Lang Resources & Evaluation
https://doi.org/10.1007/s10579-018-9432-0

ORIGINAL PAPER

Applying data mining and machine learning techniques for sentiment shifter identification

Zeinab Rahimi[1] · **Samira Noferesti**[2] · **Mehrnoush Shamsfard**[1]

Abstract Sentiment shifters, as a set of words and expressions that can affect text polarity, play a fundamental role in opinion mining. However, the limited ability of current automated opinion mining systems in handling shifters is a major challenge. This paper presents three novel and efficient methods for identifying sentiment shifters in reviews in order to improve the overall accuracy of opinion mining systems: two data mining based algorithms and a machine learning based algorithm. The data mining algorithms do not need shifter tagged datasets. They use weighted association rule mining (WARM) for finding frequent patterns representing sentiment shifters from a domain-specific and a general corpus. These patterns include different kinds of shifter words such as shifter verbs and quantifiers and are able to handle both local and long-distance shifters. The items in WARM for the two designed methods are in the form of dependency relations and SRL arguments of sentences, respectively. Secondly, we implemented a supervised machine learning system based on semantic features of sentences for shifter identification and polarity classification. This method obviously needs shifter tagged dataset for shifter identification. We tested our proposed algorithms on polarity classification task for 2 domains: a specific domain (drug reviews) and a general domain. Experiments demonstrate that (1) the extracted shifters improve the performance of the polarity classification, (2) the proposed data mining methods outperform other implemented

✉ Mehrnoush Shamsfard
m-shams@sbu.ac.ir

Zeinab Rahimi
z_rahimi@sbu.ac.ir

Samira Noferesti
snoferesti@ece.usb.ac.ir

[1] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran

[2] University of Sistan and Baluchestan, Zahedan, Iran

Published online: 12 November 2018

methods in shifter identification, and (3) the proposed semantic based machine learning method has the best efficiency among all implemented methods in polarity classification.

Keywords Sentiment analysis · Opinion mining · Sentiment shifters · Shifter identification · Association rule mining · WARM · Machine learning

1 Introduction

Opinion mining, a branch of sentiment analysis, is the task of extracting and analyzing opinions, sentiments, evaluations, or feelings from user-generated contents such as reviews, discussion groups, and blogs. Due to its wide range of applications, such as analysis of customer reviews (Hu and Liu 2004) and reputation management (Wiegand et al. 2010), this field has received considerable attention both in industrial and academic research areas. One of the main subtasks of opinion mining is polarity classification, which aims at classifying opinions into predefined classes (usually positive and negative). Existing approaches to polarity classification can be grouped into two main categories: lexicon-based and machine learning approaches. Lexicon-based approaches mainly rely on linguistic resources containing polar terms and concepts such as SentiWordNet (Esuli and Sebastiani 2006), Senticnet (Cambria et al. 2016), General Inquirer (Stone et al. 1966), and Subjectivity Lexicon (Wilson et al. 2009). For example, "this drug is amazing" is a positive sentence since the term "amazing" is positive in sentiment lexicons. However, these resources are not sufficient since the polarity classification is a challenging task that needs to tackle many subtle phenomena such as sentiment shifters.

Sentiment shifters, also called valence shifters, are words and expressions that affect the polarity of an opinion by changing its magnitude or its direction. For example, in the sentence "I do not like this drug", the shifter word "not" before the positive word "like" changes the text polarity to negative. Therefore, ignoring sentiment shifters can lead to a noticeable decline in overall accuracy of opinion mining systems. There are two types of shifter words or shifter trigger words: (1) words that reverse the polarity of the given text (e.g., "no" and "never") and (2) words that change sentiment values by a constant amount (e.g., "severe" and "mild"). In this paper, we only focus on the first type, i.e., reversing words. Reversing words are not limited to negation words. Some kinds of verbs (e.g., "reduce") and quantifiers (e.g., "less") can act as the first type of sentiment shifters.

From another perspective, sentiment shifters can be classified into two main groups: local shifters indicating shifter words, which are directly applied to polar words (e.g., "Accutane doesn't help"), and long-distance shifters, which allow longer distance dependencies between the shifter words and the polar words (e.g., "No one ever likes this drug"). Although sentiment shifter identification plays a fundamental role in recognizing polarity of textual expressions, it has not been completely solved.

This paper presents three novel and efficient approaches to identify sentiment shifters using polarity-tagged sentences: two data mining approaches and a semantic machine learning approach. The proposed approaches use syntactic and semantic relations in a sentence and are able to handle both local and long-distance shifter words.

In the proposed data mining approaches, patterns for different kinds of shifter words are extracted; for example, for negation structures (e.g., "no" and "not"), shifter verbs (e.g., "decrease", and "eliminate"), and shifter quantifiers, i.e., words, which express a decreased/increased value of quantity (e.g., "less") while most of the existing approaches just focus on negation words. In addition, the proposed approaches are language-independent. Thus, although we tested it only in English, it can be used for other languages, as well. We also incorporate the extracted patterns into a lexicon-based method for polarity classification. In addition to the data mining approaches, we proposed a semantic machine learning based method that can be used in both shifter identification and polarity classification tasks.

This paper is an extension over the study conducted by Noferesti and Shamsfard (2016), but both semantic based systems (ML and SRL based data mining systems) are the new contributions of this paper compared to the original work.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces our proposed approaches in detail. Section 4 presents the details of evaluations and discussion of the paper. Finally, Sect. 5 concludes the paper.

2 Related work

Identifying shifter words and determining their scope (i.e., part of the sentence that is affected by the shifter) are the main tasks followed in this study. Existing shifter identification approaches can be classified into two main categories: (1) lexicon-based and rule-based methods and (2) statistical and machine learning approaches.

2.1 Lexicon based and rule-based methods

Lexicon based methods mainly rely on a list of common shifter words that is built manually (Huang et al. 2014; Marrese-Taylor et al. 2014). The main limitation of these methods is that such lists in many languages may be incomplete and, hence, there is always a need to propose a way to deal with words that are not in the lists. Furthermore, due to the language dependency nature of shifter words, it is difficult to adapt these lists to other languages.

Moreover, some researchers have proposed simple heuristic rules that define the scope of a shifter word using a window of fixed size (Hu and Liu 2004; Heerschop et al. 2011; Huang et al. 2014). In (Shaikh et al. 2007; Asmi and Ishaya 2012) the scope of negation was identified by using the dependency tree, which indicates how a negation word interacts with other words of the sentence.

In (Simancík and Lee 2009), a linguistic system for sentence-level valence annotation is presented. This system employs the formalism of Combinatory

Categorial Grammar (CCG) to represent words as functions acting on their syntactic arguments. In this work, CCG was used to determine the structural dependencies between individual terms in a sentence and consequently to determine shifter scope. Also, shifter lexicons are used to estimate the valence of individual terms.

2.2 Statistical and machine learning based methods

In Statistical and machine learning methods, the shifter detection problem is solved via statistical approaches that mostly have a computational complication due to involving a large amount of data. Yu et al. (2016) proposed MTSA, a data-driven sentiment analysis framework, to enable polarity predictions of the same word in reviews of different themes. This framework focuses on discovery and quantification of contextual valence shifters. MTSA addresses the shifter effect learning problem as a logistic regression. To rigorously formulate the problem, a series of intuitive assumptions are proposed.

Xia et al. (2016) proposed a method for shifter detection. In this method, each document is split into a set of sub-sentences and then a hybrid model that employs rules and statistical methods is built to detect polarity shifts. For detecting inconsistent sentiment in the text, the method employs the weighted log-likelihood ratio (WLLR) algorithm to find relevance between sub-sentences and sentence polarity.

Boubel et al. (2013) presented a method that automatically identifies contextual valence shifters. This method relies on a Chi square (χ^2) test applied to the contingency table representing the distribution of a candidate shifter in a corpus of reviews of various opinions. The system depends on two resources—a corpus of reviews and a lexicon of valence terms—to build a list of contextual valence shifters.

Along with the above statistical methods, some researchers have used shifter words and their scopes as a feature for polarity classification using machine learning approaches (Pang et al. 2002; Kennedy and Inkpen 2006; Jia et al. 2009; Wilson et al. 2009; Morante and Blanco 2012). Although these approaches can capture some aspects of the shifters effectively, they depend upon the availability of an annotated corpus in which shifter words and their scopes are tagged. Manual construction of such corpora is a tedious, expensive, and time-consuming task.

In most of machine learning shifter identification algorithms, a lexicon of polar words is used to find sentences that probably have shifters and the main difference of systems is among their feature vectors. Kennedy and Inkpen (2006) introduced a machine learning approach that examines three types of valence shifters: negations (reversing the polarity), intensifiers, and diminishers (increasing and decreasing the value of polarity, respectively). This system uses bigrams as features that consist of a valence shifter and another word to capture the type of the valence shifter and uses SVM classifier. For example, it selects bigrams such as “very good” where very is an intensifier and identify the bigram as int-good, where “int” indicates any intensifier.

Ikeda et al. (2010) proposed a machine learning method for modeling polarity-shifters in which the local context of three words to the left and right of the target polar word has been used as the feature representation. This model is a kind of

binary classification model that determines whether the polarity is shifted by its context. The model assigns a score to the polar word according to its surrounding words.

Li et al. (2010) introduced a machine learning approach to incorporate polarity shifting information into a document-level sentiment classification system. In this system using a binary classifier, each document in training data is split into two partitions of polarity-shifted and polarity-unshifted, which are used to train two base classifiers. They used n-grams, document frequency of terms in one category, and the ratio of document frequency in one category divided by other categories as classification features.

Morsy and Rafea (2012) proposed a machine learning method for shifter identification in order to improve the performance of document-level sentiment analysis. The proposed feature sets refine the traditional sentiment feature extraction method and take contextual valence shifters into consideration from a different perspective than the earlier research. These feature sets include (1) a feature set consisting of 16 features for counting different categories of contextual valence shifters (intensifiers, negators, and polarity shifters) as well as the frequency of words grouped according to their final (modified) polarity and (2) another feature set consisting of the frequency of each sentiment word after modifying its prior polarity.

Recently, some interesting papers have been published on dual training and prediction in sentiment analysis and polarity shift. In (Xia et al. 2013, 2015), the polarity of train and test sentences are reversed during some steps such as negating the sentence considering the negation scope and reversing polar words and also train labels are reversed. Then, the system training procedure is performed using both original and reversed datasets.

In this section, we employ no data mining category because, to the best of our knowledge, no data mining method has been presented for shifter identification so far. The methods proposed in this work will fit in data mining based and machine learning categories. Moreover, in the proposed machine learning algorithm, a semantic feature set based on semantic role labeling representation of sentences is presented.

3 The proposed approach

As mentioned before, sentiment shifters can reverse the polarity of the given text and, hence, are vital for precise polarity classification task. In particular, in the drug domain, most of the medical terms such as “pain” and “depression” are negative, but they occur frequently in positive sentences.

Consider the following examples:

- “Accutane eliminated my cystic acne”.
- “It reduced my pain”.
- “No pain”.
- “Less acne”.

We assume that when the polarity of a sentence is different from the polarity of the majority of its words, that sentence may have a valence shifter. This assumption is commonly used among other shifter identification approaches. In the above examples, the valence shifters "eliminated", "reduced", "no", and "less" invert the polarity of the corresponding polar terms. Therefore, capturing such shifters will improve the performance of polarity classification. Given this insight, the idea behind our approaches is to identify sentiment shifters.

Due to this target, for shifter identification task, we can have two different approaches for whether we have shifter tagged dataset or not. To make a shifter tagged dataset, we use the above assumption. So, we propose 3 systems to identify shifters in a sentence and determine the polarity in an opinion or review:

- Two data mining based shifter pattern extraction systems
- A semantic-based machine learning system

These systems are elaborated in the following subsections.

3.1 Data mining based shifter pattern extraction systems

Our proposed data mining approaches for shifter identification consist of two main steps: candidate sentence extraction and frequent shifter pattern mining. In the first step, we extract candidate sentences from a corpus. Two corpora; a domain-specific one (in the medical-drug domain) and a general one are selected in order to compare the efficiency of our algorithm on both cases. Then, in the second step, we mine frequent patterns in a set of candidate sentences. The patterns are extracted once from dependency trees and once from semantic roles of the candidate sentences, which are in line with our two proposed data mining methods. These two sub-approaches are then compared. In the following subsections, we describe each of these steps in detail. Flowchart of our data mining based systems is presented in Fig. 1.

3.1.1 Extracting candidate sentences

To extract candidate sentences for shifter identification, we use a corpus of polarity-tagged sentences that contains two sets: A set of positive sentences (P) and a set of negative sentences (N). We divide each set into two subsets: positive sentences with negative terms (PN), positive sentences with positive terms (PP), negative sentences with negative terms (NN), and negative sentences with positive terms (NP).

If a sentence includes both positive and negative terms, we use term-counting method. For example, PN includes positive sentences that have more negative terms than positive ones. This is the case for other sets as well. To determine polar terms,

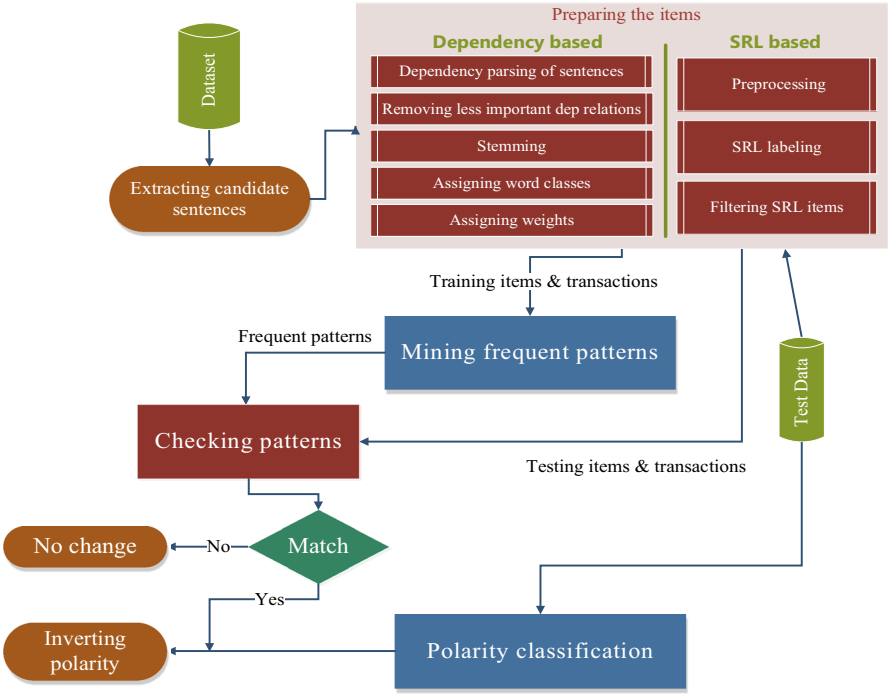

Fig. 1 Flowchart of data mining based systems

we use a sentiment lexicon of 5330 words in the medical domain, which is built manually (Noferesti and Shamsfard 2016), and a set of 6000 general polar words that are downloaded online.[1] Finally, we select PN and NP sets as candidate sets; i.e., sentences including sentiment shifters.

3.1.2 Mining frequent patterns

The second step is to extract patterns that appear significantly more frequently in the PN (or NP) than other sets. The idea is that these patterns represent shifter patterns since they are frequent in sentences with shifters but not frequent in sentences without shifters.

Thus, in this step, frequent patterns will be extracted as shifter patterns. In the proposed method, this process is performed on two groups of elements: dependency relations and semantic roles of the candidate sentences. So, we can have two sets of shifter patterns based on syntax and semantic information.

In order to extract shifter patterns, we use weighted association rule mining (WARM). The ARM is one of the key data mining techniques that have been used to tackle a variety of applications (Agrawal and Srikant 1994). The ARM consists of two subtasks. The first subtask is the frequent itemset mining, which generates all

[1] http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html.

items whose supports are higher than a predefined threshold called minimum support. The second subtask generates association rules that satisfy the minimum support and minimum confidence thresholds. WARM generalizes the classical model to the case where different items have different weights to reflect their different importance. To extract shifter patterns, we present a two-step procedure. First, we extract important dependency or SRL[2] relations of the sentences and then we adopt WARM to find frequent shifter patterns.

3.1.2.1 Dependency-based patterns To extract important dependency relations of a sentence, we perform the following steps (Noferesti and Shamsfard 2016):

- Extracting dependency relations of the sentence: For each sentence, a set of dependency relations is obtained from the Stanford dependency parser.[3] Each dependency relation represents a relation between two words. We show a dependency relation with a triplet r (relation-name, word1, word2).

- Removing less important relations: Less important relations—i.e., the dependency relations containing very common words (stopwords)—are stripped out.
- Stemming: For each remaining dependency relation, we use Stanford stemmer to reduce different forms of a word to one canonical form or lemmatize it.
- Assigning word classes: For each dependency relation, we replace the polar words involved in the dependency with their classes. The class indicates the part of speech (POS) tag and the polarity of that word. For example, the class "A_POS" is assigned to positive adjectives like "good". In this way, we generalize the dependency relations and, as a result, the extracted shifter patterns. Generalized shifter patterns have higher coverage than specific ones, and so have a greater chance of matching a context.
- Assigning weights: In this step, each sentence is described by a set of dependency relations that is represented as a vector v = {(r, w), (r, w), …, (r, w)}, where r is a dependency relation and w is its weight. Also, m is the number of dependency relations in the sentence. The weights can be determined in a number of ways. In this paper, we simply use one of the most widely used weighting approaches called TF-IDF.[4] In this approach, the weight of relation r in a sentence is defined as follows:

$$w_{ij} = tf_{ij} \cdot log_2^{\left(\frac{n}{df_{ij}}\right)} \quad (1)$$

where tf is the number of occurrences of relation r in the PN (or NP), df is the total number of occurrences of relation r, and n is the total number of sentences. In fact, TF-IDF is intended to reflect the importance of a dependency relation as a shifter pattern.

[2] Semantic Role labels.

[3] https://nlp.stanford.edu/software/stanford-dependencies.html.

[4] Term frequency-inverse document frequency.

3.1.2.2 Semantic-based patterns To extract important semantic patterns for shifter identification, we perform the following steps:

- Providing semantic representation of candidate sentences: For this purpose, we use SENNA[5] tool as a semantic role labeler. After applying SENNA, we have predicates and the semantic arguments of candidate sentences.
- Selecting semantic roles that usually affect sentiment shifts. In the proposed system, we used the following roles as items for pattern mining:PredicatesPredicates
    - Predicates
      Semantic class of predicates in WordNet[6] (Miller 1995); for example, clear off–verb.change
      Upper semantic class of predicates in VerbNet[7] (Kipper et al. 2000); for example, "become" for "come,turn,get,go"
    - Arguments containing polar words
    - Semantic class of polar word in WordNet
    - AM-NEG label
      This label usually illustrates negation, which helps in shifter detection.
    - AM-TMP label
      Some temporal information such as "no longer" will help to find sentiment shifters, especially in the medical domain.
    - AM-DIR label
      This label indicates the direction in sentences, for example, up or down, and will be useful in shifter identification.
    - Providing POS tags and stems of words for more accurate processing.

3.1.3 Extracting shifter patterns

In this step, we use an Apriori-like algorithm to explore frequent weighted relations (Zhang and Zhang 2002). Apriori is a well-known algorithm for ARM (Agrawal and Srikant 1994). Given a set of transactions, where each transaction is a set of items, Apriori algorithm aims to find frequent itemsets; i.e., item sets whose occurrences are greater than a user-specified minimum support. In the first step, Apriori finds the frequent individual items and, in each next step, it extends each subset with one item at a time to generate frequent groups of items. We use a modified implementation of an Apriori-like method to mine frequent shifter patterns. Dependency relations (or semantic role labels) and sentences become "items" and "transactions", respectively, in the frequent itemset mining framework. The first scan finds weighted frequent individual relations whose supports (weights) are greater than the minimum support threshold. In the first scan, we impose a restriction; we only mine weighted frequent individual relations that contain at least one polar word.

[5] http://ronan.collobert.com/senna/.

[6] https://wordnet.princeton.edu/.

[7] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html.

This set is called 1-relation set. Each subsequent scan starts with the set of frequent relation sets found in the previous scan. This set is used to generate a set of new potential shifter patterns. Candidates whose weights are greater than the threshold form the set of newly found shifter patterns, called k-relation set. The algorithm terminates when no candidate relation set can be generated or no candidate pattern can be found. Among extracted frequent patterns (relation sets), we only select those whose confidences are higher than a specific threshold, called minimum confidence. The confidence of a pattern presents its accuracy; i.e., the ratio of correct shifters detected by this pattern in a set of instances matching it. The confidence is computed as follows:

$$\text{Confidence} = \frac{\text{No. of correctly detected instances}}{\text{No. of instances match the pattern}} \quad (2)$$

We employ particle swarm optimization (PSO) (Kennedy 2011) to adjust the values of minimum support and minimum confidence parameters. In the area of association rule mining, PSO is successfully used for the determination of these threshold values (Kuo et al. 2011). PSO tries to find the best values with which we gained the best performance in shifter identification on the development set. In this way, we can have different values for minimum supports in each iteration. Finally, the extracted frequent relation sets represent shifter patterns.

3.1.4 Incorporating shifter patterns into lexicon-based approaches for polarity classification

To incorporate the extracted shifter patterns into a lexicon-based approach for polarity classification, we first tag the polarity of the given sentence using a sentiment lexicon. Then, we produce items for WARM using Stanford parser or SENNA semantic role labeler and make the vector of items for it. Finally, if the vectors match with a shifter pattern the polarity of the sentence will be reversed.

3.2 The semantic-based machine learning system

Here, we propose a machine learning system to reach two targets:

1. Comparing the performance of a direct machine learning system with our shifter identification systems, considering the fact that most of the recent sentiment shifter identification systems focus on machine learning process.
2. Incorporating our extracted shifter patterns in a machine learning sentiment classification system along with a lexicon-based one

We decided to train a semantic-based machine learning system, considering two issues: (1) Have a better vision for analyzing the effect of semantic information in machine learning based polarity classification and shifter detection and (2) as far as we know, none of the existing machine learning shifter identification systems

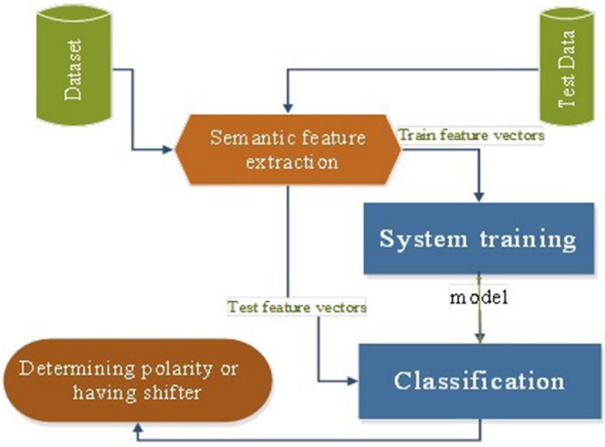

Fig. 2 Flowchart of a semantic-based machine learning system

directly rely on semantic features. In this regard, we extracted semantic features from the same data we used for associate rule mining.

In machine learning approaches for shifter identification, the main task is typically addressed as a 2-class classification problem (having a shifter or not). Here, there is a similar procedure. After generating feature vectors for candidate sentences, a Naïve Bayes machine learning classifier is trained and the associated model is used for the test process. Training and testing processes are performed by WEKA toolkit.

The feature set, as mentioned earlier, consists of a predicate, semantic class of predicates in WordNet, upper semantic class of predicates in VerbNet, arguments containing polar words Semantic class of polar word in WordNet, AM-NEG label, AM-TMP label, and AM-DIR label. Also, polarity feature can be added to the feature set.

According to the mentioned two targets and using the defined features, two distinct systems are trained:

- For shifter identification (determining whether a sentence polarity is shifted), we used semantic feature vectors of training data (once without considering "Polarity" feature and once with it) to train and test the system. To tag sentences for having a shifter, we counted polar terms and computed expected polarity as the number of positives minus negatives. If the polarity of the sentence is different with the expected polarity, then we assume that there is shifter in the sentence. Although this heuristic is not necessarily correct, it works well in many cases.
- For polarity classification (determining sentence polarity), we used semantic feature vectors of training data (once without considering "Shifter" feature and once with it) to train and test the system. To have shifter (feature value), we used SRL patterns extracted in Sect. 3.1.3.

Flowchart of semantic-based machine learning system is presented in Fig. 2.

4 Experiments

As mentioned above, in this article, three main algorithms are implemented for sentiment shifter identification task. We performed experiments to evaluate three issues:

- Comparing shifter identification algorithms
- Studying the effect of shifter identification in polarity classification task
- Studying the effect of changing the dataset domain (general or specific) on the performance of the proposed algorithms

In this section, the evaluation procedure and performance analysis of the proposed systems are presented.

4.1 Data and metrics

To train and test the work, we employed two corpora; a domain-specific corpus (in medical drug domain) and a general corpus. We used them to extract candidate sentences for shifter identification (both data mining methods) and also for our machine learning method:

1. A corpus of polarity-tagged sentences was collected from www.druglib.com website. This corpus contains 2776 reviews for 85 drugs. Sometimes, different parts of a compound sentence have different polarities. Thus, to achieve more accurate shifter patterns, compound sentences are broken down into simple sentences. Splitting is done by exploiting dependency tree and conjunction structure of the sentence (De Marneffe et al. 2006). This dataset is also used as the training set of the machine learning algorithm.

For test data, we used a test set of 1500 sentences collected from www.druglib.com and www.askapatient.com for all systems. In order to have a precise evaluation, this test set is manually labeled by an expert to find out whether they have a shifter. The details of the annotated corpus are presented in Sect. 4.1.1.

2. A general corpus of polarity-tagged sentences containing 4000 sentences labeled with the positive or negative sentiment was extracted from reviews of products, movies, and restaurants. To be more specific, the sentences are collected from three different websites/fields: imdb.com, amazon.com, and yelp.com. We used 800 sentences as test data and the rest for training and developing the datasets.

Evaluation measurements in shifter identification and polarity classification tasks are similar to most of NLP tasks and consist of precision, recall, and F-measure, which are defined as follows:

$$\text{Precision(P)} = \frac{\text{correct system decisions}}{\text{all system decisions}} \tag{3}$$

$$\text{Recall(R)} = \frac{\text{correct system decisions}}{\text{what system should have decided}} \tag{4}$$

$$F\text{−}measure = \frac{2 \cdot P \cdot R}{(P + R)} \tag{5}$$

4.1.1 Annotated test corpus

As mentioned above, to have a precise shifter identification evaluation, test data set is manually labeled by an expert for having a shifter or not. In this regard, an annotation guide is prepared to define 4 kinds of tags that were used for annotating polar phrases and shifters. The sentences in this corpus, were already polarity tagged and we added the following four tags to their words and phrases:

1. Positive polar word or phrase
2. Negative polar word or phrase
3. Intensifier shifter (quantifiers)
4. Negation shifter

In our guideline, the needed instructions toward each tag selection are prepared and some examples and usage explanations of annotation toolkit are provided. Also, it has been noted that patterns should be considered in annotation process in addition to the common sense knowledge and if an element appears in two separated parts in the sentence, an index should be assigned to it.

In this corpus, we annotate polar phrases to determine exact shifter scope. Also, numbers are assigned to polar phrases and shifters such that the associated pairs (polar phrases and their shifters) have identical numbers. Thus, it clear that which shifter belongs to which polar phrase in a sentence containing more than one polar phrase. The statistics of our shifter tagged test corpus is presented in Fig. 3.

We simplify the annotation process using our implemented annotation tool. Annotators just need to right click on the selected words and select the appropriate tag. A screenshot from our annotation tool is presented in Fig. 4.

In the evaluation procedure, we extract polar and shifter gold tags for each test sentence (Fig. 5) and then compare the implemented systems outputs with them. It is noteworthy that our WARM algorithms detect both negation and intensifiers.

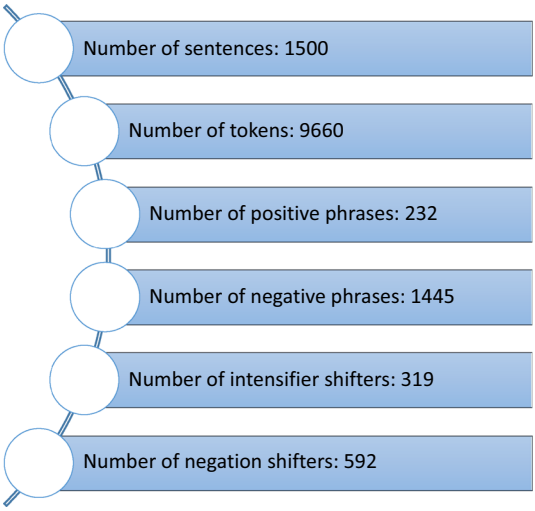

Fig. 3 Statistics of shifter annotated

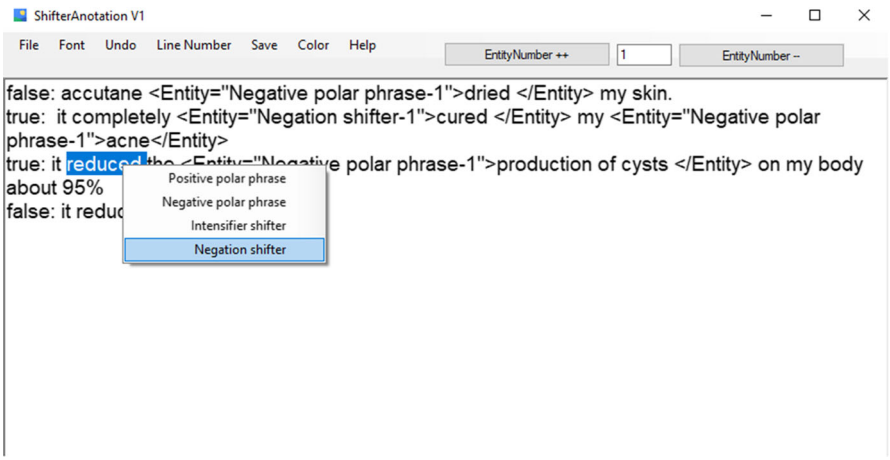

Fig. 4 GUI of the implemented annotation tool

Fig. 5 Summary of shifter annotated test corpus

```
1  shifter-amp-1 neg-1 neg-2
2  shifter-neg-1 neg-1
3  shifter-neg-1 shifter-amp-1 neg-1
4  shifter-neg-1 neg-1 neg-1
5  shifter-neg-1 neg-1
6  shifter-amp-1 neg-1 neg-2
7  shifter-neg-1 neg-1
8  shifter-neg-1 neg-1
9  shifter-amp-1 pos-1
10 neg-1 neg-2
11 neg-1
12 neg-1
```

4.2 Results

4.2.1 Polarity classification evaluation (drug domain)

As mentioned in Sects. 3.1.4 and 3.2, we incorporate the polarity shift information of WARM into sentiment classification in 2 polarity classification methods: (1) lexicon-based method and (2) semantic machine learning method.

Here, we present the experimental results of polarity classification task using four algorithms. Two of these algorithms are based on data mining; one considering dependency relations as the items (Proposed-DM-Dependency)[8] and the other considering semantic roles as the items (Proposed-DM-SRL). The other two algorithms employ machine learning techniques using semantic features; one without shifter as a feature (Proposed-ML) and the other one with shifter as a feature (Proposed-ML + Shifter).

To assess the effectiveness of incorporating shifter patterns into lexicon-based methods for polarity classification, we first determine the polarity of each sentence in a test set using two sentiment lexicons: a domain-specific (in drug domain) and two general-purpose (a general polar words lexicon and SentiWordNet) lexicons. Then, if that sentence matches a shifter pattern, its polarity is inverted.

Table 1 illustrates the performance of the proposed approaches and compares them with the baselines. The first two rows of the table show the performance of two lexicon-based methods without shifter identification (using SentiWordNet and Domain Specific Lexicon). The third and fourth row show the performance of our two proposed data mining algorithms added to the basic lexicon-based method (domain specific).

Moreover, we evaluated the effect of shifter identification in machine learning method by feeding the binary feature of "having shifter" to the system and used the WARM information to help statistical learning method in sentiment classification. Here, the fifth row (ML-unigram) shows the performance of a machine learning baseline (using just unigrams as features) and the last two rows indicate our two proposed semantic machine learning methods (without and with shifter).

Table 1 summarizes the effect of adding WARM in a lexicon based (domain specific) and machine learning method in sentiment classification. As can be expected, including shifter patterns has a considerable effect on the performance of polarity classification. Also, Table 1 illustrates that among data mining methods, the dependency based one has better precision and SRL based one has better recall and F-measure.

Table 1 also compares the efficiency of machine learning based methods with each other and with data mining methods. As can be seen, the proposed semantic based machine learning method outperforms the simple one. Moreover, the

[8] The Proposed-DM-Dependency method originally was presented in our conference paper (Noferesti and Shamsfard 2016) and we used it here as baseline to compare with our new methods.

Table 1 Comparison of the approaches proposed for polarity classification on drug reviews

Approach	Method	Precision	Recall	F-measure
Lexicon-based	SentiWordNet	58	30	39.5
	Domain Specific Lexicon	61.05	31.4	41.4
Lexicon based + WARM	Lexicon based + proposed-DM-Dependency	67.25	33	44.5
	Lexicon based + proposed-DM-SRL	62.8	36.2	45.9
Machine learning	ML-unigram	70.6	70.8	70.5
	Proposed-ML	76.5	75.9	76.1
	Proposed-ML + shifter	86.3	86.1	86.2

proposed machine learning method with shifters as features is the best in drug review domain among all those compared in the table. However, as it is a supervised method and needs a shifter tagged corpus to be trained, it may not be suitable in other domains. In such cases, the proposed data mining methods are good enough to identify the polarity.

Overall, these results show that using semantic information has a positive effect on the polarity classification task. Also, according to the two last rows of Table 1, involving shifters can improve polarity classification systems.

4.2.2 Shifter identification evaluation (drug domain)

As one of the main contributions of this paper, the WARM algorithm, which is actually a sentiment inconsistency detection method, is evaluated directly in this section. In this evaluation, we measure the precision and effectiveness of WARM based on a manually labeled sentiment inconsistency test corpus and compare the proposed shifter identification methods with 5 other existing sentiment inconsistency detection (shifter identification) methods (Fig. 3): (1) a baseline method, where each appearance of valence shifters inverts the polarity of text, (2) NegEx algorithm (Chapman et al. 2001), (3) (Huang et al. 2014), (4) a rule-based approach (Asmi and Ishaya 2012), and (5) WLLR (Xia et al. 2016). NegEx is a negation detection algorithm in biomedical texts that is based on regular expressions and a dictionary of medical terms. NegEx usually correctly detects negated terms; however, it is not able to detect other kinds of shifters such as shifter quantifiers and shifter verbs. Huang et al. (2014) used a set of simple heuristic rules to define the scope of a shifter word using a window of fixed size. Also, Asmi and Ishaya (2012) identified the scope of negation using dependency tree. WLLR detects sentiment inconsistency by computing a relevance score for each sentence based on the occurrence of its words in positive and negative sentences.

In Table 2, rows 6–7 are our proposed WARM based methods and the two last rows are our proposed semantic based machine learning methods without and with using "polarity of the sentence" as a feature.

Table 2 Comparison of the proposed shifter identification approaches with other methods

No.	Method	Precision	Recall	F-measure
1	BaseLine	51.7	12.8	20
2	NegEx	51.8	12.5	20.2
3	Huang et al.	52.7	13	20.7
4	Asmi and Ishya	54.3	14.8	23
5	WLLR	50.6	19.1	27.3
6	Proposed-DM-dependency	80.3	47.1	59.4
7	Proposed-DM-SRL	59.8	74.5	66.3
8	Proposed-ML	60.6	58.3	58.8
9	Proposed-ML + polar	62	59.9	60.4

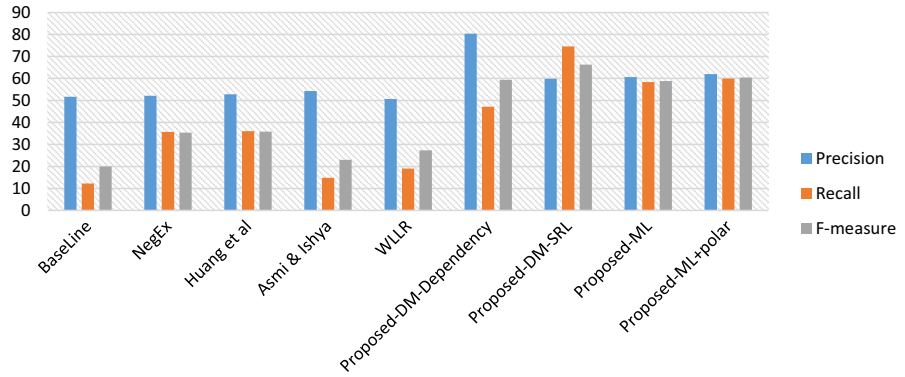

Fig. 6 Comparison of the proposed approaches with other methods of shifter identification

Figure 6 shows that the proposed data mining approaches for shifter identification outperform all other methods. The dependency-based system has a remarkable precision while the SRL-based one has the best F-measure. Also, it indicates that the semantic machine learning method has a close performance in shifter identification task. Certainly, machine learning methods have the limitation of needing shifter tagged dataset, which is discussed in the Sect. 4.2.4.

4.2.3 Studying the effect of dataset domain on performance

In order to analyze the effect of changing dataset on the proposed algorithms, main experiments on polarity classification are repeated for a general domain dataset. The evaluation results are presented in Table 3.

As illustrated in Table 3, the overall results of polarity classification in the proposed systems are higher in general dataset. In other words, applying the same algorithm on general dataset leads to a better performance and the general dataset has simpler shift patterns to detect compared to drug domain dataset.

Besides, by considering the general lexicon system as a base system, it is found that that the proposed systems outperform base system. For a deeper analysis, we can consider the performance difference of the base system and proposed systems to compare new results with the former evaluations. In this regard, as Table 3 indicates, a 3.8 unit improvement in the Proposed-DM-Dependency (compared to 3.1 in drug dataset) illustrates that there exist more accurate dependency patterns in general domain compared to the specific drug domain. However, in the proposed semantic based systems (ML and SRL-DM), the difference is less. Apparently, semantic features are more useful in drug domain.

4.2.4 Discussion

Applying the proposed approaches for shifter identification, we extracted two sets of shifter patterns (frequent dependency and SRL patterns). Tables 4 and 5 depict

Table 3 Polarity classification performance of the proposed approaches applied on general domain dataset versus drug-domain dataset

Method	Dataset	Precision	Recall	F-measure
SentiWordNet	Drug	58	30	39.5
	General	71	60	65.5
Domain Specific Lexicon	Drug	61.05	31.4	41.4
General lexicon	Drugs	63	21	31
	General	73	57.2	64.5
Proposed-DM-dependency	Drug	67.25	33	44.5
	General	86.8	56.3	68.3
Proposed-DM-SRL	Drug	62.8	36.2	45.9
	General	81	56	66
Proposed-ML	Drug	76.5	75.9	76.1
	General	76	75.9	76

some examples of the extracted shifter patterns. The second column of Tables shows a shifter pattern and the third column illustrates an example sentence for each shifter pattern. As can be seen in Tables 4 and 5, the proposed approach is able to handle some kinds of shifter verbs (e.g., "reduce" and "go away" in examples 2 and 6, respectively). Likewise, there are some patterns (e.g., example 7) to detect shifter quantifiers (e.g., "less").

Furthermore, the proposed approach can detect some kinds of long-distance shifters (e.g., examples 3 and 5).

In addition, most of the extracted patterns are not domain-specific. After applying the shifter identification method in drug review and general domain, it was found that the extracted patterns can be used in any other domain as well. However, for having a more general pattern set, we can extract shifter patterns from several domains for which polarity-tagged corpora are available.

It is noteworthy that frequent item sets data mining approaches are usually concerned with scalability. This issue is mostly related to the growth of the number and types of features and samples, which extends the search space of the problem in the creation of frequent itemsets and increases the complexity of the problem, specifically in the formation of the decision tree part. Although we almost doubled the size of training data in our general evaluation dataset and there was not any major problem, some thresholds should be specified in this regard for training dataset to avoid such problems.

Also, to the best of author's knowledge, shifter identification task is not solved via deep learning method yet probably due to lack of a large amount of shifter tagged data. Nevertheless, our proposed data mining method can properly work with limited amount of data and also can simply determine shifter scope.

Another point to mention is that the machine learning method has a close performance with rule mining methods in shifter identification and outperforms others in polarity classification. When testing the semantic features directly through

Table 4 Examples of shifter dependency patterns

No.	Shifter patterns	Example sentences or phrases
1	{det(n_neg,any), dobj(experience,n_neg), neg(experience,not)}	I did not experience any side effects from taking this medication
2	{nsubjtpass(reduce,n_neg), auxpass(reduce,be)}	My leg pain was reduced
3	{nsubj(v_pos,one),det(one,no)}	No one ever likes this drug
4	{prep(lack,n_pos)}	Lack of energy
5	{advmod(n_neg,longer),advmod(longer,no)}	I no longer have panic attacks
6	{nsubj(go,n_neg),prt(go,away)}	Also joint swelling and pain in my legs have gone away
7	{amod(n_neg,less), dobj(have,n_neg)}	I have less stress

Table 5 Examples of shifter SRL patterns

No.	Shifter pattern	Example
1	{polar word semantic class = 'DISO', predicate semantic class = 'see', NEG}	No change was observed in my diabetes
2	{polar word = 'good', predicate wordnet class = 'verb.change', NO NEG}	The good impact of medication changed after a while
3	{polar word = 'DISO', polar word wordnet class = 'adj.ppl',predicate = 'get', predicate semantic class = ','become',NO NEG}	My husband DISO got worse

a machine learning algorithm, although the obtained result is acceptable, it is not possible unless if we have shifter tagged corpora. For this purpose, we tagged the corpus using polar word counting (which is not necessarily correct) and then used that corpus as machine learning train dataset.

Considering the effects of semantic features in both proposed data mining and machine learning methods, proper performance of semantic features is also proved in the experiments; however, they are more useful in drug domain.

5 Conclusion

This paper proposed three novel methods for sentiment shifter identification based on dependency relations and semantic arguments. First, we employed data mining techniques to mine sentiment shifter patterns from a domain-specific corpus of polarity-tagged sentences. These patterns include different kinds of sentiment shifters such as negation structures (e.g., "no", "not", "no longer"), shifter verbs (e.g., "reduce" and "eliminate"), and quantifiers (e.g., "less"), and can detect both local and long-distance shifters. Afterward, we incorporated the extracted shifter patterns into lexicon-based and machine learning approaches for polarity classification. Experimental results showed that the proposed approaches improve the performance of those approaches significantly.

Furthermore, we compared the performance of the proposed approaches on a dataset of drug reviews to that of a baseline method and other approaches for shifter identification. Experimental results indicate that the proposed approaches outperform other methods and between two data mining based proposed methods, the dependency-based system is more precise but SRL based one has a better overall performance. Although the shifter identification method is tested on drug review domain, its results can be used in any other domain as well.

In addition, we implemented a semantic-based machine learning system both for shifter identification and polarity classification which has a very good performance. We conclude that our approach is appropriate for sentiment shifter identification, although extra knowledge is required to increase the performance. As a future work, it is possible to tag a corpus by our data mining methods and then use the corpus to train a machine learning approach.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of 28th international conference on very large data bases (VLDB)* (Vol. 8(14), pp. 92–108).

Asmi, A., & Ishaya, T. (2012). Negation identification and calculation in sentiment analysis. In *IMMM 2012, the second international conference*, *(c)* (pp. 1–7).

Boubel, N., François, T., & Naets, H. (2013). Automatic extraction of contextual valence shifters. In *Proceedings of the international conference recent advances in natural language processing RANLP 2013* (98–104).

Cambria, E., Poria, S., & Bajpai, R. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 2666–2677).

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics, 34*(5), 301–310.

De Marneffe, M.-C., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th international conference on language resources and evaluation (LREC 2006)* (pp. 449–454).

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th conference on language resources and evaluation* (pp. 417–422).

Heerschop, B., Van Iterson, P., Hogenboom, A., Frasincar, F., & Kaymak, U. (2011). Analyzing sentiment in a large set of Web data while accounting for negation. *Advances in Intelligent and Soft Computing, 86,* 195–205.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD 04, 4* (p 168).

Huang, S., Niu, Z., & Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems, 56,* 191–200.

Ikeda, D., Takamura, H., & Okumura, M. (2010). Learning to shift the polarity of words for sentiment classification. *Transactions of the Japanese Society for Artificial Intelligence, 25,* 50–57.

Jia, L., Yu, C., & Meng, W. (2009, November). The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 1827–1830).

Kennedy, J. (2011). *Particle swarm optimization. In Encyclopedia of machine learning* (pp. 760–766). Boston, MA: Springer.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. In *Computational Intelligence* (Vol. 22, pp. 110–125).

Kipper, K., Dang, H. T., & Palmer, M. (2000). Class-based construction of a verb lexicon. In *AAAI/IAAI*, 691, 696.

Kuo, R. J., Chao, C. M., & Chiu, Y. T. (2011). Application of particle swarm optimization to association rule mining. *Applied Soft Computing*, *11*(1), 326–336.

Li, S., Lee, S. Y. M., Chen, Y., Huang, C. R., & Zhou, G. (2010). Sentiment classification and polarity shifting. In *Aclweb.Org, (August),* (pp. 635–643).

Marrese-Taylor, E., Velásquez, J. D., & Bravo-Marquez, F. (2014). A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications, 41*(17), 7764–7775.

Miller, George A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Morante, R., & Blanco, E. (2012, June). *SEM 2012 shared task: Resolving the scope and focus of negation. In *Proceedings of the first joint conference on lexical and computational semantics-volume 1: proceedings of the main conference and the shared task, and volume 2: proceedings of the sixth international workshop on semantic evaluation* (pp. 265–274).

Morsy, S. A., & Rafea, A. (2012, June). Improving document-level sentiment classification using contextual valence shifters. In *International conference on application of natural language to information systems* (pp. 253–258).

Noferesti, S., & Shamsfard, M. (2016). Using data mining techniques for sentiment shifter identification. In *Tenth International Conference on Language Resources and Evaluation, LREC 2016* (pp. 2716–2720).

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up: sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 79–86).

Shaikh, M. A. M., Prendinger, H., & Mitsuru, I. (2007, September). Assessing sentiment of text by semantic dependency and contextual valence analysis. In *International conference on affective computing and intelligent interaction* (pp. 191–202).

Simančík, F., & Lee, M. (2009). A CCG-based system for valence shifting for sentiment analysis. *Advances in Computational Linguistics, 41,* 99–108.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: The MIT Press.

Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing, (July)*, (pp. 60–68).

Wilson, T. A., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics, 35*(3), 399–433.

Xia, R., Wang, T., Hu, X., Li, S., & Zong, C. (2013, August). Dual training and dual prediction for polarity classification. In *ACL2* (pp. 521–525).

Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. *Information Processing and Management, 52*(1), 36–45.

Xia, R., Xu, F., Zong, C., Li, Q., Qi, Y., & Li, T. (2015). Dual sentiment analysis: Considering two sides of one review. *IEEE Transactions on Knowledge and Data Engineering, 27*(8), 2120–2133.

Yu, H., Shang, J., Hsu, M., Castellanos, M., & Han, J. (2016). Data-driven contextual valence shifter quantification for multi-theme sentiment analysis. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 939–948).

Zhang, C., & Zhang, S. (2002). Association rules mining: Models and algorithms. In *Lecture notes in computer science* (Vol. 2307, p. 243).