]. Dashtipour et al. [63] developed a new lexicon for Persian language called Per-Sent. The lexicon contains 1500 Persian words accompanied by their respective polarity values, based on a numeric scale ranging from -1 to $+1$, and their parts of speech tags. The words and phrases used in Per-

² Objectivity score = $1 - (\text{positivity score} + \text{negativity score})$

³ A more detailed description of FarsNet and the other Persian WordNets is provided in [Persian WordNet](#) section.

Sent were taken from multiple resources, such as movie review websites, blogs, and Facebook. The majority of the values in Per-Sent were assigned manually.

The structure of the remainder of this paper is as follows: **Methods** section introduces the general architecture of sentiment classification systems. The methods of constructing Persian sentiment lexicon in this study are then described. In this section, the popular features in the literature and various state-of-the-art methods of selecting superior features for the sentiment classification of reviews will be expanded upon. In **Experimental Results** section, the quality assessment results of several text processing tools, various Persian WordNets, and the state-of-the-art methods of sentiment classification for different features will be presented and compared. The final section is the conclusion.

Methods

In order to classify reviews, they first must be pre-processed. Then, with the help of the constructed sentiment lexicon, their features are extracted. After converting the text into numeric vectors, superior features of opinions are identified and selected using feature selection methods. Finally, applying different classification methods, positive and negative opinions are separated. The architecture of sentiment classification system is shown in.

Normalization, segmentation of text (into sentences, phrases and words), tagging, and annotation have significant impacts on the processing and extraction of information, classification and other applications of natural language processing [24]. This paper utilizes Ferdowsi Persian text processing tools. The tools were developed for non-commercial use and are available on the website of the Web Technology Laboratory of Ferdowsi University.⁴ In the rest of this section, first, a few studies aimed at Persian WordNet construction are introduced. Then, the proposed approach for sentiment lexicon generation is discussed. Two approaches in the present study which are applied to extract sentiment words are explained and the qualities of the results are compared in the following sections. In the first method, the links between FerdowsNet, as described in Section 0, and English WordNet synsets and sentiment weights of existing words in the SentiWordNet dictionary are used for Persian sentiment lexicon generation (Fig. 1).

In the second method, experts first labeled a set of reviews with sentiment tags and other specified tags. Then, using a learning algorithm based on HMM (Hidden Markov Model), patterns for expressing sentiment phrases were found and the list of these words was extended. More details on this approach (PSWM) are included in Section 0. Additionally, the state-of-the-art methods for extracting and selecting superior features are presented.

⁴ <http://wtlab.um.ac.ir>

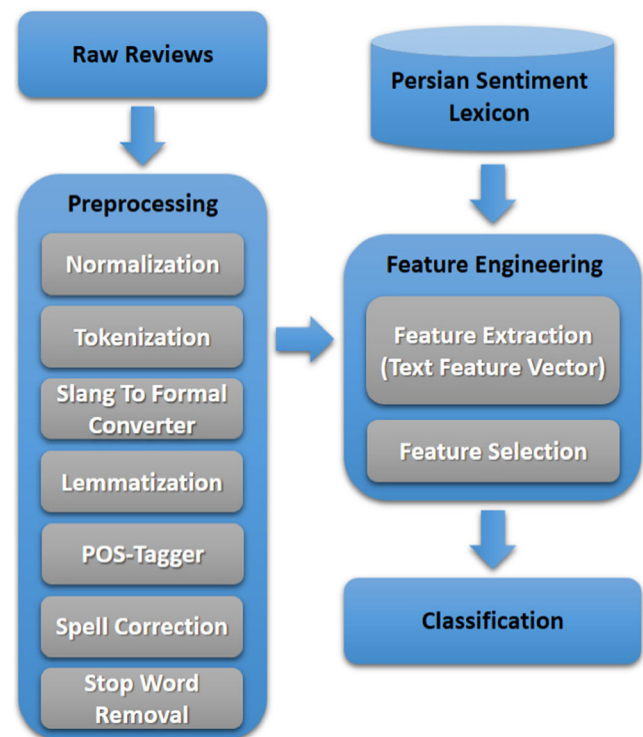


Fig. 1 Persian sentiment classification architecture

Sentiment Lexicon Generation

Sentiment lexicon generation is an essential part of sentiment detection and its intensity. In previous studies, two methods have been applied to generate lexicons of sentiment words: 1—development or translation of expressions from the available lexicons [56, 64, 65]; 2—expansion of the list of seed sentiment words using a knowledge base (e.g., WordNet) or a corpus of opinions (statistical approach) and other linguistics resources [49, 57, 66, 67]. In this paper, a combination of both approaches is used by creating a complete WordNet for Persian and establishing links between its concepts and the English WordNet ones.

Persian WordNet

Princeton WordNet (PWN) is an electronic lexical database for the English language. It is comprised of a natural language vocabulary in the form of synonymous sets (synsets), which are classified into categories according to their parts of speech, such as verbs, nouns, adjectives, and adverbs. These synsets are connected to each other by semantic relations, such as synonymy, antonymy, hypernymy, hyponymy, and meronymy. The latest version of PWN⁵ contains approximately 155,327 words, which were organized into 117,597 synsets. WordNet has recently been used in some papers to extract sentiment words and features [31, 44]. Researchers have made

⁵ WordNet 3.1 database statistics

attempts to automatically or semi-automatically construct a Persian WordNet [68–73]. However, FarsNet [71] and PersianWN [70] are the only published Persian WordNets which are available to use.

FarsNet FarsNet (the first Persian WordNet) is a lexical database that contains information on words and language combinations (concepts), their syntactic information (POS), and the semantic relations between them. The latest version of this database (FarsNet version 2.0) is available to researchers.⁶ The EuroWordNet concepts were used to create the WordNet in this study. That is, the initial core of Persian WordNet was first produced manually and then it was completed in a top-down process using a semi-supervised method. The initial core of FarsNet was developed with the help of translating BalkaNet concepts and some common Persian concepts. It was then completed using a semi-supervised method, various Persian resources, and bilingual resources (Persian-English) [71].

Persian WordNet of Tehran University (PersianWN) The latest version of Persian WordNet, provided by Tehran University [70], is available on the multilingual WordNet website.⁷ It was created by running an unsupervised Expectation Maximization (EM) algorithm and implementing a text corpus and English WordNet (PWN). The FarsNet version 1.0 is used to calculate the primary probability of each word in each synset. Then, an iterative method of EM is employed to maximize the probability of each word.

Constructing a Comprehensive Persian WordNet (FerdowsNet)

As shown in Table 7, the current Persian WordNets contain an insufficient number of synsets. Also, the synsets have a small overlap with PWN in English. Furthermore, the semantic relations between synsets in the Persian WordNets are fewer than those in the PWN. The inadequacy of the current Persian WordNets called for the development of a new Persian WordNet (FerdowsNet), which covers most synsets and semantic relations in the PWN.

To construct FerdowsNet, the following language resources and knowledge bases were implemented:

- Princeton WordNet
- Various bilingual dictionaries (English-Persian)
- Google Text Translator
- Wikipedia, the Encyclopedia, and the Yago Ontology [74] to link Wikipedia and PWN

- Persian corpora (several newsgroups and Persian Wikipedia)
- Persian encyclopedias and dictionaries
- Pre-existing Persian WordNets.

The construction of FerdowsNet consists of nine steps (Fig. 2):

- Step 1: All synset words are translated by different bilingual dictionaries.
- Step 2: A bipartite graph is formed, in which there is a node on the left side (X_i) for each English word (synset words) and a node on the right side (Y_i) for each Persian word (list of translations). Then, each English word x_i is connected to its Persian translation y_j by an edge in the form of (x_i, y_j) between them. The weight of each edge depends on the number of occurrences of the words in the translation list (by various dictionaries) and their translation rating (in most available dictionaries, translated words are sorted according to their importance).
- Step 3: In this step, Persian words related to a synset are first extracted from Wikipedia knowledge bases and other Persian WordNets. Then, using the Yago ontology [74], concepts relevant to the synset words selected from Wikipedia and the equivalent of these concepts in Persian are extracted (if a Persian Wikipedia page for that concept is available). Then, using the links between FarsNet, Persian WN, and PWN, synset words equivalent to the corresponding English synsets are extracted (if any equivalent synset is available in these WordNets).
- Step 4: Using the words extracted from the previous step, translated words are extended (adding new words to the right of the bipartite graph) or the weight of the edges related to the translated words in the bipartite graph (formed in Step 2) is modified.
- Step 5: Using the Hungarian algorithm,⁸ the best match (the most proper Persian equivalent in English) is extracted from the formed weighted bipartite graph and the list of candidate words S for this synset is obtained. Then, Persian words that were not selected and whose relation weights (relation with one of the English words) are more than the average weight of the selected ones are chosen as the second candidate translation.
- Step 6: Gloss and the example of each synset are translated using Google Translator.
- Step 7: The required preprocessing of the translated text is performed and the keywords are extracted.

⁶ <http://dadegan.ir/catalog/farsnet>

⁷ <http://compling.hss.ntu.edu.sg/omw/>

⁸ https://github.com/KevinStern/software-and-algorithms/blob/master/src/main/java/blogspot/software_and_algorithms/stern_library/optimization/HungarianAlgorithm.java

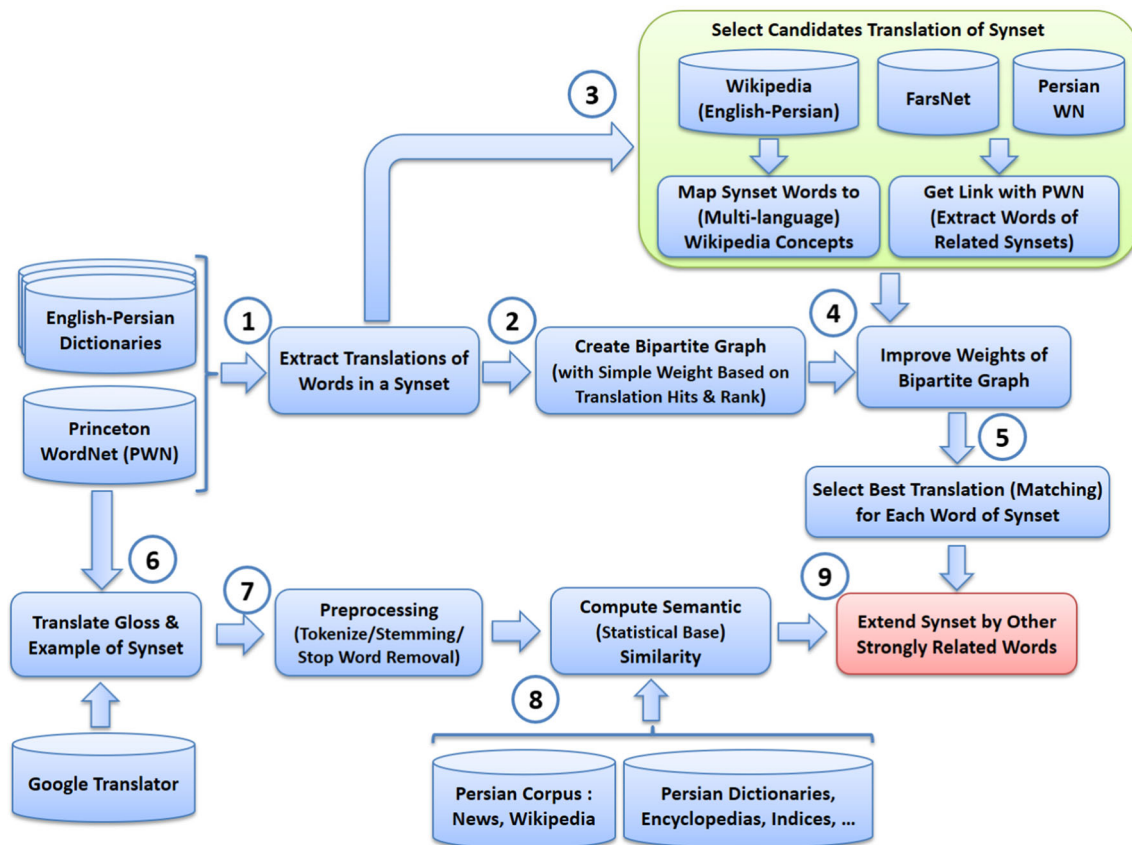


Fig. 2 The system architecture for construction of the synsets of FerdowsNet

Step 8: The synonymous and equivalent words (using the PMI⁹ measure [2, 75]) with the words selected from Step 5 (word set S) are extracted using Persian corpora (Hamshahri online newspaper, Alef and Tabnak [76–78] and the contents of Persian Wikipedia pages), and available Persian dictionaries and encyclopedias.

Step 9: The words extracted in the previous step, whose similarities with the words of S are more than a certain threshold, are added to the list of final words.

After translating and extending the synsets, the relations between synsets in PWN are used for the relations between FerdowsNet synsets.

Construction of Persian Sentiment Lexicon

As mentioned, this paper employs two methods to construct the sentiment lexicon. In the first method, after construction of FerdowsNet and establishment of links to PWN, the polarity of each synset can be obtained using SentiWordNet. Thus, a

Persian sentiment lexicon can be constructed by translating SentiWordNet. However, this method yields two types of errors. The first occurs because of the disambiguation in the synset translation when constructing the Persian WordNet. Moreover, the specified polarities of synsets in SentiWordNet also contain errors, which further decreases the accuracy of the Persian sentiment lexicon.

To resolve this, in the second method (PersianSentiWordMiner or PSWM), the sentiment lexicon is extracted by a semi-supervised learning method (without the use of SentiWordNet).

Persian Sentiment WordNet (PSWN) To develop the Persian Sentiment WordNet, concepts (synsets) in Princeton WordNet are first mapped onto their equivalent synsets in FerdowsNet. Then, the calculated polarity for each synset in English SentiWordNet is mapped to its corresponding synsets in the Persian SentiWordNet (PSWN).

Persian Sentiment WordNet can be used as a comprehensive sentiment lexicon for Persian. The obtained Persian sentiment lexicon is derived from the sentiment words whose polarity is more than 0.5.¹⁰ Moreover, given the degree of confidence for the words in each synset in FerdowsNet, there will be confidence of

⁹ point-wise mutual information

¹⁰ If the total positive and negative sentiment polarity of a word is more than 0.5, the word is subjective.

Table 1 Positive (Pos#) and negative (Neg#) sentiment words in PS

		Noun	Adjective	Adverb	Verb	Total
No restriction	Pos#	11,198	9491	2334	3675	26,698
	Neg#	13,175	11,022	698	4173	29,068
Confidence above 0.8	Pos#	1749	1372	408	383	3912
Confidence above 0.8	Neg#	2100	1624	119	471	4314
Confidence above 0.8 and sentiment polarity above 0.5	Pos#	366	483	24	39	912
Confidence above 0.8 and sentiment polarity above 0.5	Neg#	553	750	15	87	1405

The italicized items are the best results obtained

correctness for each sentiment word in addition to positive and negative polarity. The number of sentiment words with different POS tags and confidence is shown in Table 1.

The effect of confidence on the precision and recall of FerdowsNet and its impact on quality of sentiment lexicon are demonstrated in the assessment (evaluation) section.

Persian Sentiment Word Miner (PSWM) The PSWM is an HMM-based sequence learning method employed to extract the sentiment lexicon after manually tagging some reviews. The methodology is similar to that of OpinionMiner [79]. After manually labeling opinions, the sequence learning approach based on HMM is applied to extract the sentiment words (rather than sentences as in OpinionMiner [79]).

In PSWM, some sentences, consisting mostly of adjectives and adverbs that are often used to express sentiment sentences or change the polarity, are labeled with special tags. For this purpose, in order to extract sentiment words, after manually tagging some reviews, the sequence learning HMM-based method is employed. Hence, a number of sentences in texts about a specific subject (digital products) are labeled with tags specified in Table 2, manually using the tagger tools provided for this purpose. Experts specified the polarity (rating) of sentiment words and the total polarity of each sentence (a number between -5 and 5). Finally, the words with tags other than the ones shown in Table 2 are labeled as <BG> (background word).

In Persian, negative forms of verbs, usually indicated by adding the prefix “نـ”/N/ to the beginning of the verb, are often used to reverse the polarity of a sentence. Negative verbs are

detected in the preprocessing phase by the lemmatizer and are automatically tagged as “Reverse.”

After tagging reviews, tagged sentences are used to develop the set of sentiment words according to the PSWM algorithm applied. Before implementing the learning method, a list of sentiment seed words is extracted from those tagged as sentiment words (positive or negative). In order to extract new sentiment expressions in the opinion corpus, a list of existing sentiment words by semantic relationships in FerdowsNet is then expanded (Fig. 3).

Next, the learning algorithm is executed to expand the list of sentiment words in a corpus of unlabeled reviews. The Viterbi approach [80] is used in order to implement the HMM learning method and select the best path with the maximum score. The purpose of the learning method is to find the most probable tag {<BG>, Pos, Neg, Reverse, Decrease, Intensity, Feature} for each word in a given sentence. Words with tags such as *Reverse*, *Decrease*, and *Intensity* are extracted from a predetermined list of words from the training corpus which was labeled manually. The list of words is assumed to be fixed in the current study, due to the limited number of these types of words in Persian. Thus, the challenge is to determine tags {*Feature*, *Neg*, *Pos*, <BG>} for each unlabeled word, which does not have *Intensity*, *Reverse*, or *Decrease* labels or any of the POS tags, such as *Number*, *Delim*, and *Prep* in the sentence. Details of this algorithm are presented below.

PSWM Algorithm

1. The synset related to the list of words with sentiment

Table 2 Sentimental tags and their descriptions

Tag	Meaning of label/tag	Example
<i>Pos</i>	Words with a positive polarity	Good/خوب/xu:b, suitable/مناسب/monaseb, excellent/عالی/ali:, resistant/مقاوم/mogavem
<i>Neg</i>	Words with a negative polarity	Bad/بد/bæd, ugly/زشت/zæft, unsuitable/نامناسب/namonaseb,
<i>Feature</i>	Various Entities or Their Features	Phone/گوشی/gu:ʃi, camera/دوربین/du:rbi:n, lens/لنز/lenz, screen/صفحه نمایش/sæfe næmajæf
<i>Intensification</i>	Words which increase the sentiment intensity	Extremely/بشدت/bedædæt, very/زیاد/besiar, much/بسیار/besi:ja:r, more/بیشتر/biʃtær,
<i>Reducer</i>	Words which can reduce sentiment intensity	Rare/کم/kæm, sometimes/گاهی/gahi, fairly/نیمه/ni:me, quite/نسبتاً/nesbætæn
<i>Reverse</i>	Words which reverse the sentiment of sentences	Not/نه/næ, lacking/فایده/faqæd, loss/عدم/ædæm, miss/نبودند/naebu:d

- polarity tags (positive and negative) is extracted from FerdowsNet. The list of sentiment words is then extended by those related to the selected synset in FerdowsNet.
- The HMM learning method is trained using tagged sentences by reviewers or in the previous iteration of the algorithm.
 - Part of the unlabeled opinion texts is randomly selected from the review corpus.
 - The words of an unlabeled review sentence are POS tagged.
 - A set of tags {Feature, Intensity, Decrease, Reverse, Neg, Pos} from the initial SentiWordNet and the current sentiment lexicon for words, extracted from the previous iteration of the algorithm, are considered and if there is a word with a corresponding tag it is labeled accordingly.
 - Other unlabeled words of the reviews are labeled by the HMM learning algorithm with *Feature*, *Neg*, *Pos*, and *<BG>* tags.
 - If there are new sentiment words among the labels, the list of sentiment words will be updated and the algorithm will return to the first stage. Otherwise, the algorithm will terminate.

Feature Engineering

Generation and selection of relevant features of a dataset play vital roles in improving the quality and efficiency of machine learning methods. Feature selection is especially essential in high-dimensional data, such as text, gene expression data, image, and audio video. [81, 82]. The objective of the feature selection is to extract a set of relevant features from natural language texts, before they are sent to the sentiment classification methods. In general, feature selection is performed with two objectives: (1) increasing the efficiency and speed of classification methods by reducing the size of data (number of

dimensions), it is particularly essential when using classification methods whose training phase has cost and time overheads or high memory usage (such as SVM) and (2) enhancing the generalization by reducing the redundant or irrelevant features. The irrelevant input features may lead to overfitting [83].

Features

In the current paper, the list of features applied to classify sentiments includes the following:

N-Gram Features This feature has been the baseline in most related research [6, 15]. In this paper, various unigrams, bigrams, and trigrams are applied as feature sets. In order to reduce the feature space an informal to formal word converter, normalizer and stemmer were developed. Feature space is pruned by a minimum n-gram occurrence empirically set to the value of five.

TFIDF-Based Word Weighting Instead of using merely word presence, a variety of Delta-TFIDF-based versions [84] have been implemented and tested for the purpose of weighting such as Augmented TF, LogAve TF, BM25 TF, BM25 + TF, Delta smoothed IDF, and DeltaProb. IDF, Delta smoothed Prob. IDF, and Delta BM25 IDF. In Delta-TFIDF, a simple TFIDF weighting method is applied separately for each class (positive and negative) [85].

Character N-Gram Features Similar to n-Gram features of words, N-Gram features of characters are also applied according to the [86] approach. Different characters of 3-Grams to 6-Grams, available in the opinion texts are applied as a feature set. The minimum occurrence of a particular character n-gram is set to five, according to the corpus size, in order to prune the feature space.

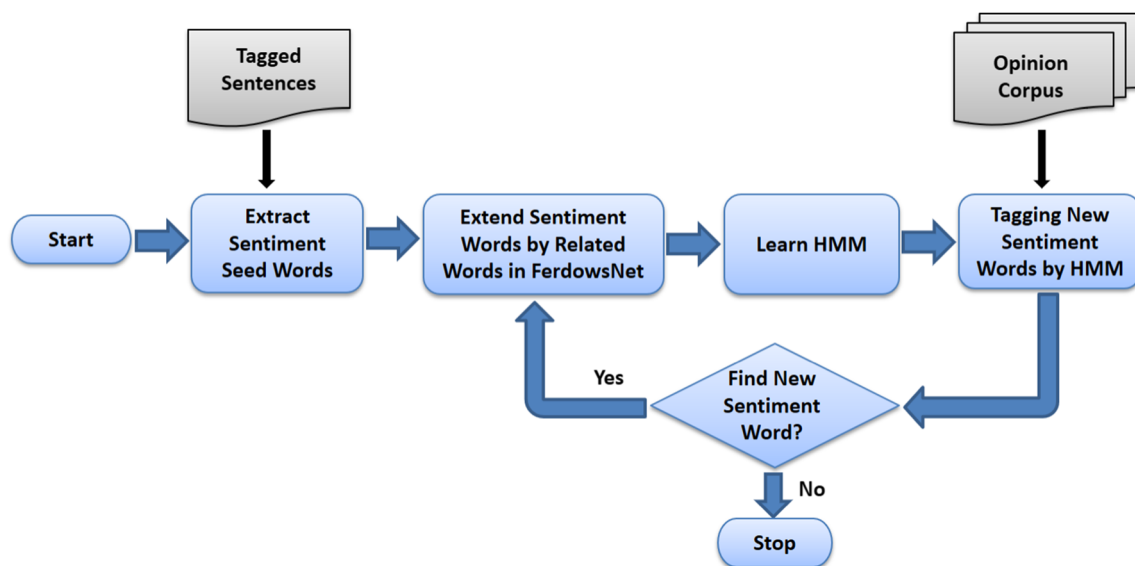


Fig. 3 Sentiment lexicon generation (PSWM) algorithm

Table 3 Common bi-tagged patterns to express sentiments in Persian

POS/senti tag pattern	Example
N + Adj	good book/کتاب خوب/ketabe xu:b
Adj + V	is great/عالی است/ali: æst
Reverse + (feature pos neg)	worthless/فاقد ارزش/faged æræzef lack of/امکانات نبود/næbu:d emkanat no/عدم وجود/ædæm vødʒu:d very good/خوب خیلی/xeili: xu:b really stylish/واقعا شیک/vageæn
(Booster Reducer) + (pos neg)	fairly complete/تقریباً کامل/tægri:bæn high quality/بالا کیفیت/keifi:æt bala
(pos neg) + (Booster Reducer)	low speed/کم سرعت/søræte kæm not bad/بد نیست/bæd ni:st
neg + reverse reverse + neg	no fault/بدون نقص/bedu:ne nægs not interesting/نبود جالب/dʒaleb næbu:d
(pos Feature) + reverse reverse + (pos Feature)	no USB 3.0 port/عدم وجود درگاه USB 3.0/vødʒu:d dærgah
(pos neg) + (pos neg)	comfortable and pleasant/راحت و دلچسب/rahæt væ deltʃæsb
(pos neg) + V	is beautiful/زیبا است/zi:ba æst

Part-of-Speech (POS) Tag Feature Given the various meanings and usages of words in different parts of speech, the POS tag feature is used to classify sentiments in most related studies. In this paper, only words with noun, verb, adjective, and adverb tags are used. Additionally, the number of occurrences of each POS tag is considered (similar to [87]).

Sentiment Words Features The extracted sentiment word lexicon is one of the key features. The polarity of the sentiment words in each sentence is calculated. However, the polarity of sentiment phrases may change after analyzing the sentiment reverse or intensifier tags. In order to calculate the overall sentiment of a review, the polarities of different sentences are averaged. Features related to words with roles of intensifying, reducing, and reversing the sentiment are also considered in this calculation. These words directly affect the polarity or sentiment intensity. Therefore, it is necessary to apply them in sentiment analyses. In this study, elongated words and repeated punctuation are also treated as features, similar to relevant state-of-the-art features used to analyze sentiments on Twitter (like STATE- features) [87, 88]. In this approach, the collection of sentiment features is obtained from the semi-supervised method (PSWM).

Bi-Tagged Feature This feature has been proposed by [2] to extract the relevant features for expressing sentiments in English. These features include pre-defined patterns of common collocations to express a sentiment. They have been employed in most related works to classify sentiments [89, 90]. In the present study, a set of common Bi-tags are used to express a sentiment in Persian. A list of these patterns, along with some examples, are compiled in Table 3.

SWN Subjectivity Scores (SWNSS) The SWNSS method [91] uses the weights assigned to the words in SentiWordNet (SWN) to calculate their subjectivity. Considering the specified threshold, objective words (unigram features) and the words that do not exist in SWN are removed from opinion texts. In the current study, the sentiment threshold of sentiment phrases should be 0.22 based on the conducted experiments. Moreover, the SWNPD (SWN Proportional Difference) method was proposed which applies positive or negative polarity in SentiWordNet [91]. Similar to the Proportional Difference method, the polar words (negative or positive) are selected and others are removed from the features. However, it was shown that the SWNPD method is less efficient compared to SWNSS. Thus, only the SWNSS method for using the created Persian Sentiment WordNet (PSWN) is used.

Word2vec Cluster N-Grams (W2VC) Similar to the methodology applied in [92], the words of the review corpus are reduced to 100-dimensional vectors using the Word2vec tool¹¹ [93]. The K-means clustering method is then used to cluster 100,000 words (within the input corpus) into 5000 clusters. These clusters are used to represent words (n-Grams).

Sentiment-Specific Word Embedding (SSWE) Tang et al. improved the Word2vec model to propose a new method for word representation [94].¹² It was shown that sentiment classification using SSWE for the conversion of sentiment features in a continuous space yields better results than other

¹¹ Available at <https://code.google.com/archive/p/word2vec/>

¹² Available at <https://github.com/attardi/deepnl/>

similar methods, such as Word2Vec_Skip-gram [93] and ReEmb [95]. They succeeded in increasing the accuracy of the sentiment classifier in the Coooolll system by combining the SSWE features and other common features (STATE features which were used at NRC [87]) for opinion mining.

Feature Selection

Various methods have been proposed to select features in text and sentiment classification [15, 96–98]. Feature extraction methods that transfer features into a new space with fewer dimensions have also been proposed and implemented. However, [6] has shown that the results of feature extraction methods for sentiment classification, such as Principal Component Analysis (PCA) and Singular Value decomposition (SVD) are less accurate than feature selection methods, such as Information Gain (IG). Thus, the current study uses only feature selection methods. In this section, a variety of the state-of-the-art supervised and unsupervised approaches for feature engineering are introduced.

Notations used in this section are represented in Table 4. c_w , $c_{\bar{w}}$, \bar{c}_w , and $\bar{c}_{\bar{w}}$, respectively, show the number of documents in class c that contain the word (feature) w ; the number of documents in class c that have no word w ; the number of documents in \bar{c} that contain word w ; and the number of documents in class \bar{c} that have no word w . n_c and $n_{\bar{c}}$ also represent the number of documents in c and \bar{c}

Table 4 Representation of notation

	Document belongs to class c	Document does not belong to class c	
Document contains word w	c_w	\bar{c}_w	$c_w + \bar{c}_w$
Document does not contain word w	$c_{\bar{w}}$	$\bar{c}_{\bar{w}}$	$c_{\bar{w}} + \bar{c}_{\bar{w}}$
	$n_c = c_w + c_{\bar{w}}$	$n_{\bar{c}} = \bar{c}_w + \bar{c}_{\bar{w}}$	N

classes, respectively. N (i.e., $n_c + n_{\bar{c}}$) is the total number of documents.

Mutual Information (MI) In information theory, the mutual information (MI) of two random variables is a measure of the mutual dependence between the two variables. The Mutual Information metric for calculating the probability of the feature (word) occurrence in the target class in proportion to the probability of its overall occurrence is calculated as follows [99]:

$$MI_w = \frac{c_w \times N}{(c_w + c_{\bar{w}})(c_w + \bar{c}_w)} \tag{1}$$

Information Gain (IG) The Information Gain metric specifies the number of necessary information bits to predict the category (class) in the presence or absence of each feature (word) in the text [100]. The Information Gain value of a feature is calculated as follows:

$$IG_w = -P(c)\log_2P(c) + P(\bar{c})\log_2P(\bar{c}) - (P(w)(-P(c_w)\log_2P(c_w)-P(\bar{c}_w)\log_2P(\bar{c}_w))) + (P(\bar{w})(-P(c_{\bar{w}})\log_2P(c_{\bar{w}})-P(\bar{c}_{\bar{w}})\log_2P(\bar{c}_{\bar{w}}))) \tag{2}$$

where

$$P(c_w) = \frac{c_w}{c_w + \bar{c}_w} \quad P(\bar{c}_w) = \frac{\bar{c}_w}{c_w + \bar{c}_w} \quad P(c_{\bar{w}}) = \frac{c_{\bar{w}}}{c_{\bar{w}} + \bar{c}_{\bar{w}}} \quad P(\bar{c}_{\bar{w}}) = \frac{\bar{c}_{\bar{w}}}{c_{\bar{w}} + \bar{c}_{\bar{w}}}$$

$$P(w) = \frac{c_w + \bar{c}_w}{N} \quad P(\bar{w}) = 1 - P(w) = \frac{c_{\bar{w}} + \bar{c}_{\bar{w}}}{N} \quad P(c) = \frac{n_c}{N} \quad P(\bar{c}) = \frac{n_{\bar{c}}}{N}$$

Chi-square (CHI) and Variants Chi-square (χ^2) is one of the common statistical metrics to calculate the independence

between a feature and a class and is used to select the superior features. Ng et al. [101] proposed a variant of χ^2 called the

Table 5 Features of Digikala review corpus

	Entire corpus	Tagged corpus
Type of products	10	10
Number of products	7572	270
Number of reviews	31,730	3080

NGL (Ng-Goh-Low) Coefficient. They demonstrated that the feature selection results of the NGL approach for text classification, in some cases, are better than χ^2 . Moreover, Galavotti et al. presented a simplified form of χ^2 called GSS (Galavotti-Sebastiani-Simi) coefficient [102]. They proposed that GSS can produce better results than NGL and χ^2 [102].

$$GSS_w = c_w \bar{c}_w - \bar{c}_w c_w \quad (3)$$

$$NGL_w = \frac{\sqrt{N} GSS_w}{\sqrt{(c_w + c_{\bar{w}})(\bar{c}_w + \bar{c}_{\bar{w}})(c_w + \bar{c}_w)(c_{\bar{w}} + \bar{c}_{\bar{w}})}} \quad (4)$$

$$\chi^2_w = (NGL_w)^2 \quad (5)$$

Relevancy Score (RS) and Odds Ratio (OR) These two metrics are recognized statistical methods of feature selection that have been shown, in some cases to yield better results in classifying texts than IG and MI [98, 103].

$$OR_w = \frac{c_w \bar{c}_w}{c_{\bar{w}} \bar{c}_w} \quad (6)$$

$$RS_w = \frac{c_w}{\bar{c}_w} \quad (7)$$

Document Frequency (DF) One of the popular methods of feature selection in text classification is to filter features according to the number of documents in the corpus which contain said features [6]. Features that occur in less than a certain number of texts are removed. The frequency of documents (DF) including a particular feature is calculated using the following equation:

$$DF_w = c_w + \bar{c}_w \quad (8)$$

Categorical Proportional Difference (CPD) In order to determine the impact of each feature (unigrams) in representing a class, the CPD method was proposed by [104]. The frequency of each feature in each class (positive or negative) is separately calculated and the polarized words, which occur dominantly in a class, have a higher PD value, while those words distributed equally in both classes have a lower PD value. The PD value for both positive and negative classes is calculated as follows:

$$CPD = \frac{|DF_+ - DF_-|}{DF_+ + DF_-} \quad (9)$$

In this paper, all mentioned feature selection methods (CPD, CHI, GSS, IG, MI, NGL, OR, RS and DF) have been applied to extract features from the Persian reviews. The best 2000 features with higher weight scores were selected to be used for sentiment classification. It was empirically proven that the selection of over 2000 features does not notably affect the quality of the sentiment classifier of Persian reviews.

Experimental Results

In this section, we introduce our dataset (review corpus) and then compare the quantitative and qualitative assessment of sentiment lexicons, in addition to various Persian WordNets. Finally, it is demonstrated how the sentiment classification of reviews is affected by the inclusion of several text processing tools, sentiment lexicons, and the latest methods of sentiment classification for different features.

Dataset

To assess the extracted sentiment words, a corpus of opinions on the Digikala website¹³ were collected by a web crawler. Similar to Amazon, Digikala is the fifth most visited website and market leader in e-commerce in the Middle East.¹⁴ Due to the high volume of users, Digikala has relatively rich comments. This dataset contains reviews about different products. Total opinions of the collected corpus consist of 31,730 reviews on ten different types of products. There are about 3080 reviews for training supervised machine learning algorithms and assessing them. These have been tagged by experts, but the rest have no sentiment tags. Some features of this dataset are listed in (Table 5).

There are three categories of reviews: “Expert Review,” “Reviews of Active Users,” and “Short Comments.” In Table 6, the features of each category and their differences are presented.

¹³ www.digikala.com

¹⁴ <http://www.alexa.com/topsites/countries/IR>

Table 6 A variety of opinions available on the corpus

	Style	Investigated features	Subjectivity rate	Average number of opinions for each product	Average length of text (number of characters)
Expert review (website administrator)	Formal	Many	Not many (objective)	1	8045
Reviews of active users	Semi-formal	Many	Average	3.79	1743
Short comments	Informal	Few	Many	9.12	356

Table 7 Quantitative assessment Persian WordNets

	English WordNet	FarsNet 2	Persian WN of Tehran University	FerdowsNet
Number of words (unique)	155,287	31,230	18,166	100,062
Number of synsets	117,659	20,432	17,759	91,640
Number of synsets linked to English WordNet (Princeton)	117,659	17,300	17,759	91,640
Average number of words in each synset	1.759	1.853	1.715	2.033
Percentage of single-word synsets	54%	56%	58%	60%
Average number of sharing each word among different synsets	1.514	1.16	1.57	1.88

Evaluation of FerdowsNet and PSWN

First, FerdowsNet is quantitatively assessed and compared with other Persian WordNets. In Table 7, the features of various Persian WordNets are shown.

To assess and compare FerdowsNet with other Persian WordNets, first about 1000 synsets were randomly selected from the English WordNet (250 synsets from each noun, verb, adjective and adverb category). Then, the equivalent words of these synsets in different Persian WordNets were extracted and their quality in the context of natural language processing were assessed by three experts.

In order to make a precise evaluation, the accuracy and recall metrics for each wordnet were separately calculated by experts. First for each English synset, a reference set of equivalent Persian words is considered as $S^* = \{s_1^*, s_2^*, s_3^*, \dots, s_n^*\}$. Then, the set of words available in the WordNet for each synset is defined as $S^{wn} = \{s_1^{wn}, s_2^{wn}, s_3^{wn}, \dots, s_n^{wn}\}$. Finally,

the accuracy and recall metrics for each synset was calculated and their average (micro) was considered as the final accuracy and recall in the WordNet using the following formulas:

$$Recall = \frac{|S^* \cap S^{wn}|}{|S^*|} \quad (10)$$

$$Precision = \frac{|S^* \cap S^{wn}|}{|S^{wn}|} \quad (11)$$

$$F_1\text{-Measure} = 2 \cdot \frac{precision \times recall}{precision + recall} \quad (12)$$

where $|S^* \cap S^{wn}|$ represents the number of correct words selected in the intended synset words, $|S^*|$ is the number of Persian words equivalent to the main concept, and $|S^{wn}|$ stands for the number of words in the intended Persian WordNet synset.

Table 8 The qualitative assessment of Persian wordnets

Persian WordNet	Percentage of synset coverage in English WordNet	Precision	Recall	F1-measure	Existing Recall ^a	Existing F1-measure ^a
FarsNet V2.0	15%	<i>0.898</i>	0.101	0.182	0.695	0.784
Persian WN of Tehran University	11%	0.797	0.057	0.106	0.531	0.637
FerdowsNet (conf ≥ 0.8)	45%	0.826	0.199	0.321	0.483	0.610
FerdowsNet(conf ≥ 0.4)	56%	0.801	0.310	0.447	0.627	0.703
FerdowsNet (all words)	82%	0.740	<i>0.603</i>	<i>0.665</i>	<i>0.901</i>	<i>0.813</i>

The italicized items are the best results obtained

^a The criterion is calculated for each synset existing in Persian WordNet and other synsets, whose equivalence does not exist in WordNet, have been ignored

Table 9 Qualitative assessment of the PSWN

	Accuracy	Recall	F1-measure
Per-Sent [63]	0.926	0.45	0.61
UTIIS [47]	0.877	0.653	0.75
Persian Sentiment Word Net (PSWN)	0.86	0.75	0.8

For a quality evaluation and comparison of FerdowsNet with other Persian WordNets, 1000 synsets of English WordNet (Princeton) were randomly selected. Then, the corresponding synsets in the Persian wordnets were extracted. Next, words and their Glosses, along with examples for each synset in English WordNet, were prepared in a list of Persian vocabulary equivalent to that synset. A single-blind trial was then conducted. To ensure fairness, Persian vocabulary list was presented in a way that the participating experts were not informed about which WordNet each word belonged to. Finally, in regard to natural language processing concepts and WordNet, the experts identified the wrong words of each synset. After the trial, the accuracy, recall, and F1-Measure rates for each synset were calculated, based on the tags applied by the experts. The results are shown in Table 8.

For the confidence level of the words in FerdowsNet synsets, a quality assessment for different confidence intervals has been calculated. In Table 7, conf represents the FerdowsNet confidence level. It is important to note that in Persian WordNets, there are no Persian synsets equivalent to some of the synsets available in English WordNet. For this purpose, in Table 7, the recall of these groups is considered once as zero. Then, the groups are left out when calculating the recall average. For example, out of a set of 1000 synonymous groups selected from the English WordNet for assessment, there are only 147 equivalent synsets (about 15%) in FarsNet. The accuracy of words in existing synsets is about 0.898 and the overall recall (from 1000 synsets) is 0.101. However, if the recall value is calculated only for the 147 synsets (synsets whose equivalents are available in FarsNet), the recall value of this WordNet is 0.695.

For quality evaluation of Persian Sentiment WordNet (PSWN), about 150 polar words were randomly extracted. Using the same technique, the accuracy and recall of PSWN are calculated and shown in Table 9.

Table 10 Feature set description

Feature Set	Features
FS1	Binary
FS2	TF-IDF based
FS3	Word n-grams
FS4	Char n-grams
FS5	W2VC
FS6	SSWE

During the calculation of the recall criterion, it was found that among the 150 sentiment phrases, only 113 existed in PSWN. Therefore, the recall value of the sentiment WordNet is about 0.75. However, most sentiment phrases not in the WordNets are related to informal language or spelling errors of sentiment words that are not corrected by the text pre-processing tools. Also, as the polarity of 97 out of 113 words in PSWN matches the sentiment determined by the experts, the accuracy is approximately 0.86. For the words that exist in several synsets, the average polarity is considered. Part of Persian Sentiment WordNet errors is related to the errors in FerdowsNet. In addition, the polarity specified for the words in SentiWordNet has some errors which are propagated into Persian Sentiment WordNet (PSWN) and, as a result, reduce its accuracy [105].

Sentiment Classification Results

In order to classify reviews, their texts will be first pre-processed with the normalizer, sentence splitter and tokenizer, stop word removal, lemmatizer, informal-to-formal converter and spell-check¹⁵ tools. To evaluate the superior feature set, a variety of classification approaches are initially assessed over various features. The F-Measure metric and tenfold cross-validation method are applied.

Finally, as shown in Table 10, four groups of features are extracted from the reviews. Due to the high number of Char n-Gram features, the number of features of this dataset are reduced to 10,000 using the CPD feature selection method. In the word n-grams approach (FS3), the POS tags of words are considered.

Different TFIDF Weighting Schemas were studied to apply various features using different classification methods. Table 11 presents the results of these widely used TFIDF-based approaches. Also, global weighting methods, such as IDF (Invert Document Frequency), GFIDF, Entropy, DeltaBM25Idf, DeltaSmoothIdf, DeltaSmoothProbIdf, and local weighting methods for words in each document were considered, such as TermFrequency (TF), LogAvg, Augnorm, and BM25Plus. The LogAvg–DeltaSmoothIdf technique and Linear SVM produced better results among the evaluated approaches.

The results of different classification methods are compared in Table 12. A KNN algorithm was used with K = 3 and log-distance-weighted nearest neighbors. Among the classification approaches, the Linear SVM method (using the LibLinear library [106])¹⁶ produces the best results.

¹⁵ The opinion corpus and Persian text-processing tools for non-commercial use are available on the website of Web Technology laboratory of Ferdowsi University (<http://wtlab.um.ac.ir>).

¹⁶ The SVM method with different non-linear kernel functions (Sigmoid, Polynomial, RBF) was also studied that compared to the Linear SVM method is less accurate.

Table 11 A comparison of the quality (F-Measures (%)) of TF-IDF variants for the sentiment classification of Persian text reviews

Local W.	Global W.	BayesNet	LibLinear	SMO	KNN	MaxEnt	RandomForest	Average
TF	Idf	82.45	83.97	86.08	82.25	78.37	84.18	82.883
	Gfldf	82.45	85.89	86.07	82.27	78.31	84.13	83.187
	Entropy	82.45	86.78	86.07	82.29	78.74	84.36	83.448
	DeltaBM25Idf	82.45	88.52	86.08	82.27	78.6	84.15	83.678
	DeltaSmoothldf	82.45	88.6	86.07	82.25	78.6	84.22	83.698
	DeltaSmoothProblfd	82.45	88.53	86.07	82.25	78.78	84.22	83.717
BM25Plus	Idf	82.91	84.06	86.14	81.83	78.76	82.93	82.772
	Gfldf	82.91	85.03	86.15	81.84	79.01	83.2	83.023
	Entropy	82.91	86.14	86.14	81.83	78.75	83.08	83.142
	DeltaBM25Idf	82.91	87.19	86.15	81.81	78.59	83.1	83.292
	DeltaSmoothldf	82.91	87.24	86.15	81.8	78.94	82.92	83.327
	DeltaSmoothProblfd	82.91	87.22	86.14	81.79	78.99	83.04	83.348
Augnorm	Idf	83.19	84.36	86.25	81.37	78.55	83.64	82.893
	Gfldf	83.19	86.12	86.26	81.28	78.34	83.29	83.08
	Entropy	83.19	87.08	86.25	81.34	78.17	83.43	83.243
	DeltaBM25Idf	83.19	89.28	86.25	81.23	78.15	83.52	83.603
	DeltaSmoothldf	83.19	89.42	86.25	81.14	78.12	83.39	83.585
	DeltaSmoothProblfd	83.19	89.27	86.25	81.55	78.39	83.61	83.71
LogAvg	Idf	82.45	86.1	86.7	81.68	79.8	84.21	83.49
	Gfldf	82.45	86.3	86.71	81.75	79.23	84.15	83.432
	Entropy	82.45	82.21	86.7	81.76	79.76	84.1	82.83
	DeltaBM25Idf	82.45	<i>91.57</i>	86.73	81.77	78.53	84.18	84.205
	DeltaSmoothldf	82.45	<i>91.57</i>	86.74	81.72	79.38	84.27	84.355
	DeltaSmoothProblfd	82.45	<i>91.57</i>	86.73	81.75	79.38	83.94	84.303
Average		82.75	87.251	86.297	81.784	78.76	83.719	83.427

The italicized items shows the best result obtained for each classifier

Then, analytical tests are performed to assess the preprocessing methods. The impact of different text pre-processing tools on the average F-Measure of sentiment classification for different feature sets is shown in Fig. 4.

As previously mentioned, the sentiment lexicon list was created using two different techniques. Similarly, two sets of sentiment features are extracted from the reviews: the SentiWords (PSW) feature set and the SWNSS feature set.

The Persian SentiWords (PSW) feature set contains the number of Bi-tagged pattern features, sentiment expressions

obtained from the PSWM method, as well as an intensifier, reverser, and reducer of the sentiment in the text. Contrarily, the SWNSS feature set is based on the sentiment words obtained from Persian Sentiment WordNet (PSWN).

To select the superior sentiment feature set, the sentiment classification results are compared with every sentiment feature set mentioned in Table 10 and combined with the sentiment feature sets (PSW and SWNSS) in Fig. 5. In this figure, the classification results of text reviews are compared with the sentiment feature sets of PSW and SWNSS.

Table 12 Results of various classification methods

F-Measure	BayesNet	LibLinear	SMO	KNN	MaxEnt	RandomForest
FS1	0.795	0.864	0.855	0.802	0.821	0.799
FS2	0.832	0.916	0.867	0.823	0.798	0.844
FS3	0.807	0.881	0.871	0.806	0.822	0.808
FS4	0.707	0.881	0.798	0.748	0.691	0.791
FS5	0.738	0.858	0.851	0.817	0.849	0.812
FS6	0.75	0.866	0.863	0.815	0.864	0.809
Average	0.772	0.878	0.851	0.802	0.808	0.811

Fig. 4 The effect of text preprocessing tools on Persian sentiment classification

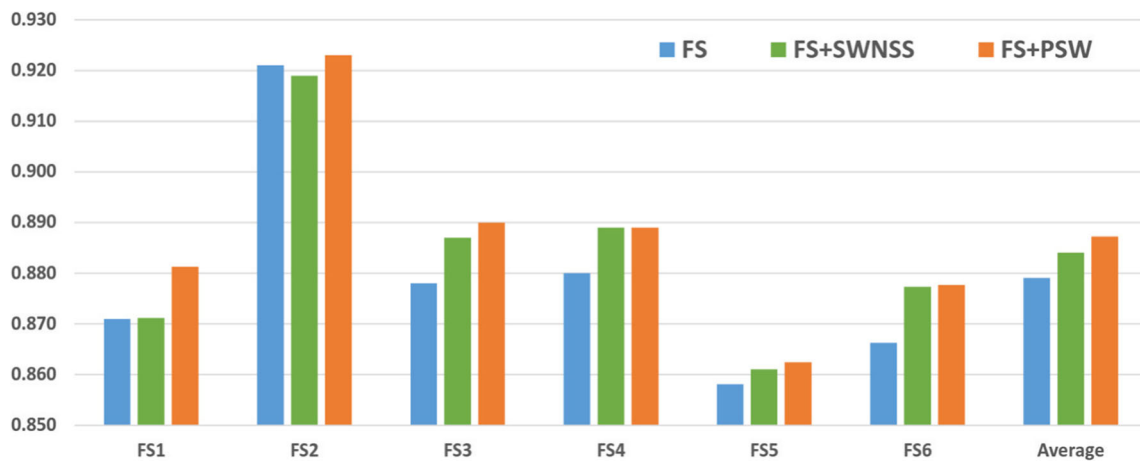
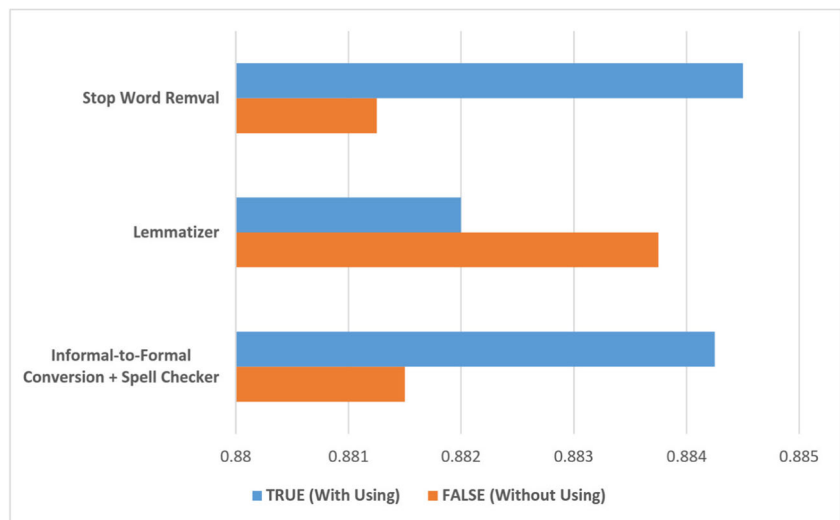


Fig. 5 The comparison of PSW and SWNSS method

Table 13 Impact of the sentiment feature set (PSW) on the quality (F-Measure) of LibLinear classification

F-Measure	FS1	FS1 + PSW	FS2	FS2 + PSW	FS3	FS3 + PSW	FS4	FS4 + PSW	Avg(FS)	Avg(FS + PSW)
All Features	0.871	0.878	0.921	0.923	0.878	0.884	0.88	0.889	0.8875	0.8935
CPD	0.789	0.859	0.691	0.847	0.765	0.866	0.772	0.873	0.7543	0.8613
CHI	0.825	0.846	0.775	0.814	0.815	0.839	0.828	0.843	0.8108	0.8355
GSS	0.809	0.85	0.78	0.816	0.808	0.847	0.81	0.834	0.8018	0.8368
IG	0.762	0.812	0.819	0.844	0.622	0.706	0.613	0.696	0.704	0.7645
MI	0.691	0.85	0.691	0.847	0.691	0.869	0.831	0.87	0.726	0.859
NGL	0.799	0.83	0.754	0.808	0.785	0.822	0.82	0.832	0.7895	0.823
OR	0.691	0.851	0.716	0.825	0.691	0.875	0.691	0.691	0.6973	0.8105
RS	0.856	0.871	0.742	0.795	0.849	0.86	0.828	0.845	0.8188	0.8428
DF	0.831	0.872	0.75	0.812	0.847	0.877	0.808	0.849	0.809	0.8525
Avg	0.7924	0.8519	0.7639	0.8331	0.7751	0.8445	0.7881	0.8222	0.7799	0.8379

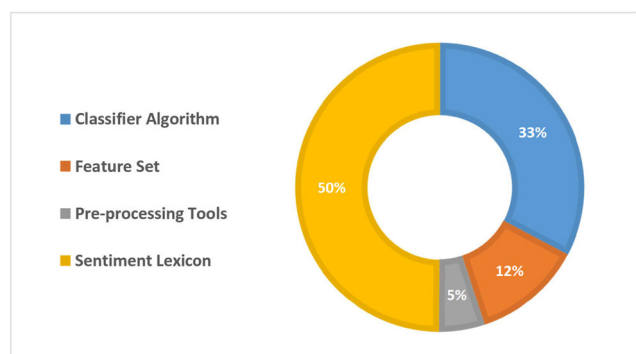


Fig. 6 Impact of different step of sentiment polarity classification in Persian reviews

For extracting sentiment words, the results demonstrate that the semi-supervised learning method PSWM used in the PSW feature set is more efficient than the one based on the Persian sentiment WordNet PSWN. The reason is that the polar words of PSW for the target domain were extracted from reviews related to commercial products, but SWNSS (Persian Sentiment WordNet) is general and domain-independent.

Different feature selection methods for extracting 2000 superior features on the selected feature set were implemented and their results are presented in Table 13. In order to select non-binary features (such as TF-IDF feature set), features were converted to binary values using the Maximum Gini-Index method, similar to the one applied in binary decision trees.

As shown in Table 13, reducing the dimensions via feature selection methods, despite significantly improving the performance of the classification algorithms, often decreases the quality of sentiment classification. Feature selection methods like CPD, RS, and MI using sentiment features and CHI and RS methods not using sentiment features are more efficient than other feature selection methods in classifying product review corpora in Persian.

The impact of each step of sentiment classification process is shown in Fig. 6. Our final analysis revealed that the sentiment lexicon feature (PSW) has the most impact on the sentiment classification results. Impact rate of a classifier algorithm and feature set was calculated by the difference between the best state and second best state.

Conclusion

In this paper, the impact of Persian NLP tools for preprocessing and sentiment classification was thoroughly examined. After developing the tools, a comprehensive Persian WordNet (FerdowsNet), and a corpus of Persian reviews, a new method (PSWN) of using English SentiWordNet was proposed for generating a Persian SentiWordNet. Moreover, the SentiWords lexicon was extracted using a semi-supervised PSWM method.

An in-depth analysis of different supervised machine learning methods was conducted to analyze the sentiments in commercial product data in Persian. In addition, a detailed assessment was done on a set of common state-of-the-art features and various methods of feature selection and classification, and the impact of different pre-processing tools on different feature sets was studied. Finally, informal-to-formal conversion and stop word removal for LogAvgTF-DeltaSmoothIDF with sentiment features and applying the Linear SVM classification method proved to produce the best results for sentiment classification of commercial products in Persian.

In addition to the development of the WordNet, the appropriate sentiment patterns, and the corpus of opinion mining for Persian language, the result of the current study could be used as a solid basis for selecting and using features, feature selection methods, and various classification approaches. The findings and developments made in this study could prove useful in the advancement of opinion mining research in Persian and other similar languages, such as Urdu and Arabic.

We are currently extending our research to aspect-based sentiment analysis, and preliminary results are encouraging. Our ultimate objective is to apply semantics in the sentiment analysis of comments by developing the opinion ontology. Therefore, a semantic framework as an integrated method will be used in all stages of aspect-based sentiment analysis.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed Consent In this paper, informed consent was not needed. We do not use any private or personal information in this research study.

References

1. Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: from unimodal analysis to multimodal fusion. *Inf Fusion*. 2017;37:98–125.
2. Turney, P.D. 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*. 2002. Assoc Comput Linguist
3. Recupero DR, Presutti V, Consoli S, Gangemi A, Nuzzolese AG. Sentilo: frame-based sentiment analysis. *Cogn Comput*. 2015;7(2):211–25.
4. Pang, B. and L. Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. 2005, Association for Computational Linguistics. p. 115–124.
5. Tang D, Qin B, Wei F, Dong L, Liu T, Zhou M. A joint segmentation and classification framework for sentence level sentiment

- classification. *IEEE/ACM Trans Audio Speech Lang Process.* 2015;23(11):1750–61.
6. Agarwal B, Mittal N. *Prominent feature extraction for sentiment analysis.* Berlin: Springer International Publishing; 2016.
 7. Liu B. *Sentiment analysis. Mining opinions, sentiments, and emotions.* Cambridge University Press; 2015.
 8. Cambria E, Rajagopal D, Olsher D, Das D. Big social data analysis. *Big Data Comput.* 2013;2013:401–14.
 9. Wang Q-F, Cambria E, Liu C-L, Hussain A. Common sense knowledge for handwritten chinese text recognition. *Cogn Comput.* 2013;5(2):234–42.
 10. Cambria E, Mazzocco T, Hussain A. Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining. *Biologically Inspired Cogn Architectures.* 2013;4:41–53.
 11. Zheng L, Wang H, Gao S. Sentimental feature selection for sentiment analysis of Chinese online reviews. *Int J Mach Learn Cybern.* 2015;1:1–10.
 12. Liao C, Feng C, Yang S, Huang H. Topic-related Chinese message sentiment analysis. *Neurocomputing.* 2016;210:237–46.
 13. Aldayel HK, Azmi AM. Arabic tweets sentiment analysis—a hybrid scheme. *J Inf Sci.* 2015;42(6):782–97.
 14. Vilares D, Alonso MA, Gómez-Rodríguez C. A syntactic approach for opinion mining on Spanish reviews. *Nat Lang Eng.* 2015;21(01):139–63.
 15. Habernal I, Ptáček T, Steinberger J. Reprint of “Supervised sentiment analysis in Czech social media”. *Inf Process Manag.* 2015;51(4):532–46.
 16. Dashtipour K, Poria S, Hussain A, Cambria E, Hawalah AY, Gelbukh A, et al. Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cogn Comput.* 2016:1–15.
 17. Balahur A, Perea-Ortega JM. Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Inf Process Manag.* 2015;51(4):547–56.
 18. Zhang, P., S. Wang, and D. Li. Cross-lingual sentiment classification: similarity discovery plus training data adjustment. *Knowl-Based Syst.* 2016.
 19. Guo, H., H. Zhu, Z. Guo, X. Zhang, and Z. Su. OpinionIt: a text mining system for cross-lingual opinion analysis. In *Proceedings of the 19th ACM international conference on Information and knowledge management.* 2010. ACM.
 20. Gao D, Wei F, Li W, Liu X, Zhou M. Cross-lingual sentiment lexicon learning with bilingual word graph label propagation. *Comput Linguist.* 2015;41(1):21–40.
 21. Balahur A, Turchi M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Comput Speech Lang.* 2013;28(1):56–75.
 22. Banea C, Mihalcea R, Wiebe J. Porting multilingual subjectivity resources across languages. *IEEE Trans Affect Comput.* 2013;4(2)
 23. Martín-Valdivia M-T, Martínez-Cámara E, Perea-Ortega J-M, Ureña-López LA. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Syst Appl.* 2013;40(10):3934–42.
 24. Duwairi R, El-Orfali M. A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. *J Inf Sci.* 2014;40(4):501–13.
 25. Prusa, J.D., T.M. Khoshgoftaar, and D.J. Dittman. Impact of feature selection techniques for tweet sentiment classification. In *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference.* 2015.
 26. Uysal AK, Gunal S. The impact of preprocessing on text classification. *Inf Process Manag.* 2014;50(1):104–12.
 27. Shamsfard, M. Challenges and open problems in Persian text processing. In *5th Language & Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics.* Poznań, Poland; 2011. p. 65–69.
 28. Feely, W., M. Manshadi, R. Frederking, and L. Levin. The CMU METAL Farsi NLP Approach. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* 2014.
 29. Hung C, Chen S-J. Word sense disambiguation based sentiment lexicons for sentiment classification. *Knowl-Based Syst.* 2016;110:224–32.
 30. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. *Comput Linguist.* 2011;37(2):267–307.
 31. Montejo-Ráez A, Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA. Ranked wordnet graph for sentiment polarity classification in twitter. *Comput Speech Lang.* 2014;28(1):93–107.
 32. Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. *Cogn Comput.* 2015;7(4):487–99.
 33. Poria S, Cambria E, Winterstein G, Huang G-B. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowl-Based Syst.* 2014;69:45–63.
 34. Dong, L., F. Wei, S. Liu, M. Zhou, and K. Xu. A statistical parsing framework for sentiment classification. *Comput Linguist.* 2015.
 35. Oliveira N, Cortez P, Areal N. Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decis Support Syst.* 2016;85:62–73.
 36. Ofek N, Poria S, Rokach L, Cambria E, Hussain A, Shabtai A. Unsupervised commonsense knowledge enrichment for domain-specific sentiment analysis. *Cogn Comput.* 2016;8(3):467–77.
 37. Wang G, Zhang Z, Sun J, Yang S, Larson CA. POS-RS: a random Subspace method for sentiment classification based on part-of-speech analysis. *Inf Process Manag.* 2015;51(4):458–79.
 38. Liu, B. and L. Zhang, A survey of opinion mining and sentiment analysis, in *Mining Text Data.* 2012, Springer. p. 415–463.
 39. Boiy E, Moens M-F. A machine learning approach to sentiment analysis in multilingual Web texts. *Inf Retr.* 2009;12(5):526–58.
 40. Cambria E. Affective computing and sentiment analysis. *IEEE Intell Syst.* 2016;31(2):102–7.
 41. Appel O, Chiclana F, Carter J, Fujita H. A hybrid approach to the sentiment analysis problem at the sentence level. *Knowl-Based Syst.* 2016;108:110–24.
 42. Catal C, Nangir M. A sentiment classification model based on multiple classifiers. *Appl Soft Comput.* 2017;50:135–41.
 43. Rushdi Saleh M, Martín-Valdivia MT, Montejo-Ráez A, Ureña-López L. Experiments with SVM to classify opinions in different domains. *Expert Syst Appl.* 2011;38(12):14799–804.
 44. Esuli, A. and F. Sebastiani, Pageranking wordnet synsets: an application to opinion mining, in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL).* 2007: Prague, Czech Republic. p. 442–431.
 45. Hassan, A. and D. Radev. Identifying text polarity using random walks. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics.* 2010. Assoc Comput Linguist
 46. Hassan, A., A. Abu-Jbara, R. Jha, and D. Radev. Identifying the semantic orientation of foreign words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2.* 2011. Assoc Comput Linguist
 47. Dehdarbehbahani I, Shakery A, Faili H. Semi-supervised word polarity identification in resource-lean languages. *Neural Netw.* 2014;58:50–9.
 48. Dehkharghani R, Saygin Y, Yanikoglu B, Oflazer K. SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Lang Resour Eval.* 2016;50(3):667–85.
 49. Baccianella, S., A. Esuli, and F. Sebastiani. *SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.* In *LREC.* 2010.

50. Esuli, A. and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. in Proceedings of 5th International Conference on Language Resources and Evaluation (LREC). 2006. Genoa: Citeseer.
51. Strapparava, C. and A. Valitutti. WordNet Affect: an Affective Extension of WordNet. in Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC). 2004.
52. Neviarouskaya A, Prendinger H, Ishizuka M. SentiFul: a lexicon for sentiment analysis. *IEEE Trans Affect Comput.* 2011;2(1):22–36.
53. Cambria, E., R. Speer, C. Havasi, and A. Hussain. *SenticNet: A publicly available semantic resource for opinion mining.* in AAAI fall symposium: commonsense knowledge. 2010.
54. Cambria, E., S. Poria, R. Bajpai, and B.W. Schuller. SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives. in Proceedings of the 26th International Conference Computational Linguistics (COLING). 2016. Osaka.
55. Pandarachalil R, Sendhilkumar S, Mahalakshmi G. Twitter sentiment analysis for large-scale data: an unsupervised approach. *Cogn Comput.* 2015;7(2):254–62.
56. Denecke, K. Using sentiwordnet for multilingual sentiment analysis. in Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on. 2008. IEEE.
57. Cruz FL, Troyano JA, Pontes B, Ortega FJ. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Syst Appl.* 2014;41(13):5984–94.
58. Basiri ME, Naghsh-Nilchi AR, Ghassem-Aghaee N. A framework for sentiment analysis in Persian. *Open Trans Inf Process.* 2014;1(3):1–14.
59. Amiri, F., S. Scerri, and M.H. Khodashahi. Lexicon-based sentiment analysis for Persian Text. in Recent Advances in Natural Language Processing. 2015.
60. Shams, M., A. Shakery, and H. Faili. A non-parametric LDA-based induction method for sentiment analysis. in Proceeding of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP). 2012. IEEE.
61. Ali-Mardani S, Aghaie A. Desinging supervised method for opinion mining in the Persian using lexicon and SVM (In Persian). *J Inf Technol Manag.* 2015;7(2):345–62.
62. Cerini, S., V. Compagnoni, A. Demontis, M. Formentelli, and G. Gandini. Micro-WNOp: a gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Language resources and linguistic theory: Typology, second language acquisition, English linguistics, 2007:* p. 200–210.
63. Dashtipour, K., A. Hussain, Q. Zhou, A. Gelbukh, A.Y. Hawalah, and E. Cambria. PerSent: a freely available Persian sentiment lexicon. in Proceedings of the 8th International Conference Advances in Brain Inspired Cognitive Systems, BICS 2016, Beijing, China. 2016. Springer.
64. Steinberger J, Ebrahim M, Ehrmann M, Hurriyetoglu A, Kabadjov M, Lenkova P, et al. Creating sentiment dictionaries via triangulation. *Decis Support Syst.* 2012;53(4):689–94.
65. Özsert, C.M. and A. Özgür. Word polarity detection using a multilingual approach, in computational linguistics and intelligent text processing. 2013, Springer. p. 75–82.
66. Chen, Y. and S. Skiena. Building sentiment lexicons for all major languages. in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). 2014.
67. Mahyoub FH, Siddiqui MA, Dahab MY. Building an Arabic sentiment lexicon using semi-supervised learning. *J King Saud Univ Comput Inf Sci.* 2014;26(4):417–24.
68. Famian A, Aghajaney D. Towards building a WordNet for Persian adjectives. *Int J Lexicogr.* 2000;2006:307–8.
69. Keyvan, F., H. Borjian, M. Kasheff, and C. Fellbaum. Developing persianet: the persian wordnet. in 3rd Global wordnet conference. 2007.
70. Montazery, M. and H. Faili. Automatic Persian wordnet construction. in Proceedings of the 23rd International Conference on Computational Linguistics: Posters. 2010. Assoc Comput Linguist.
71. Shamsfard, M., A. Hesabi, H. Fadaei, N. Mansoory, A. Famian, S. Bagherbeigi, E. Fekri, M. Monshizadeh, and S.M. Assi. Semi automatic development of farsnet; the persian wordnet. in Proceedings of 5th Global WordNet Conference, Mumbai, India. 2010.
72. Fadaee, M., H. Ghader, H. Faili, and A. Shakery. Automatic WordNet construction using Markov Chain Monte Carlo. *Polibits,* 2013(47): p. 13–22.
73. Taghizadeh N, Faili H. Automatic Wordnet development for low-resource languages using cross-lingual WSD. *J Artif Intell Res.* 2016;56:61–87.
74. Mahdisoltani, F., J. Biega, and F. Suchanek. *YAGO3: a knowledge base from multilingual Wikipedias.* in 7th Biennial Conference on Innovative Data Systems Research. 2014. CIDR 2015.
75. Turney, P. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. in 12th European Conference on Machine Learning (ECML 2001), Freiburg, Germany 2001.
76. AleAhmad A, Amiri H, Darrudi E, Rahgozar M, Oroumchian F. Hamshahri: a standard Persian text collection. *Knowl-Based Syst.* 2009;22(5):382–7.
77. Eghbalzadeh, H., B. Hosseini, S. Khadivi, and A. Khodabakhsh. Persica: a Persian corpus for multi-purpose text mining and natural language processing. in Telecommunications (IST), 2012 Sixth International Symposium on. 2012. IEEE.
78. Balali, A., A. Rajabi, S. Ghassemi, M. Asadpour, and H. Faili. Content diffusion prediction in social networks. in *5th Conference on Information and Knowledge Technology (IKT)*. 2013.
79. Jin, W., H.H. Ho, and R.K. Srihari. OpinionMiner: a novel machine learning system for web opinion mining and extraction. in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009. ACM.
80. Collins, M. Discriminative training methods for hidden markov models: theory and experiments with perceptron algorithms. in Proceedings of the ACL-02 conference on Empirical methods in natural language processing. 2002. Assoc Comput Linguist.
81. Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C, A. D.N. Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage.* 2012;60(1):59–70.
82. Tang, J., S. Alelyani, and H. Liu. Feature selection for classification: a review. *Data Classification: Algorithms and Applications,* 2014.
83. Bermingham ML, Pong-Wong R, Spiliopoulou A, Hayward C, Rudan I, Campbell H, et al. Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci Rep.* 2015;5:10312.
84. Paltoglou, G. and M. Thelwall. A study of information retrieval weighting schemes for sentiment analysis. in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. 2010. Assoc Comput Linguist.
85. Martineau, J. and T. Finin. Delta TFIDF: an improved feature space for sentiment analysis, in Proceedings of the Third International ICWSM Conference. 2009. p. 106.
86. Blamey, B., T. Crick, and G. Oatley, RU(-) or:-(? character-vs. word-gram feature selection for sentiment classification of OSN corpora, in Research and Development in Intelligent Systems XXIX. 2012, Springer. p. 207–212.
87. Mohammad, S.M., S. Kiritchenko, and X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets, in 7th International Workshop on Semantic Evaluation (SemEval 2013). 2013. p. 321–327.
88. Zhu, X., S. Kiritchenko, and S.M. Mohammad. *Nrc-canada-2014: Recent improvements in the sentiment analysis of tweets.* in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014.

89. Jain AK, Pandey Y. Analysis and implementation of sentiment classification using Lexical POS markers. *Int J Comput Commun Netw*. 2013;2(1):36–40.
90. Agarwal B, Mittal N. Semantic feature clustering for sentiment analysis of English reviews. *IETE J Res*. 2014;60(6):414–22.
91. O’Keefe, T. and I. Koprinska. Feature selection and weighting methods in sentiment analysis. in *Proceedings of the 14th Australasian document computing symposium, Sydney*. 2009. Citeseer.
92. Dong, L., F. Wei, Y. Yin, M. Zhou, and K. Xu, Splusplus: a feature-rich two-stage classifier for sentiment analysis of tweets. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015: p. 515–519.
93. Mikolov, T., I. Sutskever, K. Chen, G.S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. in *Adv Neural Inf Proces Syst* 2013.
94. Tang, D., F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin. Learning sentiment-specific word embedding for Twitter sentiment classification. in *The 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*. 2014. USA.
95. Labutov, I. and H. Lipson. Re-embedding words. in *Association for Computational Linguistics (ACL)*. 2013. Bulgaria.
96. Forman G. An extensive empirical study of feature selection metrics for text classification. *J Mach Learn Res*. 2003;3:1289–305.
97. Zheng Z, Wu X, Srihari R. Feature selection for text categorization on imbalanced data. *ACM Sigkdd Explor Newsl*. 2004;6(1):80–9.
98. Uchyigit, G. Experimental evaluation of feature selection methods for text classification. in *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2012 9th International Conference on. 2012. IEEE.
99. Manning, C.D., P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Vol. 1. 2008: Cambridge University Press.
100. Sebastiani F. Machine learning in automated text categorization. *Acm Comput Surveys (Csur)*. 2002;34(1):1–47.
101. Ng, H.T., W.B. Goh, and K.L. Low. Feature selection, perceptron learning, and a usability case study for text categorization. in *ACM SIGIR Forum*. 1997. ACM.
102. Galavotti, L., F. Sebastiani, and M. Simi, Experiments on the use of feature selection and negative evidence in automated text categorization, in *Research and Advanced Technology for Digital Libraries*. 2000, Springer. p. 59–68.
103. Fragoudis D, Meretakis D, Likiothanassis S. Best terms: an efficient feature-selection algorithm for text categorization. *Knowl Inf Syst*. 2005;8(1):16–33.
104. Simeon, M. and R. Hilderman. Categorical proportional difference: A feature selection method for text categorization. in *Proceedings of the 7th Australasian Data Mining Conference*. 2008. Australian Computer Society Inc.
105. Denecke, K., Are SentiWordNet scores suited for multi-domain sentiment classification?, in *Fourth International Conference on Digital Information Management, (ICDIM 2009)*. 2009, IEEE. p. 1–6.
106. Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. LIBLINEAR: a library for large linear classification. *J Mach Learn Res*. 2008;9:1871–4.