A semantic framework based on domain knowledge for opinion mining of drug reviews

S. Noferesti[a]* • M. Shamsfard[b]

[a]*Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran*
[b]*Natural Language Processing Lab., Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran*

Abstract: Opinion mining has attracted increasing attention in recent years. Existing approaches that address general domains face two major challenges concerning polarity classification of drug reviews. Firstly, indirect opinions frequently occur in the drug domain, while the existing methods have mainly focused on direct opinions and ignored indirect ones. Secondly, previous works are not sufficient for polarity classification of ambiguous concepts in the drug domain.

This paper proposed a semantic framework based on domain knowledge to construct and exploit resources for indirect opinion mining of drug reviews. Accordingly, some methods were introduced, developed, and compared for building and exploiting a combined knowledge base, polarity-tagged corpus, and context-aware resources to detect the polarity of drug reviews. The test results showed that the proposed methods reached a precision of 89.18% and 80.4% in the application of the combined knowledge base and the polarity-tagged corpus for polarity detection of indirect opinions, respectively. Also, a precision of 79.93% was achieved with the use of context-aware resources that were constructed for polarity detection of ambiguous concepts. Overall, the results demonstrated the greater performance of the proposed methods compared to the existing methods.

*Corresponding author.
E-mail address: snoferesti@ece.usb.ac.ir (S. Noferesti).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

In recent years, the development and growing popularity of social media and review sites have led to the emergence of a rich source of patients' opinions about drugs on the web. These opinions are useful in various fields. They help users to be aware of others' experiences about the outcomes and side effects of medications in order to make informed decisions before using a product. They also enable pharmaceutical companies to evaluate the qualities of their products. However, because of the ever-increasing amount of opinions on the web, there is a need to develop automated methods to analyze them. Opinion mining, also known as sentiment analysis in some studies, is the field of study that provides automated methods to explore, analyze, classify, and summarize opinions.

One major challenge in opinion mining of drug reviews is the huge number of indirect opinions. Opinions are divided into two general categories of direct and indirect opinions (Liu, 2012). In a direct opinion, an entity or one of its aspects is described. In contrast, an indirect opinion expresses the impact of an entity called "the effective entity" or one of its aspects on another entity called "the affected entity" (Liu, 2012). This effect can have a positive or a negative polarity. The existing approaches have mainly focused on analyzing direct opinions, particularly explicit direct opinions. These opinions describe an entity by using sentiment words. In addition, some works have been done on analyzing implicit opinions that lack sentiment words (Greene & Resnik, 2009; Zhang & Liu, 201;). Although the importance of indirect opinions has been pointed out in some earlier studies, they do not provide any good solution to analyze such opinions (Inui et al., 2008; Liu, 2012; Wilson, 2008). In fact, in contrast to direct opinion mining that has been extensively studied, indirect opinion mining is largely unexplored. Meanwhile, in the drug domain, patients usually write about their experience of drug effectiveness and/or side effects instead of expressing a direct opinion using explicit sentiment words. Patients' experiences are often described without any explicit expression of opinion. Rather, the desirable or undesirable effects of the drug implicitly indicate a positive or negative sentiment toward the drug. Thus, ignoring indirect opinions reduces the precision of opinion mining systems in the drug domain.

Another challenge in opinion mining of drug reviews is related to ambiguous concepts, i.e., the concepts whose polarities change according to the context. Thus, polarity classification of ambiguous concepts needs taking into account the context information. The existing opinion mining methods are divided into the two categories of lexicon-based methods and machine learning approaches. Lexicon-based methods rely on lexicons (Asghar et al., 2016; Goeuriot et al., 2012; Huang et al., 2014; Na & Kyaing, 2015) or knowledge bases containing polar concepts (Cambria et al., 2010; Tsai et al., 2013). The main problem of current lexicons and knowledge bases is that they are static. In other words, the polarity of concept is considered static in these resources. It is evident that the symbolic methods based on static knowledge bases fail to detect the polarity of ambiguous concepts in some cases.

In contrast, machine learning approaches can partly manage ambiguous concepts by examining their surrounding texts (Araque et al., 2017; Pang et al., 2002; Bobicev et al., 2015; Gopalakrishnan & Ramaswamy, 2017; Habernal et al., 2015; Hasan et al., 2018). However, these methods suffer from the problem of data sparsity. In other words, the effectiveness of machine learning approaches is highly dependent on the quality of the training data and the features used for learning. This is while the users of review sites often have a tendency to write short texts. On www.druglib.com, there is a significant number of reviews that contain only an ambiguous sentence or phrase. In these cases, machine learning approaches do not have the required performance due to the sparsity of the text surrounding an ambiguous concept.

In regard to the above-mentioned challenges, the purpose of this paper is (1) to examine the shortcomings of the existing opinion mining methods in analyzing indirect opinions and ambiguous concepts; (2) to provide a semantic framework based on domain knowledge in order to construct the resources needed for indirect opinion mining of drug reviews, and use them for polarity classification (i.e., to determine the positivity or negativity of indirect opinions); and (3) to present a suitable solution for polarity classification of ambiguous concepts.

In summary, the main contribution of this paper is to provide a semantic framework based on domain knowledge for indirect opinion mining of drug reviews. In this regard, we integrated our previously proposed methods of constructing static resources (Noferesti & Shamnfard, 2015a; 2015b; Noferesti & Shamnfard, 2016) and advanced some new methods to build context-aware resources. To the best of our knowledge, this paper is the first that combines built-in semantic knowledge in the domain-specific resources with textual clues and side information in order to provide a more abstract and richer model of the context. Our proposed framework has four main features that distinguish it from other works in this field:

1) The proposed methods for constructing static resources are semi-automatic. In other words, they only rely on the structure of review sites and use the information contained in the domain knowledge without the need for human intervention.
2) Unlike most existing resources, the constructed knowledge base contains polar facts that are essential for the analysis of

indirect opinions. Many indirect opinions express the facts that are known in the domain knowledge. Specifically, at the drug domain, most of the patients' experiences are expressed in terms of the effectiveness or side effects of a given drug. Many available resources on the web include drug effectiveness and side effects. Therefore, these resources can be used to construct a knowledge base of polar facts.

3)Various contextual clues are used simultaneously to identify the polarity of ambiguous concepts. In the few studies on context-aware opinion mining (Weichselbraun et al., 2013a; 2014b), only the textual content of an opinion was used for the disambiguation of ambiguous words, i.e., assigning a correct polarity to an ambiguous word. These methods do not provide the required performance when the text surrounding the ambiguous concept is sparse. In the proposed framework, along with the review text, we have used the side information available on the review sites as well as the semantic knowledge derived from domain-specific resources. The results of the tests show that the problem of data sparsity can be largely overcome by the simultaneous use of different contextual clues.

4)The semantic knowledge derived from domain-specific resources plays an important role in the extraction of polar facts and the classification of their polarities. In the proposed framework, some mechanisms are suggested for using built-in knowledge in domain-specific ontologies and other sources available on the web in order to (1) recognize entities and extract their relations, (2) determine semantic classes of entities, (3) group synonymous relations, (4) mine semantic patterns, and (5) organize these patterns in a hierarchical structure. In addition, domain knowledge plays an effective role in the detection and disambiguation of ambiguous concepts.

The results of the tests reveal that the proposed methods for polarity classification of indirect opinions can dramatically increase the precision of the current opinion mining systems.

2. Methodology

The existing methods of opinion mining are mainly limited to the identification of direct opinions. Among them, successful methods depend on concept-level knowledge bases and/or machine learning classifiers. In this paper, we intend to use these methods in the analysis of indirect opinions. To this end, we propose a semantic framework for resource construction and exploitation that facilitates polarity classification of indirect opinions. A schematic view of the proposed framework is displayed in Figure 1. This framework has two major modules: construction of resources and their exploitation for polarity classification of opinions.

The module of resource construction consists of two parts. The first part (upper box) focuses on constructing static resources, while the second part (lower box) deals with the construction of context-aware resources.

In the module of resource exploitation, the constructed resources are used to classify the polarities of indirect opinions. More precisely, the training corpus and knowledge base are utilized in machine learning approaches to train the classifiers and in symbolic methods for polarity classification of indirect opinions.

2.1. Construction of static resources
Figure 1 shows that the construction of static resources is composed of three modules: extracting indirect opinions from input text, polarity classification of the extracted opinions, and generalization. In the following, we describe each of these modules to construct a knowledge base of indirect opinions and a polarity-tagged corpus.

2.1.1. Knowledge base construction
An indirect opinion is represented by a quadruple$< e_i, r_{ij}, e_j, p >$, where e_i is the effective entity, e_j is the affected entity, r_{ij} is the impact of the effective entity on the affected entity, and p is the polarity of the opinion. Of course, depending on the requirements of each situation, other components can be added to this quadruple. The developed knowledge base contains both quadruples indicating indirect opinions and semantic patterns extracted from those quadruples.

As shown in Figure 1, the information available on the web, including those expressed on the review sites or domain knowledge, can be used as the input of the resource construction module. In our previous works, we proposed two methods for constructing a knowledge base of polar facts, called FactNet, from domain knowledge (Noferesti & Shamsfard, 2015b), and a knowledge base of indirect opinions, called OpinionKB, from patients' opinions about drugs (Noferesti & Shamsfard, 2015a). In this paper, we combine these two knowledge bases. In this section, we briefly describe the proposed methods for knowledge base construction, and in Section 4, we present the methods of combining these knowledge bases.

In (Noferesti & Shamsfard, 2015b), we introduced a method that uses three LOD[1] datasets (i.e., *SIDER*, *Drugbank*, and *Dailymed*[2]) as domain knowledge to extract polar facts (indirect opinions). These three sets contain many positive and negative facts about drugs. For example, the side effects mentioned in *SIDER* or the facts mentioned in *Drugbank* in the field of 'toxicity' are negative. On the contrary, the facts mentioned about drug effectiveness in the field of 'indication' are positive. Similarly, 'indication' and 'warning' fields in *Dailymed* contain positive and negative facts, respectively.

---

[1] Linked Open Data.
[2] https://old.datahub.io/dataset provides access to Drugbank, DailyMed and SIDER

According to Figure 1, the three steps of constructing a static knowledge base of polar facts by using the domain knowledge are as follows:

- *Opinion extraction:* The main problem is that the mentioned facts are represented in free text, which makes them difficult to process by machine. Therefore, there is a need to convert the unstructured text into RDF tuples, i.e., <subject, predicate, object> by means of NLP techniques. In (Noferesti & Shamsfard, 2015b), we proposed a combination of lexico-syntactic patterns and a rule-based method to extract polar facts from existing knowledge in LOD in the structured form of RDF triplets. Then, in the sub-module of opinion representation within the module of opinion extraction (Figure 1), the extracted RDF tuples are displayed as a triplet $< e_i, r_{ij}, e_j >$ where e_i is the effective entity of the same subject, r_{ij} is the effect of the same predicate, and e_j is the affected entity of the same object.

- *Polarity detection:* By default, tuples that are extracted from 'indication' or 'indications and usage' fields are assigned positive polarity, and tuples extracted from 'toxicity', 'precaution', or 'warning' fields are assigned negative polarity. Then, the quadruple extracted in the form of $< e_i, r_{ij}, e_j, p >$ is added to a knowledge base of polar facts called FactNet. Of course, sentiment shifters are also considered in detecting the polarity of facts.

- *Generalization:* By having a set of polar facts, we use a generalization method to extract generalized patterns from the facts. Generalization is performed to overcome the problem of lack of sufficient knowledge in the knowledge base and increase the coverage of the polarity classification algorithm. Generalized patterns have more coverage and are, thus, more likely to match with a text.

In (Noferesti & Shamsfard, 2015b), we developed a two-stage method based on UMLS[3] taxonomy for extracting semantic patterns from polar facts and organizing them in a hierarchical structure. UMLS is a comprehensive vocabulary containing 1.7 million concepts classified into 130 groups. Each of these groups is called a semantic type. These types themselves are classified into 15 groups, each of which is called a semantic group. In the first step, the effective and affected entities of each tuple $< e_i, r_{ij}, e_j, p >$ in FactNet are replaced by their semantic types. Then, the frequent patterns of these semantic tuples, i.e., those seen more than a predefined threshold, are selected as semantic type (ST) patterns. In the end,

a polarity is assigned to ST patterns. A pattern will be considered positive if the number of positive tuples matched with that pattern is more than the negative tuples matched with it. The opposite is true for assigning a negative polarity to the patterns.

In the next step, the semantic group of entities is defined for each semantic tuple (i.e., the tuple whose subject and object are replaced by its corresponding ST). Then, the frequent patterns are considered as semantic group (SG) patterns. In fact, SG patterns are the same as ST generalized patterns. Finally, the generalized patterns are added to FactNet, which is a knowledge base of polar facts, their extracted semantic (ST) patterns, and generalized (SG) patterns.

In the drug domain, there are several review sites such as *webmd.com*, *askapatient.com*, and *druglib.com*, on which patients post their experiences about their consumption of drugs. In (Noferesti & Shamsfard, 2015a), we proposed a method to use opinions expressed on these review sites to construct a knowledge base of indirect opinions, called OpinionKB. According to Figure 1, the three steps of constructing the OpinionKB are as follows:

- *Opinion extraction:* In (Noferesti & Shamsfard, 2015a), we proposed a rule-based relation extraction technique to extract the relations of each sentence in the user review. Then, indirect opinions are represented in the structured form $< e_i, r_{ij}, e_j >$

- *Polarity detection:* Domain knowledge, linguistic rules, and review structure are used to detect the polarity of opinions. If the known benefits (side effects) of a drug are discussed in an opinion, the polarity of that opinion will be considered positive (negative). Two linguistic rules are used in this paper: (1) sentences and phrases that are connected with conjunctions like 'and' have the same polarities; and (2) sentences and phrases that are linked with conjunctions like 'but' often have opposite polarities. By applying these rules, if the polarity of a sentence is specified (in the previous step), the polarity of the other sentence can be determined. Then, the structures of the review sites are used to classify the polarity of untagged opinions. Most review sites have separated Pros and Cons opinions. In particular, on www.druglib.com, a user can write his/her opinion on the two parts of drug benefits and side effects. If an opinion occurs more frequently in the 'benefits' section than the 'side effects' section, it will be considered positive, and vice versa.

- *Generalization:* OpinionKB is generalized by using the same method proposed for FactNet generalization.

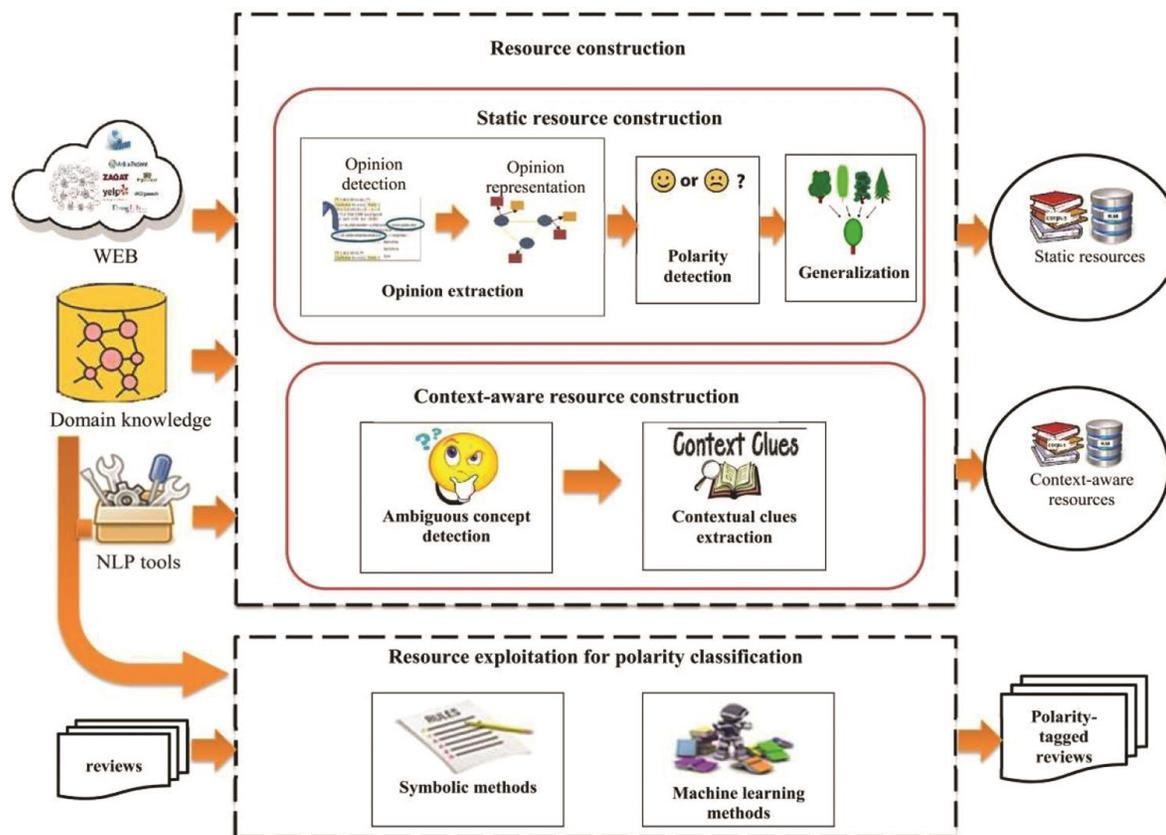
[3] www.nlm.nih.gov/research/umls/

Figure 1. A schematic view of the proposed frame work for indirect opinion mining.

2.1.2. Corpus construction

As mentioned earlier, supervised machine learning algorithms for polarity classification of opinions need to train a classifier on a polarity-tagged corpus, i.e., a corpus including sentences and phrases of specific polarities. We were inspired by distant supervision (Mintz et al., 2009) in constructing the corpus. In this method, instead of constructing a precise training set manually, a large set of automatically acquired data is used. The resulting training data will most likely be noisy. To construct a corpus of indirect opinions in the drug domain, a 3-step method will be introduced as shown in Figure 1. In the first step, the sentences indicating an indirect opinion are identified, and their positive and negative polarities are detected in the second step. In the third step, the generalization of the corpus is carried out.

• *Opinion extraction:* At first, it is assumed that any sentence containing effective and affected entities demonstrates an indirect opinion. However, this method does not elicit many indirect opinions, because most indirect opinions lack an effective entity. The users of review sites often comment on a particular entity (an entity whose name is mentioned in the title of the review), and thus they rarely mention the name of that

entity in their opinion texts. Therefore, in many opinions, only the benefits and/or the side effects of the drugs are referred to, and the names of the drugs are not mentioned. Accordingly, it is assumed that any sentence containing an affected entity exhibits an indirect opinion.

• *Polarity detection:* Domain knowledge, linguistic rules, and review structure are used (as described in Section 2.1.1) to detect the polarity of each sentence.

• *Generalization:* The selection of suitable features has a significant impact on the precision of machine learning approaches. In the field of opinion mining, several features have been introduced and applied that are often lexico-syntactic. In (Noferesti & Shamsfard, 2015a), we applied some common features of opinion mining (i.e., unigrams, bigrams, POS, and a binary feature that indicates whether the input sentence contains any sentiment shifter or not) along with a number of new features (i.e., phrase containing an affected entity, verb of the sentence, and ST as well as SG of the affected entity). The experiments revealed that applying these features enhances the precision of the polarity detection algorithm in the drug domain (Noferesti & Shamsfard, 2015a). Among these features, the two features ST and SG reflect a generalization of

the affected entity. When a corpus is being constructed, these features are maintained for each polar sentence. Figure 2 shows an overview of the proposed approach for corpus construction.

2.2. Construction of context-aware resources

In Figure 1, it can be seen that the construction of context-aware resources incorporates two major modules: ambiguous concept detection and contextual clue extraction. Domain knowledge can be applied in both modules. In the first module, a set of ambiguous concepts can be identified with the help of the domain knowledge. In the second module, the domain knowledge can be used to extract contextual clues. Also, by exploiting the built-in semantic knowledge in the domain-specific resources, we may develop a more abstract and richer model of the context. In the following, these modules will be described in more detail.

2.2.1. Identification of ambiguous concepts

Two methods are suggested to identify ambiguous concepts. The first method is based on the domain knowledge and the second one is corpus-based.

Identification of ambiguous concepts using domain knowledge: There are numerous resources on the web that contain the known side effects of drugs. *SIDER* is one of the richest resources in this area (Kuhn et al., 2010). It is a structured knowledge base capable of being processed by machine and it connects 888 drugs to 1450 known side effects. In *SIDER*, two contradictory concepts are sometimes known as a side effect. For example, the concepts of 'increased appetite' and 'loss of appetite' are both listed as side effects. In fact, these concepts are ambiguous since both an increase in and a decrease in appetite can be negative.

Based on the above discussion, the proposed approach for the identification of ambiguous concepts serves to extract opposing concepts known as side effects in the domain knowledge. For this purpose, the concepts in *SIDER* that have a common noun phrase are detected. Then, the opposing concepts are selected. Table 1 shows some of the concepts extracted by this method.

Table 1. Some opposing concepts extracted from SIDER.

increased appetite / decreased appetite
decreased libido / libido increased
glucose increased / decreased glucose
decreased white blood cell counts / white blood cell count increased
increased sweating / decreased sweating

Corpus-based approach to identifying ambiguous concepts: In (Weichselbraun et al., 2013a; 2014b), balanced occurrences of polar words (i.e., those in which polarity values assigned to them in a sentiment lexicon are higher than a predefined threshold) in both positive and negative opinions are used as indicators of ambiguity. According to the conducted tests, the above method does not have the required performance in the drug domain. The main reason for this is that sometimes an ambiguous concept occurs more frequently in the positive (or negative) set of opinions. For example, in most opinions, 'dry skin' is a negative effect for a drug. However, users sometimes refer to it as a positive effect. Therefore, another method is proposed to solve this problem.

To identify ambiguous concepts, the concepts of the knowledge bases created in Section 2.1.1 are searched in the polarity-tagged corpus of opinions expressed about the drugs. Whenever a concept has a positive (or negative) polarity in the knowledge base but has occurred in a negative (or positive) opinion in the corpus, it will be selected as an ambiguous concept. Of course, sentiment shifters are also considered in this step. Thus, a list of ambiguous concepts is obtained, some of which are shown in Table 2.

Table 2. Some ambiguous concepts extracted by the corpus-based approach.

<dry skin>
<lose weight>
<lower blood sugar>
<reduce appetite>

2.2.2. Extraction of contextual clues

The most obvious contextual clue is the text surrounding an ambiguous concept. However, contextual clues are not limited to textual clues. Social networks and review sites provide more information than the text of an opinion. This information is called side information or metadata and can be instrumental for a more precise analysis of opinions. For example, in the drug domain, 'type of disease', 'reason for drug consumption', or 'patient's condition when using a drug' can be effective clues. These clues are sometimes observed explicitly or implicitly. On some review sites, such as *www.druglib.com*, the user is asked to enter the reason(s) for taking a drug and the name of the disease afflicting him/her in a separate place. Even if this information is not available, most users will point it out in the review text. In fact, many hidden clues can be found by using information extraction techniques (Denecke & Bernauer, 2007).

The contextual clues used in this paper are textual clues besides metadata, i.e., complementary information the user gives to the system in addition to the text. For this purpose, the three fields of 'drug name', 'reason for drug consumption', and 'other conditions' available on *www.druglib.com* are exploited. Furthermore, the semantic information available in the domain knowledge is used to enrich the contextual clues. To this end, 'drug class' and 'disease category' extracted from *Drugbank* and *Diseasome* data sets, respectively, are placed in the contextual clues vector.

Contextual clues are extracted through the following procedure (Figure 3). Each ambiguous concept is first searched in the set of user reviews on revelant sites such as *www.druglib.com*. Then, the desired textual clues, metadata, and semantic features are extracted from the set of opinions containing the ambiguous concept. To extract text features, the words of the text containing the ambiguous concept are determined. Then, the stop words are removed, and any remaining words are stemmed and lower-cased. Finally, a vector is obtained from these words and their frequencies.
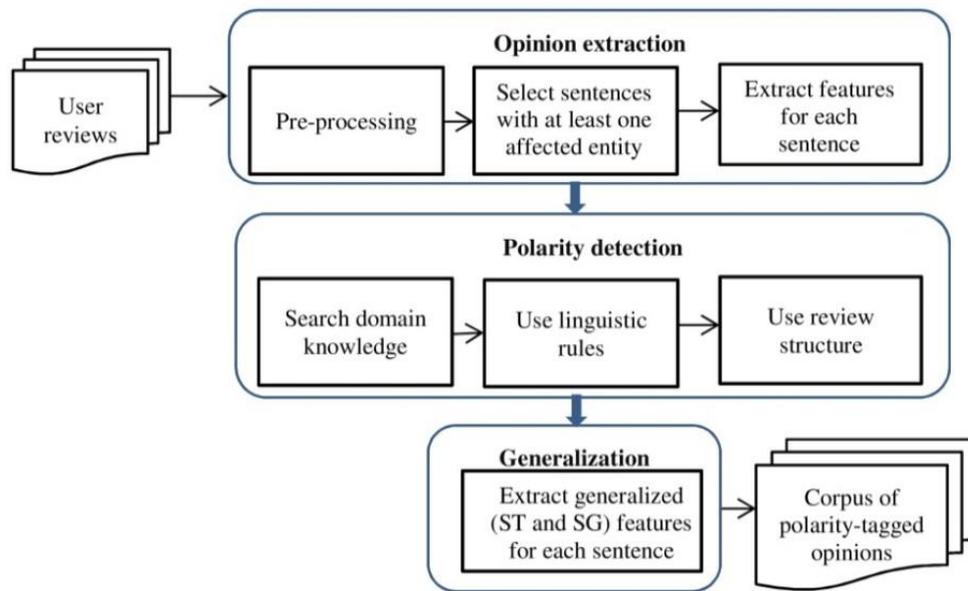

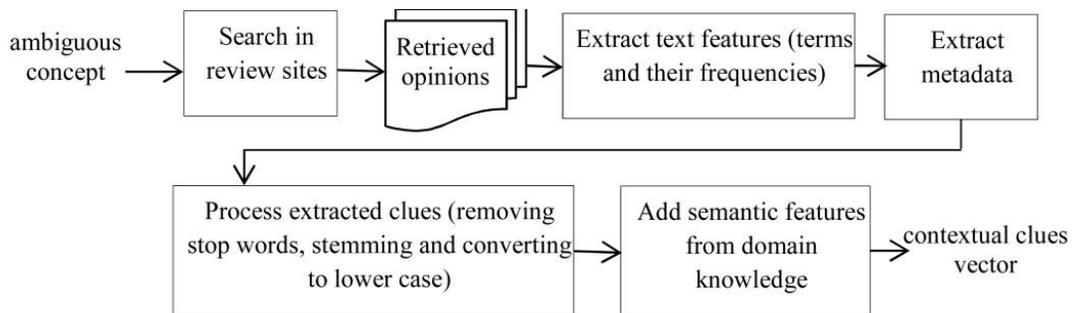Figure 2. An overview of the proposed approach for corpus construction.

Figure 3. Contextual clues extraction for the ambiguous concepts.

As mentioned earlier, to extract the metadata, we use the contents of the fields of 'drug name', 'reason for drug consumption', and 'other conditions'. The field 'drug name' contains only one word, while the other two fields may mention several reasons or conditions, which are usually separated by commas. The reasons and conditions are segregated and placed next to the name of the drug in the metadata vector.

In the end, textual clues, metadata, and semantic features are integrated into a feature vector that is used in the polarity detection algorithm. Contextual clues can be used in both symbolic methods and machine learning approaches. In the symbolic approaches, two feature vectors are constructed for each ambiguous concept by using positive and negative opinions containing that concept. These two feature vectors, together with the ambiguous concept, are kept in the knowledge base to be used for polarity detection of ambiguous concepts (see Section 2.3.2). In the machine learning approaches, contextual clues are applied as features for training the classifiers.

2.3. Resource exploitation for polarity classification of opinions

The following describes the methods of applying the constructed resources for polarity detection of opinions.

2.3.1. Application of static knowledge bases

To use the constructed knowledge bases, we first applied a relation extraction method to extract the triplet $< e_i, r_{ij}, e_j >$

from the input text. If the relation extraction module is able to extract a relation from the input text, then the following steps will be followed:

• For each triplet $< e_i, r_{ij}, e_j >$, two other triplets are extracted: ST triplet, which contains semantic types of entities, and SG triplet, which includes semantic groups of entities.
• These triplets will be searched in the knowledge base. If one of the triplets $< e_i, r_{ij}, e_j >$, ST, or SG exists in the knowledge base, the tuple polarity retrieved from that knowledge base is returned as the polarity of the input text. It is clear that polarity is reversed in the presence of a sentiment shifter in the input text.
• Otherwise, a semantic approach is suggested to enhance the coverage of the polarity classification algorithm. In this regard, when the three mentioned triplets do not exist, we search the knowledge base for triplets in which the verb (r_{ij}) is replaced

with one of its synonyms obtained from an external source of knowledge such as WordNet (Miller, 1998) and BabelNet (Navigli & Ponzetto, 2012).

2.3.2. Application of context-aware knowledge base

As noted earlier, in the context-aware knowledge base, two feature vectors are stored for each ambiguous concept. These vectors are used for polarity detection of the ambiguous concept. To do this, by using the Naïve Bayes technique (Eq. 1), the probability of positivity or negativity of the ambiguous concept C shown as $p(C_{+/-})$ is calculated depending on the probability vector of its contextual clues, i.e., $c = \{c_1, c_2, \ldots, c_n\}$. To detect the polarity of the ambiguous concept C, the (positive or negative) class that has the highest degree of probability will be selected.

$$p(C_{+/-}|c) = \frac{p(C_{+/-}) \cdot \prod_{i=1}^{n} p(c_i|C_{+/-})}{\prod_{i=1}^{n} p(c_i)} \tag{1}$$

Unlike the method introduced in (Weichselbraun et al., 2013a), the feature vector in the proposed method contains both metadata and semantic information, in addition to textual clues.

2.3.3. Corpus application

The constructed corpus is used as the training set in machine learning approaches for polarity detection of opinions. As already mentioned, other context features including 'drug name', 'reason for drug consumption', 'other conditions', 'drug class', and 'disease category' are also utilized to disambiguate ambiguous concepts in addition to the textual content.

3. Results

In this section, we evaluate the performance of both static resources and context-aware resources.

3.1. Assessment of static resources

By applying the proposed method in (Noferesti & Shamsfard, 2015b) for extracting polar facts from the domain knowledge, we built a knowledge base (called FactNet) containing 9,703 tuples, 1,436 ST patterns, and 224 SG patterns. Also, by applying the proposed method in (Noferesti & Shamsfard, 2015a) for extracting indirect opinions from user reviews, a knowledge base (called OpinionKB) containing 302 tuples, 44 ST patterns, and 10 SG patterns was created. The proposed methods achieved precisions of 92.26% and 88.08% for the construction of FactNet and OpinionKB, respectively (see (Noferesti & Shamsfard, 2015a; 2015b) for more details). A comprehensive evaluation of the proposed methods for constructing FactNet and OpinionKB is presented in (Noferesti & Shamsfard, 2015a; 2015b). In this section, however, we just focus on evaluating and comparing the proposed methods for polarity detection of opinions belonging to a test set.

Since there is no public dataset specifically designed for the problem at hand, we created a test set of 495 indirect opinions. This test set was collected from www.askapatient.com and *www.druglib.com*[4], two popular websites for reviewing drugs. This test set contained sentences and phrases not used in the construction of resources. Three annotators were asked to identify indirect opinions, and annotate them with positive or negative tags. Considering the obtained Fleiss' kappa (0.77) (Fleiss & Paik, 2013), we observed a substantial agreement between the annotators. At first, we only chose those sentences that were tagged by all the annotators. Then, to extend the test set, we asked two other annotators to tag all instances where disagreement occurred. Finally, we chose those instances that were tagged by the majority of the annotators. This test set was used in all the experiments that were conducted to evaluate the proposed methods.

Figure 4 compares the performance of the proposed methods of polarity classification through OpinionKB and FactNet knowledge bases. As can be seen, the precision of polarity classification via OpinionKB is higher compared to that of FactNet. Nevertheless, this degree of precision was achieved with much lower recall. The main reason for the low recall associated with OpinionKB is the low volume of knowledge stored in it. This problem could be overcome by reviewing more opinions and extending OpinionKB.

A combination of OpinionKB and FactNet resources is employed to detect the polarity of opinions. To this end, we developed three approaches: In the first approach specified as *combination approach1* in Figure 4, the test set of opinions are initially classified on the basis of polarity with the help of OpinionKB. Then, FactNet is applied to classify the polarity of the remaining opinions (opinions whose polarity was not determined in the previous step). In the second approach, the opposite of these steps is taken. In the third approach, all the test set of opinions are classified on the basis of polarity with the help of both OpinionKB and FactNet. Then, contradictory polarity-tagged opinions are discarded. It can be seen in Figure 4 that the greatest precision was obtained via the third approach, i.e., combining FactNet and OpinionKB (henceforth referred to as CFO). However, due to refraining from contradictions, this approach has a lower recall compared to the two other approaches. Furthermore, the combination of OpinionKB and FactNet for polarity classification of opinions helps attain a higher F-measure compared to independent use of these resources.

By using the method proposed in Section 2.1.2, we created a corpus of indirect opinions called OpinionCorpus. The characteristics of this corpus are presented in Table 3.
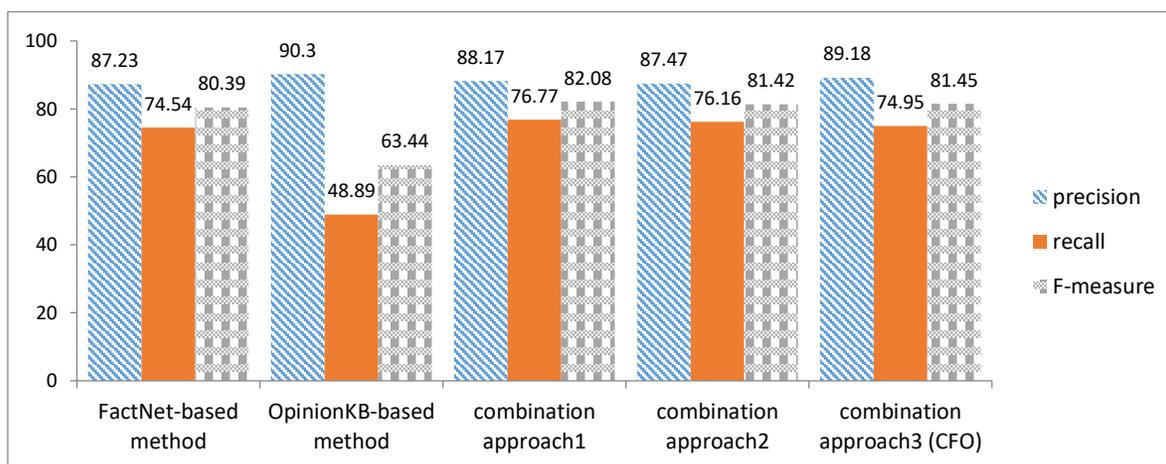

Figure 4. Results of polarity classification using FactNet and OpinionKB.

[4] Accessed March 2017.

Table 3. Characteristics of the constructed corpus.

No. of sentences	3856
No. of words	35064
No. of positive sentences	1332
No. of negative sentences	2524

Different pre-processing trends were followed to train the machine learning classifiers on the constructed corpus. Also, various features were used for classification. The experiments demonstrated that the highest precision for polarity detection is obtained by implementing the pre-processing sequence of removing the stop words, converting the text into lowercase, stemming, and applying the following set of features: unigrams, bigrams, POS features, the main verb of the sentence and its affected entity, semantic type and semantic group of the affected entity, as well as a binary feature that indicates whether the sentence includes a sentiment shifter or not. In addition, by applying different machine learning classifiers such as Naïve Bayes (NB), Support Vector Machine (SVM), Maximum Entropy (MaxEnt), and Decision Tree, we found that SVM provides the highest precision. The precision of the machine learning approach that uses SVM algorithm and the above-mentioned features is displayed in Figure 5.
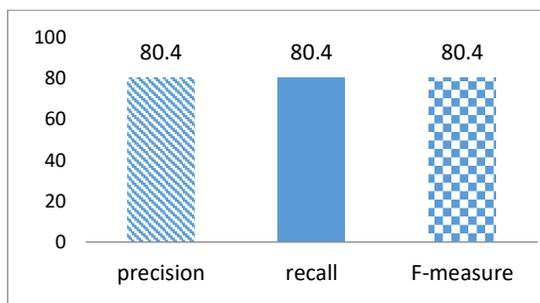

Figure 5. Performance of using OpinionCorpus to train SVM classifier for polarity classification of drug reviews.

Figures 4 and 5 show that the knowledge-based methods achieved greater precision in classifying the polarity of opinions compared to the machine learning approaches. Nonetheless, the recall of machine learning approaches was higher compared to the knowledge-based methods. In fact, one of the advantages of machine learning techniques is the classification of samples not already seen, while knowledge-based methods are only able to classify samples that can match with the existing samples or patterns in the knowledge base. However, one of the strengths of knowledge-based methods is that samples and patterns are assessed by humans. Besides, the extracted patterns can be used for other purposes. For example, FactNet patterns can help identify drug benefits and side effects in the texts.

In the final experiment, the proposed methods were compared with several baseline and existing methods. As the baseline, a lexicon-based method and a distant supervision approach were selected. In the lexicon-based method, SentiWordNet (Esuli & Sebastiani, 2007) was applied as a well-known and popular lexicon in the field of opinion mining. In this method, to classify the text polarity, we first extracted the polarity of its words from SentiWordNet. Then, based on the polarity of the individual words, the total polarity of the text was detected. Two common methods for calculating the overall polarity of a text are as follows: (1) majority voting, which counts the number of positive and negative words of the text and selects the highest number, and (2) sum of predictions, in which text polarity is computed as the sum of polarity values of its words. It needs to be added that we also considered sentiment shifters in this experiment.

In the distant supervision approach, it was assumed that the opinions written on the benefits of a drug were positive, and those written on its side effects were negative. Thus, a noisy training corpus of positive and negative opinions was constructed. Together with the features introduced by Pang et al. (2002), i.e., unigrams (including a specific word presence and frequency), bigrams, POS, and adjectives, this corpus was applied to train the classifiers NB, SVM, and MaxEnt. These classifiers were chosen because of their frequent use in the analysis of direct opinions and their effectiveness. The conducted experiments showed that using the SVM classifier could lead to maximum performance. Figure 6 compares the performances of the proposed and baseline methods in detecting the polarity of opinions in the test set. As can be seen, the CFO method achieved a higher precision in the polarity detection of indirect opinions.

Figure 7 compares the precision of the proposed methods and the three existing methods for polarity classification of indirect opinions of the test set. The first method in this area was suggested by Cambria et al. (2012), and it was successfully applied as a criterion for assessing health care quality and classifying patients' opinions. This method first extracts a number of concepts from the input text by using a semantic parser. Then, it employs SenticNet (Cambria et al., 2010) as a rich concept-level source in opinion mining for assigning polarity to the extracted concepts. Finally, the overall polarity of the text is calculated by averaging the polarity values of its concepts. We implemented this method using the sematic parser and SenticNet 4 (Cambria et al., 2016), which are available on *sentic.net*.

The second existing method is a supervised machine learning approach introduced by Habernal et al. (2015). Here, different methods are evaluated for text pre-processing, and numerous features are used for classification. Next, the effects of different feature selection and classification algorithms are assessed. By using this method, the best performance was

achieved with the following settings. The pre-processing step involved tokenization, stemming, and lowercasing the text. The set of features applied included unigrams, bigrams, and POS-based features. The emoticons were withdrawn since they rarely occur in the domain of medicine. Among the different feature selection methods, *Mutual Information* yielded the best results and slightly improved the classification precision. The weighting method introduced on the basis of TF-IDF[5] did not improve the performance and hence, was skipped. In addition, SVM led to better results compared to NB and MaxEnt. To implement the approach proposed by Habernal et al. (2015), we used WEKA 3.6.9[6] in the default setting.

The third existing method is sentic patterns (Poria et al., 2014), a concept-level method that merges linguistics, common-sense computing, and machine learning to classify the polarity of opinions. We used Sentic demo[7] for polarity classification of opinions in the test set. Figure 7 shows that the proposed methods outperform all the three current methods.

3.2. Context-aware resource assessment

At first, seven ambiguous concepts including 'increasing appetite', 'lowering of blood sugar', 'lowering of blood pressure', 'weight loss', 'weight gain', 'drying skin', and 'increasing libido' obtained by the methods proposed in Section 2.2.1 were considered. Then, each of these concepts was searched in *www.druglib.com* and www.askapatient.com. During this search, user reviews containing the related concept were retrieved. Accordingly, a corpus of 294 opinions containing ambiguous concepts was created. For each concept, the numbers of positive and negative opinions available in the corpus were considered to be identical.

To construct a context-aware knowledge base, we divided the above-mentioned dataset into two parts. One part was used to construct the knowledge base, and the other one was applied as a test set. To divide the data set, we used the 10-fold cross validation technique. Then, two feature vectors were extracted for each ambiguous concept in the test set. As noted earlier, the feature vector consists of three parts: metadata, semantic features, and review text.

Figure 8 exhibits the performance of the polarity detection algorithm in different modes: when only textual clues were used; when textual clues were applied together with the metadata; and when the semantic features extracted from the domain knowledge were added to the textual clues and the metadata. As can be seen, the combination of various contextual clues dramatically increased the precision and recall of the disambiguation algorithm. In fact, adding the metadata to the text features enhanced the precision and recall of the disambiguation algorithm from 53.45% and 49.68% to 59.26% and 53.47%, respectively. Furthermore, the addition of semantic features led to improved performance of the polarity detection algorithm for the ambiguous concepts.

The above-mentioned dataset was randomly divided into two parts. The first part containing 202 opinions was used to construct the knowledge base. The second part of 92 remaining opinions was applied to create a test set. The test set was a balanced set, in which the numbers of positive and negative opinions for each ambiguous concept were the same. According to this test set, the precision of the proposed method was 64.29%. Therefore, the proposed method yielded better results as compared to the baseline method of polarity detection on the basis of static sentiment words. Since the numbers of positive and negative opinions for each concept were identical in the test set, the baseline method allocated correct tags only to half of the opinions. Thus, its precision was 50%.

In the final experiment, for polarity classification via machine learning techniques, we used the 10-fold cross validation to divide the dataset into a training and a testing part. Figure 9 displays the performance of SVM and NB classifiers in detecting the polarity of ambiguous concepts through different combinations of the features introduced in Section 2.2.2. These combinations included only text features introduced in (Weichselbraun et al., 2013a), only metadata, a combination of text features and metadata, and the integration of semantic clues with text features and metadata.

As shown in Figure 9, the combination of text features and metadata achieved higher precision as compared to those with the basic methods, i.e., the sole use of either textual clues or metadata. Specifically, the combination of text features and metadata in the SVM classifier led to an approximate increase of 3% in precision compared to when only text features were used. Also, the addition of semantic features to the combination of text features and metadata improved the precision of the NB classifier by 0.68%.

---

[5] Term Frequency-Inverse Document Frequency.
[6] www.cs.waikato.ac.nz/ml/weka/downloading.html
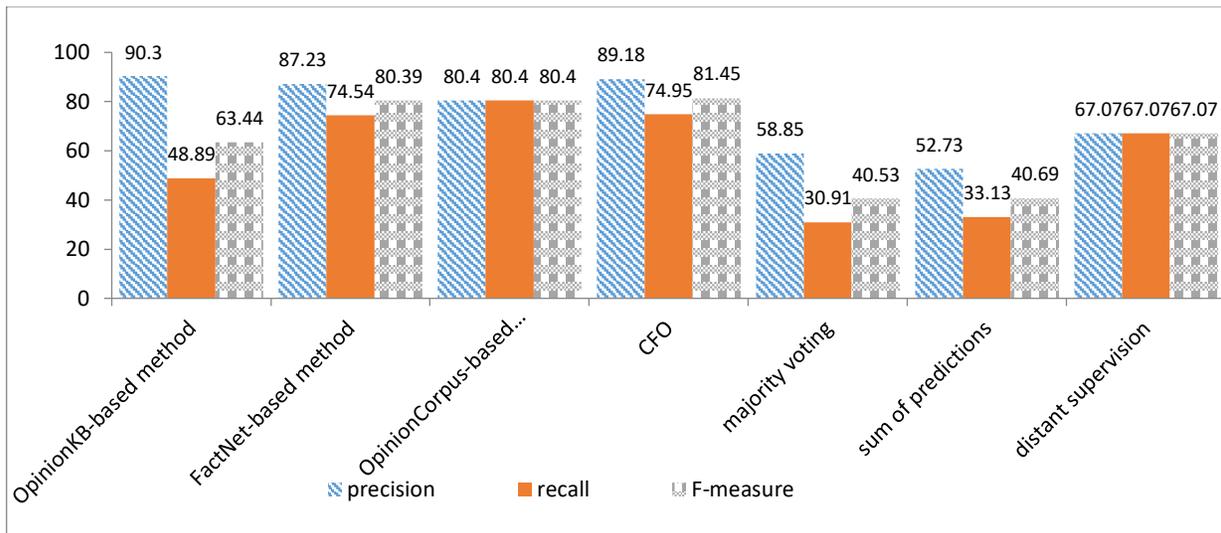[7] Sentic.net/demo.

Figure 6. Comparison of the proposed methods for polarity
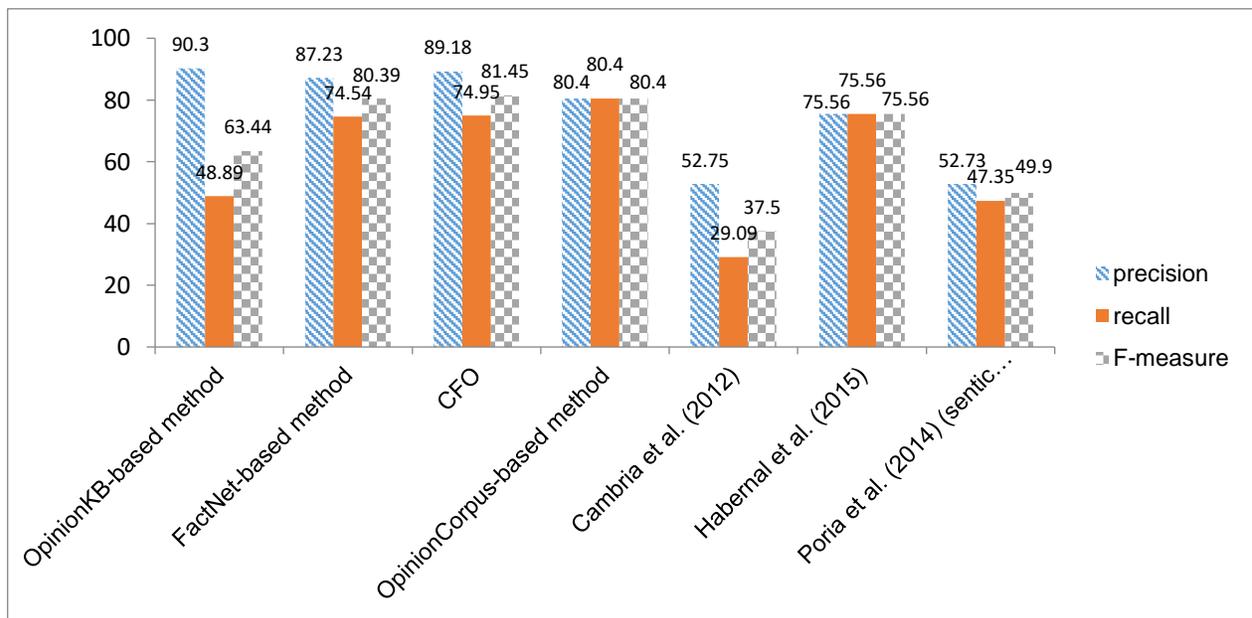classification with the baseline methods.

Figure 7. Comparison of the proposed methods for polarity
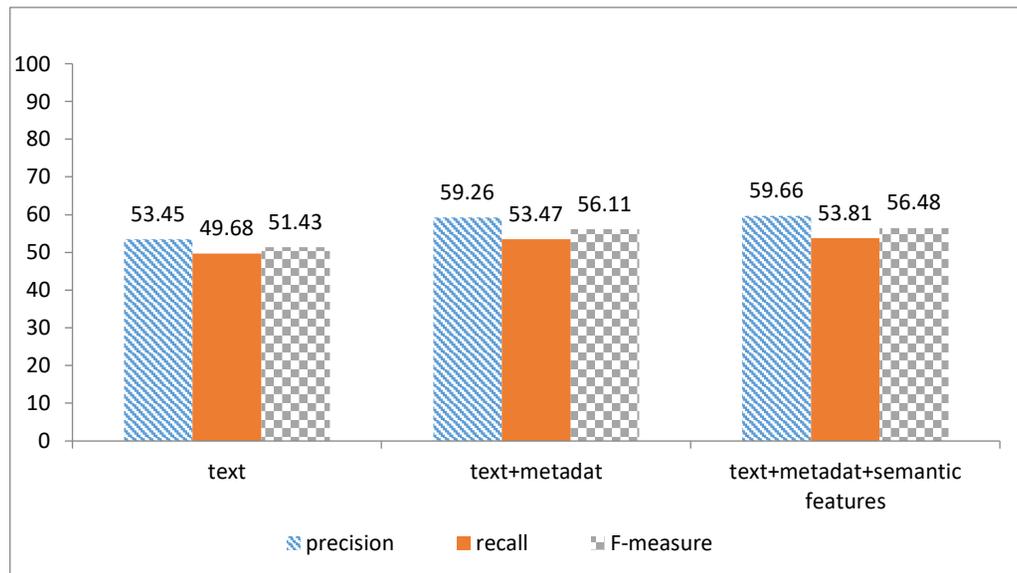classification with the existing methods.

Figure 8. Performance of the proposed knowledge-based method for
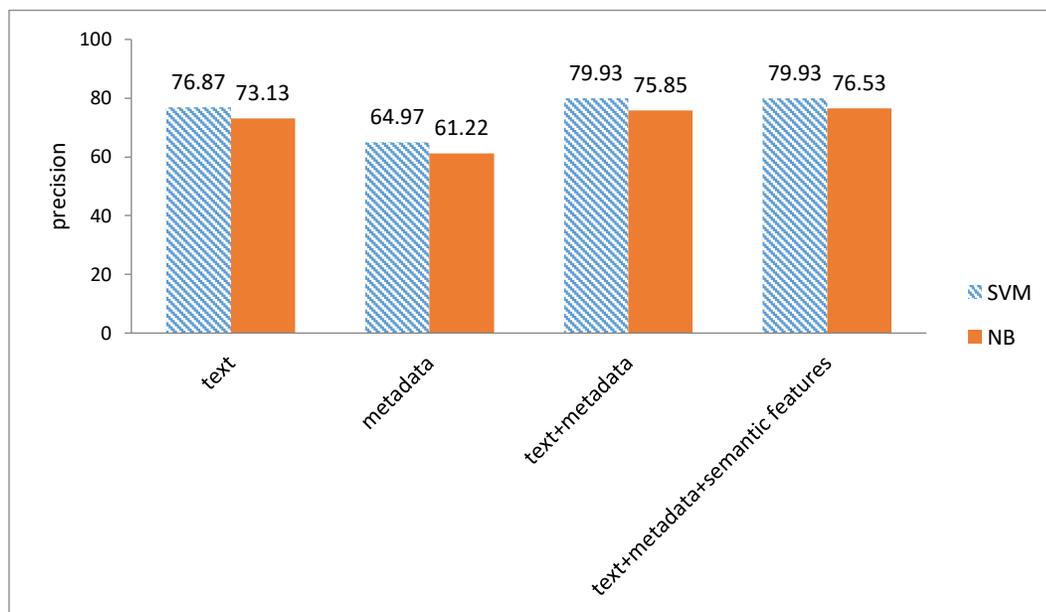polarity classification of ambiguous concepts.

Figure 9. Precision of the proposed method for adding contextual
clues in machine learning approaches.

4. Discussion

Experimental results show that the constructed resources are appropriate for polarity classification of drug reviews. As can be seen in Figure 7, compared to the existing methods, our proposed methods achieve a better precision in classifying the polarity of opinions. In fact, the existing methods are generally insufficient for predicting the polarity of indirect opinions in the drug domain. A major problem of current opinion-mining systems is that the resources they used either lack indirect opinions (e.g., the cases in which the facts have been withdrawn and only subjective sentences have been examined) or are sparse in terms of the number of facts and indirect opinions. In particular, the training corpora available for use in machine learning approaches are very sparse in terms of indirect opinions. This causes a decline in the precision of the classification algorithm, which is highly dependent on the volume and quality of training data.

More accurately, the lexicon-based methods suffer from two major problems in the analysis of indirect opinions. First, most of the existing lexicons are at word level. This means that in these lexicons, polarity is assigned to each word or each of its senses. This is while the polarity and meaning of a sentence are not conveyed through words alone, and the interaction between words is also effective in detecting the polarity of a sentence. An indirect opinion indicates the effect of an entity on another entity, and this effect will be often expressed by the verb of the sentence. Therefore, the verb of a sentence has a significant impact on the polarity detection of that sentence. Consider the following examples:

'This drug eliminated my acne completely'.

'This drug reduced my pain'.

Lexicon-based methods fail to correctly detect the polarity of such examples since terms such as 'acne' and 'pain' are negative in the existing lexicons like SentiWordNet. This causes the above two sentences to be considered as negative, while the overall polarity of both sentences is positive. In general, in the medical domain, most words such as 'disease', 'fever', and 'pain' are negative, whereas these words frequently occur in positive sentences. Thus, lexicon-based methods do not have sufficient precision in such domains.

The second problem is related to sentences that do not contain explicit sentiment words. Consider the following example:

'This drug decreased my vision'.

In SentiWordNet, the term 'vision' lacks a polarity. This makes the above example to be considered as neutral, while it is clear that the polarity of this sentence is negative.

Also, in conventional methods of constructing concept-level resources, the verb of a sentence is often withdrawn when extracting the concepts, even though it plays a vital role in the analysis of many indirect opinions. Thus, this will limit the positive effect of using indirect opinions in the opinion mining systems.

In addition, the existing methods of developing concept-level knowledge bases for opinion mining frequently rely on assigning polarity to the concepts available in the common-sense knowledge bases such as ConceptNet and Open Mind (Cambria & Hussain, 2012; Tsai et al., 2013). Although these resources are rich in terms of common-sense concepts, technical and domain-specific concepts play a major role in opinion mining. Moreover, as noted above, besides the concepts that represent text entities, the verb of a sentence is essential in the polarity classification of indirect opinions and, therefore, requires special attention.

Eventually, it should be mentioned that previous works on context-aware opinion mining (Weichselbraun et al., 2013a; 2014b) have only used the text surrounding as an ambiguous concept. These methods suffer from two problems:

1. Sometimes, consecutive sentences do not have opinions of the same polarity. In such cases, the use of the surrounding sentences for disambiguation can be misleading.

2. Sometimes, the text of an opinion only contains an ambiguous sentence or phrase. In these cases, the current methods do not have the necessary performance due to the sparsity of the opinion text.

To solve this problem in the proposed framework, we used various contextual clues simultaneously to detect the polarity of ambiguous concepts. The experimental results show that the precision of disambiguation algorithm is higher when a combination of text features and metadata is applied as opposed to the single use of textual clues.

5. Conclusions

In this paper, a semantic framework based on domain knowledge was presented to construct and exploit resources for opinion mining of drug reviews. Within this framework, two methods were suggested for building knowledge bases of indirect opinions by using domain knowledge and user reviews, and a method was presented for the semi-automatic construction of a corpus of indirect opinions. The three methods were integrated afterward. In addition, an approach was proposed to create context-aware resources in order to disambiguate ambiguous concepts. Ultimately, the application method of each constructed resource was presented to facilitate polarity detection of indirect opinions.

In the proposed methods, domain knowledge plays a key role in identifying the concepts and domain-specific entities, grouping them, extracting the relationship between them, and classifying their polarity. This knowledge is also applied in the generalization module to determine the semantic class of the concepts, extract semantic patterns from them, and find their semantic similarities for classifying their polarities.

The results of the experiments conducted to evaluate the proposed methods revealed that using the combined knowledge base significantly enhanced the precision of the polarity detection algorithm compared to the baseline scenarios and existing methods. Furthermore, applying the constructed corpus and the proposed semantic features in machine learning approaches to detect polarity of indirect opinions led to much better results than those obtained by the existing methods. The tests carried out to evaluate the performance of context-aware resources also indicated that the simultaneous use of various contextual clues boosted polarity detection of ambiguous concepts both in symbolic and machine learning approaches.

Conflict of interest

The authors do not have any type of conflict of interest to declare.

Financing

The authors received no specific funding for this work.

References

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236-246. https://doi.org/10.1016/j.eswa.2017.02.002

Asghar, M. Z., Ahmad, S., Qasim, M., Zahra, S. R., & Kundi, F. M. (2016). SentiHealth: creating health-related sentiment lexicon using hybrid approach. *SpringerPlus*, *5*(1), 1139. https://doi.org/10.1186/s40064-016-2809-x

Bobicev, V., Sokolova, M., & Oakes, M. (2015). What goes around comes around: learning sentiments in online medical forums. *Cognitive Computation*, *7*(5), 609-621. https://doi.org/10.1007/s12559-015-9327-y

Cambria, E., & Hussain, A. (2012). Sentic computing: Techniques, tools, and applications. *A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, Springer Science & Business Media*. https://doi.org/10.1007/978-94-007-5070-8

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In 2010 AAAI fall symposium series.

Cambria, E., Benson, T., Eckl, C., & Hussain, A. (2012). Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. *Expert Systems with Applications*, *39*(12), 10533-10543. https://doi.org/10.1016/j.eswa.2012.02.120

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (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).

Denecke, K., & Bernauer, J. (2007). Extracting specific medical data using semantic structures. In *Conference on Artificial Intelligence in Medicine in Europe* (pp. 257-264). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-73599-1_35

Esuli, A., & Sebastiani, F. (2007). SentiWordNet: a high-coverage lexical resource for opinion mining. Evaluation, 17,1-26.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). Statistical methods for rates and proportions. john wiley & sons.

Goeuriot, L., Na, J. C., Min Kyaing, W. Y., Khoo, C., Chang, Y. K., Theng, Y. L., & Kim, J. J. (2012). Sentiment lexicons for health-related opinion mining. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (pp. 219-226). https://doi.org/10.1145/2110363.2110390

Gopalakrishnan, V., & Ramaswamy, C. (2017). Patient opinion mining to analyze drugs satisfaction using supervised learning, *Journal of Applied Research and Technology, 15*(4), 311-319. https://doi.org/10.1016/j.jart.2017.02.005

Greene, S., & Resnik, P. (2009). More than words: Syntactic packaging and implicit sentiment. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 503-511).

Habernal, I., Ptáček, T., & Steinberger, J. (2014). Supervised sentiment analysis in Czech social media, *Information Processing & Management*, *50*(5), 693-707. https://doi.org/10.1016/j.ipm.2014.05.001

Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, *23*(1), 11. https://doi.org/10.3390/mca23010011

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. https://doi.org/10.1016/j.knosys.2013.11.009

Inui, K., Abe, S., Hara, K., Morita, H., Sao, C., Eguchi, M., ... & Matsuyoshi, S. (2008). Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 314-321). IEEE. https://doi.org/10.1109/WIIAT.2008.373

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, *6*(1), 343. https://doi.org/10.1038/msb.2009.98

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1),1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Miller, G. A. (1998). WordNet: An electronic lexical database. MIT press.

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 1003-1011).

Na, J. C., & Kyaing, W. Y. M. (2015). Sentiment analysis of user-generated content on drug review websites. *Journal of Information Science Theory and Practice*, *3*(1), 6-23. https://doi.org/10.1633/JISTaP.2015.3.1.1

Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial intelligence*, *193*, 217-250. https://doi.org/10.1016/j.artint.2012.07.001

Noferesti, S., & Shamsfard, M. (2015a). Resource construction and evaluation for indirect opinion mining of drug reviews. *PloS one*, *10*(5), e0124993 https://doi.org/10.1371/journal.pone.0124993

Noferesti, S., & Shamsfard, M. (2015b). Using Linked Data for polarity classification of patients' experiences, *Journal of biomedical informatics*, 57, 6-19. https://doi.org/10.1016/j.jbi.2015.06.017

Noferesti, S., & Shamsfard, M. (2016). Automatic building a corpus and exploiting it for polarity classification of indirect opinions about drugs. Signal and data processing, 2, 35-49.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10, 79-86. https://doi.org/10.3115/1118693.1118704

Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, *69*, 45-63. https://doi.org/10.1016/j.knosys.2014.05.005

Tsai, A. C. R., Wu, C. E., Tsai, R. T. H., & Hsu, J. Y. J. (2013). Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intelligent Systems*, *28*(2), 22-30. https://doi.org/10.1109/MIS.2013.25

Weichselbraun, A., Gindl, S., & Scharl, A. (2013a). Extracting and grounding contextualized sentiment lexicons, *IEEE Intelligent Systems,* 2, 39-46. https://doi.org/10.1109/MIS.2013.41

Weichselbraun, A., Gindl, S., & Scharl, A. (2014b). Enriching semantic knowledge bases for opinion mining in big data applications, *Knowledge-based systems*, 69, 78-85. https://doi.org/10.1016/j.knosys.2014.04.039

Wilson, T. 2008. Annotating Subjective Content in Meetings. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA).

Zhang, L., & Liu, B. (2011). Identifying noun product features that imply opinions. In Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies (pp. 575-580).