Journal of Biomedical Informatics 57 (2015) 6–19

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Using Linked Data for polarity classification of patients' experiences

Samira Noferesti *, Mehrnoush Shamsfard

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

ARTICLE INFO

Article history:
Received 21 October 2014
Revised 9 June 2015
Accepted 24 June 2015
Available online 23 July 2015

Keywords:
Sentiment analysis
Opinion mining
Patient experience mining
Polarity classification
Linked Data
Drug reviews

ABSTRACT

Polarity classification is the main subtask of sentiment analysis and opinion mining, well-known problems in natural language processing that have attracted increasing attention in recent years. Existing approaches mainly rely on the subjective part of text in which sentiment is expressed explicitly through specific words, called sentiment words. These approaches, however, are still far from being good in the polarity classification of patients' experiences since they are often expressed without any explicit expression of sentiment, but an undesirable or desirable effect of the experience implicitly indicates a positive or negative sentiment.

This paper presents a method for polarity classification of patients' experiences of drugs using domain knowledge. We first build a knowledge base of polar facts about drugs, called FactNet, using extracted patterns from Linked Data sources and relation extraction techniques. Then, we extract generalized semantic patterns of polar facts and organize them into a hierarchy in order to overcome the missing knowledge issue. Finally, we apply the extracted knowledge, i.e., polar fact instances and generalized patterns, for the polarity classification task. Different from previous approaches for personal experience classification, the proposed method explores the potential benefits of polar facts in domain knowledge aiming to improve the polarity classification performance, especially in the case of indirect implicit experiences, i.e., experiences which express the effect of one entity on other ones without any sentiment words.

Using our approach, we have extracted 9703 triplets of polar facts at a precision of 92.26 percent. In addition, experiments on drug reviews demonstrate that our approach can achieve 79.78 percent precision in polarity classification task, and outperforms the state-of-the-art sentiment analysis and opinion mining methods.

1. Introduction

Nowadays, people care deeply about their health and wellness. Studies reveal that people are increasingly using social media, such as blogs, forums and review sites, in order to share their health-related information, including health conditions, diseases, and medicines they take, as well as outcomes and side-effects that they experience [1,2]. This makes the Web a rich source of patients' experiences and opinions, which can be valuable for healthcare providers to provide better services, and for laypeople to be aware of others' opinions and to benefit from their experiences to make informed decisions before using a service or product. In particular, in the drug domain, Brownstein [3] claims that mining user-generated content about drugs can provide pharmaceutical companies with valuable information about the effects and side effects of new drugs. However, the ever-increasing volume of user-generated content on the Web has raised the demand for developing automatic methods of analyzing experiences and opinions and discovering hidden knowledge from unstructured text data. Sentiment analysis and opinion mining, which are sometimes used interchangeably, are the fields of study that aim to automatically extract and classify people's opinions, emotions, sentiments, evaluations, and appraisals toward entities. Opinion mining has the potential to provide the means for flows of patients' opinions and understanding the sentiment of their experiences [4].

An important subtask of opinion mining is polarity classification, which classifies texts into different categories (usually, positive or negative). Polarity classification has attracted increasing attention in recent years. Although much research has been performed in this area, most existing approaches have focused on the subjective part of text in which sentiment is explicitly expressed through the use of specific words called sentiment words (e.g., 'This drug is effective'). However, sentiments are not only expressed in subjective statements, but can also be stated in

* Corresponding author at: NLP Lab., Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran. Tel.: +98 9151630480.

E-mail address: snoferesti@ece.usb.ac.ir (S. Noferesti).

http://dx.doi.org/10.1016/j.jbi.2015.06.017
1532-0464/

objective or factual sentences (e.g., 'Accutane dried my skin') or in ironic expressions [5] which are quiet difficult to tackle. In other words, a piece of text could have polarity without being subjective; for example a patient experience without any sentiment words can be classified as positive or negative.

In the medical sphere in particular, it is very common for users to express their opinions indirectly. In the drug domain, patients usually write about their experiences of drug effectiveness or side effects instead of expressing a direct opinion using explicit sentiment words. Patients' experiences are often expressed 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.

In this paper, we use the term 'polar fact' for factual information that has polarity. Some researchers have discussed the importance of analyzing polar facts. Wilson [6] shows that sentiments can be conveyed indirectly through factual pieces of text and opinionated texts can also contain polar facts that imply an argument or evaluation for an idea. Inui et al. [7] say that factual information can indicate objective reasons for opinions, and hence are vital for decision making. Feldman [8] expresses that polar facts should be analyzed to acquire a more efficient sentiment analysis. However, few studies have been conducted on the polarity classification of factual statements. These studies are mainly based on machine learning approaches relying on manually annotated corpora of subjective, as well as factual, statements. In this paper, we propose a different method that explores the use of existing information in Linked Data sources to semi-automatically build a knowledge base of polar facts about drugs, called FactNet, which aims to be a resource for polarity classification of patients' experiences.

Linked Data refers to a set of best practices for exposing, sharing and linking structured data on the Web using Semantic Web technologies such as Uniform Resource Identifier (URI) and Resource Description Framework (RDF) [9]. The Linked Open Data (LOD) project[1] aims at publishing and connecting Linked Datasets on the Web according to Linked Data principles. LOD contains billions of facts for various domains, and so can be used as a rich knowledge base in different tasks. Recently, there have been an increasing number of studies aimed at exploiting the potential of Linked Data in different applications such as information extraction [10] and word sense disambiguation [11]. Given this insight, this paper shows how Linked Data can be exploited to alleviate the data acquisition bottleneck for extraction of polar facts.

We then propose a method for learning generalized patterns of polar facts in order to overcome the missing knowledge issue, i.e., not all polar facts are contained in LOD. Exploiting the generalized patterns, our proposed method for polarity classification is able to classify unseen polar facts.

Finally, we present a method for polarity classification of patients' experiences using FactNet. The contributions of this paper can be summarized as follows:

(1) We explore the potential benefits of polar facts in domain knowledge aiming to improve the polarity classification performance, especially in the case of indirect implicit experiences, i.e., experiences which express the effect of one entity on other ones (indirect) without any sentiment words (implicit). In contrast to the problem of mining direct opinions (i.e., opinions that are expressed directly on an entity or one of its aspects) for which there has been a great research, the problem of mining indirect opinions is almost unexplored.
(2) We propose a novel method for polarity classification of patients' experiences that exploits publicly available Linked Data sources as background knowledge to construct a knowledge base of polar facts called FactNet. As far as we know, this is the first work to extract polar facts from LOD.
(3) We propose a method to mine generalized patterns of polar facts aims at dealing with issues raised by the missing knowledge in FactNet, and hence enhance the knowledge discovery process.
(4) We also provide comprehensive evaluation of each part to show the effectiveness of the proposed methods.

The rest of this paper is organized as follows. Section 2 provides an overview of related research in the area of opinion mining, as well as in the area of experience mining. Section 3 presents some motivating examples to illustrate how the proposed approach improves existing ones. Section 4 presents the details of the proposed approach for polarity classification of patients' experiences. Experimental results are reported in Section 5. Section 6 concludes the paper and outlines future directions.

2. Related work

In this section, we briefly discuss the related work in the two areas to which our current research is related: experience mining and opinion mining.

2.1. Experience mining

Experience mining aims to extract and analyze personal experiences expressed in natural language text [4]. Experience mining is an unexplored open research problem that is motivated to be an extension of opinion mining. The term experience mining was first used by Inui et al., in [7]. In this work, four technical challenges for experience mining were presented: event mention extraction, entity–event relation extraction, factuality analysis, and experiencer identification. The authors also proposed a machine learning-based approach for factuality analysis. Chen [12] modeled two dimensions of patient experience in a health-related social networking site, DailyStrength:[2] interpersonal interactions and medication use. Jijkoun et al. [13] introduced an annotation schema for labeling experience as reporting no experience, reporting an off-target experience, or reporting an on-target experience. They also presented some linguistic features that are well suited for experience mining. Murphy [14] used machine learning approaches for identification of Twitter posts that contain cancer-related patients' experiences.

Existing approaches to experience mining are mainly based on supervised learning of experiences from hand-labeled corpora. Nevertheless, to the best of our knowledge, there is no publicly available annotated corpus of patients' experiences about drugs. Therefore, we propose an alternative approach for experience classification that does not require training data.

2.2. Opinion mining

Opinion mining has been a very active research area in the past decade. Different methods have been proposed to tackle different tasks and with diverse applications. Pang and Lee [15], Liu [16] and Cambria and Hussain [17] have provided comprehensive surveys of important research in this field. In the following subsections, we first review related work on polarity classification task.

[1] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

[2] www.dailystrength.org.

Then, we present the most relevant studies related to opinion mining in the medical domain.

2.2.1. Polarity classification

Traditional approaches that have been proposed for polarity classification can be classified into two main categories: dictionary-based approaches, which mainly focus on the construction and use of word-level sentiment lexicons such as SentiWordNet [18] or concept-level dictionaries such as SenticNet [19], and machine learning approaches [20–22], which depend upon the availability of an annotated corpus with polarity labels to detect the polarity of new examples. Both dictionary-based and machine learning approaches mainly rely on subjective parts of text in which sentiments are explicitly expressed, e.g., sentiment words and their co-occurrence frequencies, while most polar facts have no explicit sentiment word. In other words, these approaches neglect objective statements that carry sentiment and, hence, are ineffective for the polarity classification of patients' experiences where such statements occur frequently.

Current concept-level dictionaries are able to handle some of implicit expressions of sentiments based on common-sense knowledge. However, they are insufficient for polarity classification of drug reviews since the medical domain is highly technical and, hence, domain knowledge plays a critical role in polarity classification in this domain. In other words, available sentiment dictionaries do not have enough vocabulary to handle polar facts about drugs.

Although some researchers have discussed the importance of analyzing polar facts for a more efficient sentiment analysis [8,16], there have been few attempts at analyzing factual statements. The focus of these attempts has been on annotating a corpus of subjective as well as objective statements, which aims to be a training set for supervised machine learning techniques [6,23]. Wilson [6] proposed AMIDA schema which contains subjective statements, as well as objective polar utterances, i.e., negative or positive factual information about an entity, without indicating any explicit sentiment expression. Toprak et al. [23] annotated a corpus of consumer reviews that contains subjective expressions as well as polar facts.

Machine learning approaches suffer from several problems. Firstly, they require large annotated corpora, which are tedious, expensive and time consuming to construct. Secondly, due to the fact that labeled data are on a particular corpus, the resulting classifiers tend to bias toward that text domain. Finally, even when machine learning approaches are employed to learn polar facts from corpora, the performance is far from satisfactory. The main reason is that they are semantically weak, meaning that with the exception of obvious sentiment words they have little predictive value individually [17].

To overcome the above mentioned problems, the current research explores the usage of Linked Data sources for extraction of polar facts.

2.2.2. Opinion mining in medical domain

Existing researches for opinion mining in medical domain are often about a specific problem, and mainly rely on machine learning techniques. Greaves et al. [24] assessed patients' opinions about different performance aspects of hospitals in the United Kingdom. Tanvir et al. [25] used a subjectivity lexicon and machine learning algorithms to sentiment analysis of messages posted on forums dedicated to hearing loss. Sokolova and Bobicev [26] used supervised machine learning methods to analyze sentiments and opinions expressed in health-related user-written texts. Wang et al. [27] proposed a combination of machine learning and rule-based classifiers for sentiment analysis in suicide notes. Parker et al. [28] proposed a method for predicting public health trends via Twitter.

There are also some studies on opinion mining in the drug domain. Na et al. [29] proposed a rule-based system for polarity classification of drug reviews. Yalamanchi [30], in his thesis, developed a system called SidEffective to analyze patients' sentiments about a particular drug. Goeuriot et al. [31] built a health-related sentiment lexicon and used it for polarity classification of drug reviews. Wiley et al. [32] used SentiWordNet and NRC word-emotion [33] lexicons for polarity and emotion classification of health-related content of online social networks, respectively.

In a close stream of research, some studies have been conducted on polarity analysis of relationships extracted from biomedical literature articles. Yang et al. [34] proposed a supervised machine learning approach for polarity and strength prediction. They first introduced new features designed to capture lexical, semantic and structural characteristics of a relation, and then proposed a wrapper-based approach for feature selection. Then, they used SVM classifier and SVR predictor to polarity and strength analysis. Miao et al. [35] presented the main challenges related to the relation polarity analysis. In addition, they propose a combined approach based on background knowledge and domain-specific training data to polarity classification of relationships between foods and medical conditions.

This paper presents a different method to automatically build a knowledge base that contains instances and patterns of polar facts, from LOD. This knowledge base is then applied to classify the polarity of drug reviews.

3. Motivation

Our studies on drug reviews from www.druglib.com and www.askapatient.com show that approximately 50 percent of sentences are polar facts that express patients' experiences about positive or negative effects of drugs. This means that traditional approaches to polarity classification, which rely on subjective statements, only consider portions of the available data and ignore a considerable amount of valuable information. Table 1 illustrates examples of polar facts about drugs. These examples are selected form www.druglib.com.

Some polar facts contain sentiment words, and hence can be classified using traditional approaches. For instance, in Table 1 examples (1–2), the words 'successfully' and 'cure' imply positive sentiments. A dictionary-based system would be able to correctly classify such sentences as positive since the sentiment words 'successfully' and 'cure' have positive polarity in sentiment lexicons such as SentiWordNet.

However, existing approaches to polarity classification are generally not sufficient for predicting the polarity of patients' experiences for three reasons. Firstly, most medical terms such as 'pain' and 'depression' are considered negative in current sentiment resources such as SentiWordNet and SenticNet, but they occur

Table 1
Examples of polar facts about drugs.

No.	Example	Polarity
1	Taken daily, Prevacid successfully reduced my acid stomach symptoms to the point that they did not occur	Positive
2	Accutane cured the acne	Positive
3	This drug reduced my pain significantly	Positive
4	Accutane completely eliminated my acne	Positive
5	This drug decreased my vision	Negative
6	This birth control method can cause blood clots	Negative
7	The antibiotic significantly diminishes the immune system	Negative
8	After using this drug, I experienced hot flashes	Negative

frequently in positive sentences. In example (3), although the negative word 'pain' is present, the sentence should be classified as positive. In fact, verbs play an important role in analyzing the facts' polarity. For example, the expression 'reduced my pain' in example (3) is positive although the term 'pain' is negative. Likewise, example (4) is positive although the word 'acne' is negative in SenticNet. Secondly, some patients' experiences do not contain sentiment words. Example (5) has no explicit sentiment word, but clearly implies a negative sentiment about the drug, since this experience is an undesirable fact. Thirdly, the medical domain is highly technical, while current resources for polarity classification often do not contain technical and domain terms, meaning that the performance of existing approaches in this domain is low. Thus, we need a different analysis technique for polarity classification of polar facts that not only spots domain terms but also considers context words such as verbs. To address the mentioned considerations, this paper explores the task of polarity classification of patients' experiences from unlabeled textual user reviews.

4. The proposed approach

The proposed approach for polarity classification of patients' experiences consists of two steps: building FactNet, the knowledge base of polar facts, and exploiting it for the task of polarity classification. An overview of our approach is depicted in Fig. 1. In the following subsections, we describe each of these steps in detail.

4.1. Building the FactNet knowledge base

For building FactNet, we extract two categories of polar facts. The first category contains the facts that express polar relationships between entities (e.g., 'Piroxicam is used to reduce the pain'). The second category contains polar concepts. Polar concepts are words and phrases which have polarity (e.g., 'severe abdominal pain', 'dry lips' and 'anemia').

In this section, we first introduce Linked Data sources that are used for extracting polar facts, and then we present an approach for building FactNet. Our proposed method consists of two stages. In the first stage, we extract polar facts from Linked Data sources. In the second stage, a generalization method is used to extract generalized semantic patterns of polar facts.

4.1.1. Linked Data sources

In recent years, vast amount of information about drugs have been published and integrated into LOD. In the Linked Open Drug Data (LODD) project [36], several data sources about drugs, clinical trials, diseases, and pharmaceuticals were added to the LOD, which provides novel data of interest to the pharmaceutical industry and patients. Using LOD in the pharmaceutical industry is ongoing but is currently at an early stage.

In our research, we aim to exploit the knowledge in three LOD datasets to build FactNet: (1) Drugbank[3] which is a repository of FDA-approved drugs that contains detail information about a drug's chemical and pharmacological properties as well as sequence, structure, and pathway information on each of the drug's known biochemical targets; (2) DailyMed which publishes detailed information about marketed drugs including general background on the chemical structure of the compound and its therapeutic purpose, details on the compound's clinical pharmacology, indications and usage, contraindications, warnings, precautions, adverse reactions, overdose, and patient counseling. Since the Linked Data version of DailyMed does not contain some facts about drugs, we also use the DailyMed site,[4] which provides complementary information about drugs; and (3) SIDER, which contains information on marketed drugs and their adverse effects.

4.1.2. Polar fact extraction

There are many positive and negative facts about drugs in Linked Data sources. For example, in SIDER, all drug side effects are negative concepts. In Drugbank, for each drug there is an 'indication' field that presents positive facts and there is a 'toxicity' field that expresses negative facts (see Table 2). Likewise, in DailyMed dataset, 'indication' and 'warning' fields contain positive and negative facts, respectively. On the DailyMed site, the 'indications and usage' field for each drug contains positive facts. Unfortunately, these facts are represented in free text, which makes them difficult to be processed by machine. In other words, these datasets contain implicit knowledge in their text fields. As it can be seen in Table 2, the 'indication' field links the drug's URI to a literal that is an unstructured textual description. Therefore, we use data mining and natural language processing techniques to transform unstructured texts into the structured form of RDF triplets, i.e., <subject, predicate, object>. RDF is a family of World Wide Web Consortium (W3C) specifications for modeling semantic metadata to describe Web resources on the Semantic Web [37]. An RDF triplet <subject, predicate, object> shows a semantic relation represented by the predicate between the subject and object. To extract RDF triplets, we present a two-step procedure. First, we define some lexico-syntactic patterns, and then we adopt a rule-based approach for relation extraction to the task of automatically extracting RDF triplets from LOD.

4.1.2.1. Lexico-syntactic patterns. We observe that facts about drugs in LOD are usually expressed through regular patterns. This motivates us to obtain a set of those patterns from the textual content of interested fields (i.e., 'indication' and 'toxicity' fields in Drugbank, and 'indication', 'indications and usage', 'warning' and 'precaution' fields in DailyMed). In this subsection, we introduce the procedure for polar fact extraction using lexico-syntactic patterns (Fig. 2).

As shown in Fig. 2, the first step is pre-processing of input data. For each textual content, we first use the Stanford CoreNLP[5] to detect sentences. Subsequently, we perform tokenization, lemmatization, POS tagging, dependency parsing, and shallow parsing for each sentence. Exploiting the Standford coreference resolution, we replace each resolved pronoun with the origin term that it refers to. Each sentence is then cleaned by removing special characters. Finally, we convert all characters into lower case.

The second step is mining frequent n-grams from the processed texts. We employ sequential pattern mining techniques in text databases to identify frequent word patterns from LOD. Sequential pattern mining is one of the key data mining techniques that has been intensively used for knowledge discovery in a variety of domains such as healthcare, education and telecommunications [38]. Given a database of sequences, where each sequence is a list of transactions and each transaction is a set of items, sequential pattern mining aims at finding all sequential patterns with a user specified minimum-support, where the support of a pattern is the number of sequences that contain the pattern.

We mine patterns for each field of interest separately. Tokens and sentences become 'items' and 'transactions', respectively, in a sequential pattern mining framework. We use a modified implementation of an Apriori-like method, called Generalized Sequential Patterns (GSP) [39], for sequential pattern mining that does not

[3] http://www4.wiwiss.fuberlin.de provides access to Drugbank, DailyMed and SIDER.

[4] http://dailymed.nlm.nih.gov.

[5] http://nlp.stanford.edu/software/corenlp.shtml.

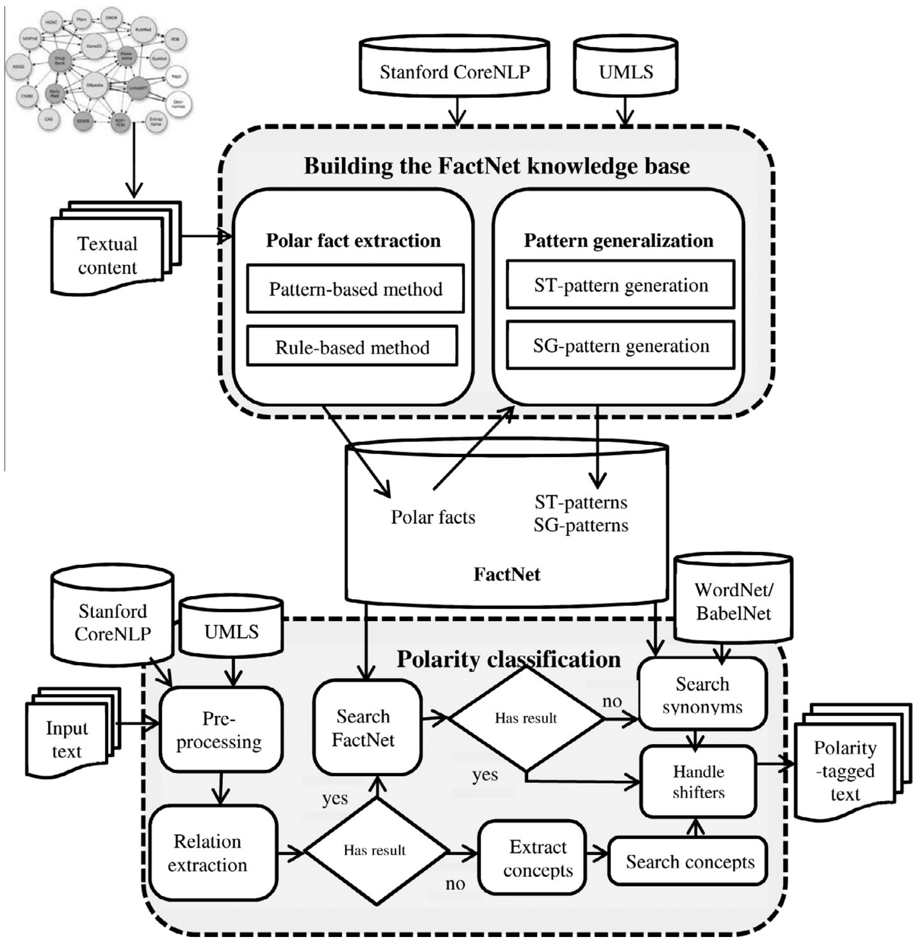

Fig. 1. An overview of the proposed approach.

allow gaps between tokens. In other words, we mine frequent n-grams with variable lengths.

GSP is a multiple-pass approach. In the first scan, it finds all of the single item frequent sequences (1-sequences). Each subsequent scan uses a seed set of frequent sequential patterns which is found in the previous scan. This seed set is used to generate candidate sequences. The candidate sequence of size $k + 1$ is generated by joining two frequent k-sequences when the prefix of one sequence is equal to the suffix of another one. Then, non-frequent candidates are removed. This process is repeated until no more frequent sequences are found.

We limit the length of n-grams to seven. This value is based on our observation that there is no meaningful pattern longer than 7. We also add another constraint that frequent tokens extracted in the first scan of the GSP algorithm should not be stop words. In this way, some lexical patterns are extracted. However, due to the automatic nature of the proposed method, some of the extracted patterns are meaningless. Thus, in the third step, to identify useful and descriptive patterns, we manually remove meaningless ones. In this step, about 63 percent of 358 extracted lexical patterns were removed.

In the fourth step, we define some lexico-syntactic patterns. Lexical patterns rely only on the surface form of text. Therefore, we manually transform them into lexico-syntactic patterns that are more expressive. To do this, we replace some words of a lexical pattern with their corresponding POS tags. We also specify the syntactic type of the phrase(s) which occur(s) after the lexical pattern.

Table 3 depicts some examples of the extracted patterns. The first column of Table 3a shows a lexical pattern. The second column is the Linked Data field that the lexical pattern is extracted from it. The third column presents the corresponding lexico-syntactic pattern. The fourth column shows the template of the relation which is extracted from the lexico-syntactic pattern. Finally, the fifth column presents the polarity label of the extracted relation. Table 3b illustrates some example sentences for each lexico-syntactic pattern.

In the fifth step, we use lexico-syntactic patterns to automatically extract a set of structured facts from LOD. To do this, for each field of interest in LOD datasets, the processed sentences of the textual content are used as the input for the pattern-matching step (Fig. 2). The pattern-matching step finds pattern(s) of the corresponding field that match the sentence. A sentence matches a pattern if there is an order preserving bijection from sequences of tokens in the sentence to items of patterns, so that each sequence of tokens satisfies corresponding items in the pattern.

For each sentence which is matched a pattern, we extract the relation among the entities using the corresponding relation template of the pattern. Then we assign a polarity label to each extracted relation. By default, relations that are extracted from 'indication' or 'indications and usage' fields are assigned positive labels and relations that are extracted from 'toxicity', 'precaution' or 'warning' fields are assigned negative labels. Finally, we add the extracted relation to FactNet.

Sometimes, authors of the review sites do not write a complete sentence. Instead, they describe their experiences with phrases, which are usually separated by comma. Consider the following examples:

'Very very dry skin'
'back pain'

Table 2
Examples of positive and negative facts about drugs in Linked Data sources.

Dataset	Field	Examples
Drugbank	Indication	http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00354 For prevention and treatment of nausea, vomiting, and dizziness associated with motion sickness and vertigo...
		http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00144 Phosphatidylserine has demonstrated some usefulness in treating cognitive impairment, including Alzheimer's disease, age-associated memory impairment and some non-Alzheimer's dementias...
	Toxicity	http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00121 Prolonged skin contact may cause irritation
		http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00034 Interferon alfa-2 may cause serious adverse effects such as anemia; autoimmune diseases, including vasculitis, arthritis, hemolytic anemia, and erythematosus syndrome; cardiotoxicity; hepatotoxicity; hyperthyroidism or hypothyroidism; transient ischemic attacks; leukopenia; neurotoxicity; peripheral neuropathy; and thrombocytopenia...
DailyMed	Indication	http://www4.wiwiss.fu-berlin.de/dailymed/resource/drugs/4293 RYZOLT is indicated for the management of moderate to moderately severe chronic pain in adults who require around-the-clock treatment of their pain for an extended period of time
		http://www4.wiwiss.fu-berlin.de/dailymed/resource/drugs/7 Fluconazole is indicated for the treatment of: Prophylaxis. Fluconazole is also indicated to decrease the incidence of candidiasis in patients undergoing bone marrow transplantation who receive cytotoxic chemotherapy and/or radiation therapy...
	Warning	http://www4.wiwiss.fu-berlin.de/dailymed/resource/drugs/90 Increased Mortality in Elderly Patients with Dementia-Related Psychosis: Elderly patients with dementia-related psychosis treated with atypical antipsychotic drugs are at an increased risk of death compared to placebo...
		http://www4.wiwiss.fu-berlin.de/dailymed/resource/drugs/114 Contains sodium metabisulfite, a sulfite that may cause allergic-type reactions including anaphylactic symptoms and life-threatening or less severe asthmatic episodes in certain susceptible people...

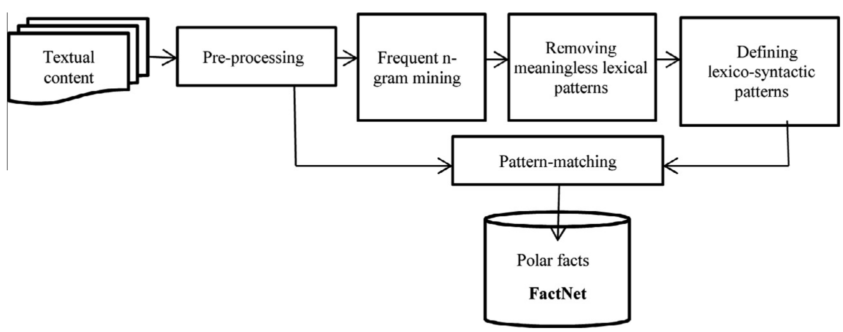

Fig. 2. Polar fact extraction using lexico-syntactic patterns.

In order to classify such examples we should know that 'dry skin' and 'back pain' are negative concepts. For this reason, we collect a set of polar concepts. All of side effects in the 'sideEffectName' field of the SIDER are negative concepts. In addition, we define some lexico-syntactic patterns to extract more negative concepts from Linked Data sources. Some of these patterns are shown in Table 4. We also collect phrases of the 'purpose' filed in DailyMed site as positive concepts.

4.1.2.2. A rule-based method for relation extraction. Since patterns match only a limited context, we use relation extraction techniques for extracting more facts from LOD. Rule-based methods for open information extraction often achieve reasonable precision. Therefore, we use a rule-based method for extract polar facts from relation-bearing sentences. We observe that most of relation-bearing sentences are in the template <subject, verb, object>. Thus, we use this template with some constraints on nouns and verbs to identify the relation between two entities. To this end, we perform the following steps:

- We only select the first sentence of textual contents of each interested field since generally the first sentence of each paragraph is the most important one and summarizes what will be developed in the paragraph.
- We ignore sentences matched with a pattern in the pattern-based method, and sentences with negations, since they are usually noise. For example, in the 'toxicity' field of Drugbank, sentences such as 'There is no experience of acute overdose.' occur frequently, but they are not actually negative facts.
- We identify the main verb(s) of the sentence using the approach proposed in [40]. We define a constraint on the main verb that it should not be a copula ('to be') verb.
- We apply some predefined rules to determine the object and subject of the sentence using the Stanford Parser. We also add a constraint on the subject and object of the sentence that they should be one of the biomedical or health-related concepts in the Unified Medical Language System (UMLS)[6] Metathesaurus. UMLS is a large vocabulary in the biomedical sciences, which contains biomedical and health-related concepts, common terms, and the relations among them. To map terms in the sentence to UMLS concepts, we use MetaMap[7] [41]. MetaMap often returns multiple mappings for a given phrase. Named entity disambiguation is the task of linking entity mentions in a text to a knowledge base that has attracted increasing attraction in recent years [42]. To disambiguate named entities and find the most appropriate concept for the given phrase, MetaMap uses some weighted heuristics. In this way, a given phrase may be mapped to several concepts with different mapping scores. Some of previous works selected the concept (sematic type) with the highest mapping

6 www.nlm.nih.gov/research/umls/.
7 http://metamap.nlm.nih.gov/.

Table 3
Examples of the extracted patterns.

Lexical pattern	Field	Lexico-syntactic pattern	Relation template	Polarity
(a) Examples of lexical patterns, their corresponding lexico-syntactic patterns and their relation template with their polarity labels				
Is indicated for the treatment of Is indicated in the management of Is indicated for the control of	Indication	Is indicated in/for [the][a] [ADV] NN of NP	(subject[b], verb form of NN[c] (if available), NP)	Positive
Helps	All fields	Help[s] VP NP	(subject, verb stem of VP, NP)	Positive
Cause	Toxicity, warning and precaution	Cause[s] NP	(subject, cause, NP)	Negative
Result in	Toxicity, warning and precaution	Result[s] in NP	(subject, result in, NP)	Negative
Increase in the risk of	All fields	Increase [in] the risk of NP	(subject, increase, risk of NP)	Negative
		Increase [in] the risk of VP NP	(subject, verb stem of VP, NP)	Negative

Lexico-syntactic pattern	Example sentence	Extracted relation
(b) Example sentences for each lexico-syntactic pattern		
Is indicated in/for [the] [ADV] NN of NP	Transdermal nitroglycerin **is indicated for the prevention of** angina pectoris due to coronary artery disease	(transdermal nitroglycerin, prevent, angina pectoris)
	DUTOPROL **is indicated for the management of** hypertension	(dutoprol, manage, hypertension)
	COSOPT **is indicated for the reduction of** elevated intraocular pressure in patients with open-angle glaucoma	(cosopt, reduce, elevated intraocular pressure)
Help[s] VP NP	This **helps** eliminate blood clots	(retavase, eliminate, blood clots)
Cause[s] NP	It may **cause** an allergic reaction	(bleomycin, cause, an allergic reaction)
Result[s] in	Excessive administration of dextrose injections may **result in** significant hypokalemia	(excessive administration of dextrose injections, result in, significant hypokalemia)
Increase [in] the risk of NP	Venlafaxine may significantly **increase the risk** of suicide	(venlafaxine, increase, the risk of suicide)
Increase [in] the risk of VP NP	Numerous drug classes **increase the risk of** developing serious cardiac arrhythmias	(numerous drug classes, develop, serious cardiac arrhythmias)

[a] We use [.] for an arbitrary argument.
[b] In order to determine the subject of a sentence, we use the sentence dependency tree. For sentences which do not have a subject, by default, we consider the drug name as the subject. Each textual content is linked to a URI. We use SPARQL query on this URI to find the drug name.
[c] Some patterns require the verb form of a noun. To this end, we use the 'getDerivationallyRelatedForms' function in the Java API for WordNet Searching (JAWS). JAWS is an API that provides Java applications with the ability to retrieve data from the WordNet database. http://lyle.smu.edu/~tspell/jaws/ provides access to JAWS.

Table 4
Examples of lexico-syntactic patterns for extracting negative concepts.

Lexico-syntactic pattern	Example	Negative concepts
Symptoms such as NP, NP, ... or/and NP	Symptoms such as severe abdominal pain or cramping	Severe abdominal pain Cramping
Symptoms include NP, NP, ... and NP	Symptoms include facial flushing, nausea, vomiting and hypotension	Facial flushing Nausea Vomiting Hypotension
The signs and symptoms of NP, NP, ... and NP	The signs and symptoms of ulcerations and bleeding	Ulcerations Bleeding

score as the appropriate concept for the ambiguous phrase. In this paper, we use the method proposed by Yang et al. [34] for named entity disambiguation that, according to their experiments, outperforms the mapping function of MetaMap. This method uses MetaMap to create a list of semantic type candidates for a give phrase. Each candidate is assigned a mapping score using the mapping function of MetaMap. It then selects all candidates whose mapping score is in [sh-ϴ, sh], where sh is the highest mapping score of candidates, and ϴ is a constant parameter that is set to 50. Finally, this method selects the most frequently mentioned semantic type among selected candidates of the previous step to label the given phrase.
- The extracted relation from the input sentence is encoded using RDF template <subject, predicate, object>, in which 'predicate' is the stem of the main verb of the sentence.
- We assign positive label to the triplet extracted from 'indication' or 'indications and usage' fields, and assign negative label to the triplet extracted from 'toxicity', 'precaution' or 'warning' fields.
- The extracted triplet is added to FactNet.

4.1.3. Mining generalized patterns of polar facts

Having a set of polar facts, we apply a generalization method to extract generalized patterns of polar facts. Generalization is intended to overcome the issues raised by lack of knowledge in FactNet and improve recall of the polarity classification algorithm. Generalized patterns have higher coverage than specific ones, and so have a greater chance of matching a context.

In this section, we introduce a two-step approach relying on UMLS taxonomy for generalizing polar facts and organizing them into a hierarchy. The UMLS Metathesaurus includes 1.7 million concepts, grouped into more than 130 semantic types. Each semantic type belongs to one of the 15 semantic groups [43]. For example, the concept 'hypertension' belongs to the semantic type 'Disease or Syndrome (dsyn)' which in turn belongs to the semantic group 'Disorders (DISO)'.

In the first step, for each triplet in FactNet, we replace the subject and object with their corresponding semantic types. In this way, we have a database of semantic triples in which the subjects and objects are semantic types. We then find frequent triplets, i.e., triplets that occur more than a predefined threshold called the minimum support. We call these triplets 'ST-patterns'.To filter out less reliable patterns, we calculate a confidence score for each ST-pattern. The confidence score is defined as follows:

$$confidence\ (p) = \frac{|PMT - NMT|}{PMT + NMT} \quad (1)$$

where p denotes a pattern, and PMT and NMT are the number of positive and negative triplets matched with the pattern, respectively. When a pattern only matches with positive (or negative) instances, confidence would be 1, and when it matches with the equal number of positive and negative triplets, confidence would be 0. We filter out patterns whose confidence scores are below a predefined threshold called the minimum confidence.

Finally, we assign a polarity label to each ST-pattern. The ST-pattern is positive if the number of positive triplets matching the pattern is more than that of the negative ones, and vice versa.

In the next step, for each triplet in the database of semantic triples, we replace each semantic type by its associated semantic group and find frequent triplets called 'SG-patterns'. SG-patterns are generalizations of ST-patterns. Fig. 3 illustrates a simple example of pattern generalization. In a similar way, we remove less reliable SG-patterns using a confidence score, and assign a polarity label to them.

Finally, we add generalized patterns to FactNet. In fact, FactNet is a knowledge base of polar fact instances and their generalized patterns. Some examples of the extracted triplets are illustrated in Table 5.

4.2. Exploiting FactNet

In this section, we introduce an algorithm for exploiting the discovered knowledge to the task of polarity classification, which takes a sentence or a phrase as input and outputs its polarity (Fig. 1).

In the pre-processing step, we employ natural language processing techniques, including tokenization, lemmatization, named entity recognition, POS tagging, dependency parsing and shallow parsing to process text.

In the second step, a set of predefined rules adopted from [44] is used for relation extraction from the processed text. If the relation extraction module extracts a relation from the input text, we perform the following steps:

- Each extracted relation is encoded in the RDF template, and the resulting triplet is called RDF-triplet.
- For the RDF-triplet, we produce two other triplets: the ST-triplet, which contains semantic types of entities, and the SG-triplet, which includes semantic groups of entities.
- Extracted triplets are searched in FactNet. A search in FactNet can be performed using two strategies: top-down or bottom-up. The top-down method evaluates general patterns first and then moves down to specific patterns and instances. The bottom-up method moves up from instances to specific patterns and then to general patterns. In Section 5.3, we evaluate the performance of the proposed approach using each of these search strategies.
- If one of the RDF-triplet, ST-triplet or SG-triplet exists in FactNet, the corresponding polarity label is returned.
- Otherwise, in order to extend the coverage of the proposed approach, we exploit two external sources of knowledge: WordNet [45] and BabelNet [46]. WordNet is a lexical database for English language which groups words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members. BabelNet is a multilingual semantic network obtained from the automatic integration of WordNet and Wikipedia.[8] BabelNet encodes knowledge as a graph, in which nodes are babel synsets, and edges are semantic relations between them. Each babel synset represents a given meaning and contains a group of synonym terms. When the RDF-triplet <subject, verb, object>, and corresponding ST- and SG-triplets do not appear in FactNet, we search FactNet for triplets in which the verb is replaced with one of its synonyms obtained from the external source of knowledge, i.e., WordNet or BabelNet. If one of these triplets exists in FactNet, the corresponding polarity label is returned. In order to obtain the synonyms of a word from BabelNet, we first perform word sense disambiguation using Babelfy [42], and then, we extract the synonyms for the correct sense of the word. Babelfy is a multilingual state-of-the-art approach to word sense disambiguation and entity linking based on BabelNet.

  We also employ BabelNet to determine the semantic group of concepts (i.e., subject/object of a RDF-triplet) which are not tagged by MetaMap (e.g., 'logy' and 'skinny'). To this end, we first find the synonyms of the untagged concept using BabelNet. Then we find the semantic groups of these synonyms using MetaMap. Finally, we select the most frequently mentioned semantic group among synonyms to label the untagged concept.

Sometimes, the relation extraction module does not extract any relation from the input text. In these cases, we use polar concepts of FactNet. As we mentioned earlier, FactNet contains some positive and negative concepts which can be used to the polarity classification task. If the input text is not a relation-bearing sentence, we perform the following steps:

- Concepts of the input text are extracted using some linguistic patterns such as ADJ + NOUN and NOUN + NOUN.
- FactNet is searched to find these concepts.
- If this search leads to a result the corresponding polarity label will be returned.

The polarity of a triplet or a concept can be affected by a set of valence shifters. In this research, we consider two types of valence shifters which occur frequently in drug reviews: negators and quantifiers. Negation words such as 'not' and 'no' can change the polarity of a text. For example, although 'dry lips' is a known side effect of 'Accutane', but the negation word 'not' implies that the sentence 'Accutane did not dry my lips.', has positive polarity. Likewise, quantifiers which express a decreased/increased value of quantity can also change the polarity. For example, the phrase 'less acne' has positive polarity, although the word 'acne' is negative.

Negation is difficult to detect in text, especially when it is expressed implicitly (i.e., without use of an explicit negation marker) in an ironic expression [5]. To handle this issue, at the first stage, we use the negation detection tool of MetaMap. MetaMap includes an implementation of NegEx, a negation detection algorithm that is based on regular expressions and a dictionary of medical terms [47]. Although NegEx usually correctly detects negated terms, it is not able to detect other kinds of valence shifters such as quantifiers. Thus, we combine NegEx with a dependency-based approach for valence shifter detection. In this approach, we first detect negations and quantifiers of the text using a list of valence shifters. However, since not all appearances of valence shifters reverse the polarity of the enclosing sentence, we determine the scope of valence shifters with the help of chunk dependency tree. To do this, we first obtain the dependency tree of

[8] http://www.wikipedia.org.

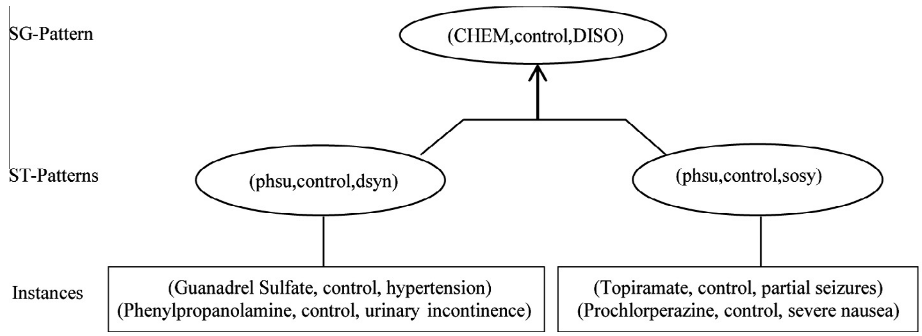

Fig. 3. A simple example of pattern generalization.

Table 5
Examples of the extracted triplets.

Polarity	Polar facts	ST-patterns	SG-patterns
Positive	(domperidone, manage, dyspepsia)	(phsu, eliminate, patf)	(CHEM, help, DISO)
	(pantoprazole, treat, erosive esophagitis)	(phsu, lessen, ftcn)	(CHEM, clean, ANAT)
	(syntocinon, improve, uterine contractions)	(phsu, treat, sosy)	(CHEM, manage, CONC)
	(esmolol, control, ventricular rate)	(phsu, improve, ftcn)	(CHEM, prevent, DISO)
Negative	(amoxicillin, increase, morbidity)	(phsu, develop, dsyn)	(CHEM, cause, DISO)
	(carisoprodol, cause, dizziness)	(phsu, damage, tisu)	(CHEM, damage, ANAT)
	(cognex, decline, cognitive function)	(phsu, worsen, dsyn)	(CHEM, decrease, PROC)

the sentence using Stanford parser. Then we construct a chunk dependency tree, in which a node represents a lexical chunk in the input sentence, and a relation denotes a dependency relation between words in the dependency tree. There are two kinds of relations in a chunk dependency tree. When two words are in the same chunk, their dependency relation is added intra the corresponding chunk node, and when two words belong to different chunks, their dependency relation is added inter the corresponding chunk nodes [48]. We assume that a sentence has a polarity shifter if there is an inter- or intra-relation between the interested entities (i.e., subject/object of the extracted RDF triplet for relation-bearing sentences, and polar concepts of phrases) or main verb of the sentence and a valence shifter word. In Section 5.3, we show that this approach outperforms the baseline method, where each appearance of valence shifters inverts the polarity of text. However, more effective valence shifter detection needs deep analysis and can be considered as future work.

In summary, a sentence has a polarity shifter if it is recognized as a negated sentence by NegEx algorithm or by the proposed dependency-based approach. After obtaining the polarity label of a relation or a concept, if the input text contains a polarity shifter, the polarity will be reversed.

Finally, if there is more than one concept or relation in the input text, we use the majority voting mechanism to determine the overall polarity of the text.

5. Experiments

In this section, we first describe briefly the dataset used in our experiments and then present and discuss the experiments that we conducted to evaluate our approach. To evaluate the performance of the proposed approach, we use three measures: precision, recall and *F*-measure. Precision is the percentage of classified instances that are correct, recall is the percentage of instances that are correctly classified, and *F*-measure is the harmonic mean of precision and recall. The precision, recall and *F*-measure are defined as follows:

$$\begin{aligned} \text{precision} &= \frac{\textit{No. of correctly classified instances}}{\text{Total No. of classified instances}} \\ \textit{recall} &= \frac{\textit{No. of correctly classified instances}}{\text{Total No. of instances}} \\ F\text{-measure} &= \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \end{aligned} \tag{2}$$

5.1. Construction of the evaluation dataset

Since there is no public dataset specifically designed for polarity analysis of patients' experiences of drugs, we first created a dataset of drug reviews. This dataset was collected from www.askapatient.com and www.druglib.com, two popular websites for reviewing drugs. This dataset contains 850 reviews for 75 drugs which were chosen randomly from the list of the most frequently rated drugs at the first page of the *druglib.com* website. 700 of these reviews were used for building a dataset of relation-bearing sentences. Three annotators were asked to identify polar sentences that express a relation between a drug and another entity and annotate them with polarity tags (positive/negative). We measured the inter-annotator agreement using Fleiss' kappa [49]. Fleiss' kappa is expressed as a number between 0 and 1, where 1 indicates a perfect agreement. In this stage, the substantial agreement was observed (kappa = 0.77).

At first, we only chose those sentences that were tagged by all annotators. Then, to extend our dataset, we asked two other annotators to tag any instances where disagreement occurred. Finally, we chose those instances that were tagged by the majority of the annotators.

We discarded comparative sentences such as 'Yasmin is better than Desogestrel in terms of weight control' since they require different analysis techniques.

In this way, we created an evaluation dataset comprising 1401 polarity tagged relation-bearing sentences. We also collected a set of 1006 polar phrases about drugs. In fact, a collection of 2407 polar facts was created, of which 49.11 percent were labeled as positive and the rest (50.89 percent) were negative.

5.2. FactNet evaluation

All polar fact instances and respective generalized patterns are stored in FactNet. The statistics of FactNet are illustrated in Table 6. To extract ST- and SG-patterns we set the minimum support and the minimum confidence to 2 and 0.6, respectively, by using a trial-and-error approach. We evaluate FactNet from two aspects: quality of polar facts, and quality of patterns.

5.2.1. Evaluation of the proposed method for polar fact extraction

Applying the proposed approach for polar fact extraction, we constructed a knowledge base of 9703 triplets from LOD (redundant triplets were discarded). The lexico-syntactic patterns found 5989 triplets, while the rule-based method extracted 3714 ones.

Semantic triplets are core component of FactNet. Thus, success of the polarity classification algorithm depends on the correctness of these triplets. In the first experiment, we evaluated the correctness of the extracted triplets. We picked a random sample of the extracted polar facts including 310 triplets, in which 70 tokens were extracted by the rule-based model and 240 ones extracted by the lexico-syntactic patterns. We manually checked these facts, and reported the precision as shown in Table 7. As can be seen, the results obtained using the proposed approach are promising. Table 7 also compares the performance of the lexico-syntactic patterns and the rule-based method for polar fact extraction. As can be seen in Table 7, polar facts are extractable with good precision using lexico-syntactic patterns whereas our analysis on the extracted facts show that the rule-based method contributes greatly to the coverage of relation extraction and extracts a large variety of relations.

5.2.2. Analysis of patterns

In order to show the effectiveness of the extracted patterns, we assessed three possible configurations to detect the polarity of sentences in the evaluation dataset described in Section 5.1; using polar fact instances in isolation, using ST-patterns in isolation, and using SG-patterns in isolation. The results of the evaluation are presented in Table 8. As can be seen, polar fact instances can classify experiences with higher precision than ST- and SG-patterns, but that precision is achieved at a lower level of recall. ST- and SG-patterns have significantly better recall rate (42.5% and 58.77% against 13.01%). As we expected, generalized patterns have higher coverage than specific ones, and so have a greater chance of matching a context. In fact, generalized patterns are able to classify unseen polar facts. Table 8 also presents that the combination of polar fact instances, ST- and SG-patterns using bottom-up strategy (see Section 4.2) outperforms each of them in isolation in terms of both recall and *F*-measure. In particular, the recall value (66.67%) is significantly high compared to the corresponding recall rates obtained by using polar fact instances, ST- and SG-patterns in isolation.

To evaluate the quality of the extracted patterns, we also analyzed them to verify if they are reasonable patterns or not. We observed that 70 percent of SG-patterns are common sense patterns, i.e., people already know them. For example, the SG-pattern <CHEM, control, DISO>, in which DISO means a disorder, describes a common sense fact that the control of a disorder by a drug is positive. These patterns, in fact, reflect the underlying semantic structures of the drug domain.

Table 6
Statistics of FactNet.

	No. of polar facts		No. of ST-patterns	No. of SG-patterns
	Triplets	Concepts		
Positive	8275	394	853	172
Negative	1428	4169	99	47
Total	9703	4563	1436	224

Table 7
Precision of the patterns-based and rule-based methods for polar fact extraction.

	Pattern-based method	Rule-based method	Total
No. of extracted facts	5989	3714	9703
Precision (%)	94.17	85.71	92.26

Table 8
Performance of the ST- and SG-patterns for polarity classification.

	Recall (%)	Precision (%)	*F*-measure (%)
Instances	13.01	89.66	22.72
ST-patterns	42.5	83.45	56.32
SG-patterns	58.77	76.01	66.29
The proposed method using bottom-up strategy	66.67	84.52	74.54

5.3. The proposed method for polarity classification

We evaluated the proposed method for polarity classification by employing it to detect the polarity of examples in the evaluation dataset described in Section 5.1. As mentioned earlier, there are two kinds of examples in the evaluation dataset: relation-bearing sentences and simple phrases that the relation extraction algorithm cannot extract any relation from them.

At first, the performance of the proposed method for polarity classification of relation-bearing sentences was evaluated. For these experiments, we manually extracted RDF-triplets from test examples to omit the effect of relation extraction module on the performance of the proposed approach.

In the first experiment, we evaluated the performance of the proposed approach using two search strategies: top-down and bottom-up. In this experiment, we used WordNet without word sense disambiguation to find synonyms. The bottom-up strategy obtained the *F*-measure of 74.54 percent, which was about 7 percentage points better than the top-down strategy (Fig. 4). The main reason is that polar fact instances are more accurate than ST- and SG-patterns (see Table 8). When we used the top-down strategy, about 80% of test examples were classified by SG-patterns. Since SG-patterns have lower precision than ST-patterns and polar fact instances, the overall performance of the top-down strategy is lower than that of the bottom-up strategy. We also conducted this experiment using BabelNet with word sense disambiguation to find synonyms. In this experiment, the top-down and bottom-up strategies achieved the *F*-measure of 61.08 and 68.58 percent, respectively. In the next experiments, we used the bottom-up strategy.

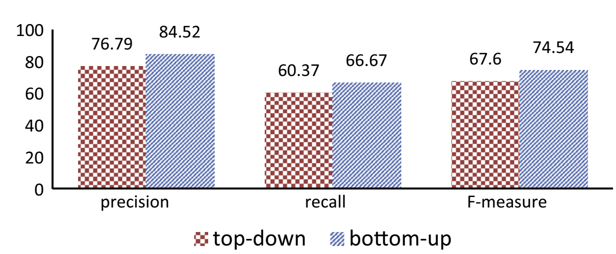

Fig. 4. Comparison of the top-down and bottom-up strategies for polarity classification.

In order to assess the effectiveness of using an external source of knowledge such as WordNet or BabelNet, we ran the proposed approach for polarity classification with and without using these resources. Fig. 5 illustrates that using WordNet without word sense disambiguation (WB) and BabelNet with word sense disambiguation (BB) can lead to an approximately 10 and 4 percent of improvement in terms of *F*-measure, respectively. As can be seen from Fig. 5, using BabelNet achieves better precision in comparison with using WordNet. The reason is that BabelNet has a larger synonym inventory. Furthermore, word sense disambiguation improves the precision to some extent.

We also evaluated the performance of the proposed method for valence shifter detection. Fig. 6 illustrates the performance of the proposed approach for polarity classification with four methods of valence shifter detection: baseline, where each appearance of valence shifters inverts the polarity of text, NegEx algorithm, dependency-based approach, and the combination of NegEx and dependency-based approach. The NegEx algorithm is not able to detect quantifiers. However, in some cases, it detects longer distance dependency between the negation trigger and the polar expression that cannot be recognized by the dependency-based approach. Thus, as can be seen from Fig. 6, the combined approach outperforms other ones.

In the next experiment, we evaluated the effectiveness of using BabelNet for determining sematic types of concepts which are not tagged by MetaMap in two possible configurations. First configuration (WB) uses WordNet without word sense disambiguation to find synonyms of verbs, and the second one (BB) exploits BabelNet with word sense disambiguation for finding synonyms. As depicted in Table 9, using BabelNet for tagging untagged concepts slightly improves the recall of these configurations.

In the next experiment, we compare the participation of the polar fact instances, ST- and SG-patterns in polarity classification. Approximately 15 percent of examples were tagged by the polar fact instances. ST- and SG-patterns classified about 38 and 47 percent of examples, respectively.

Finally, to investigate the effectiveness of the proposed method for polarity classification, we compared it to some baseline methods and some state-of-the-art sentiment analysis and opinion mining approaches. In these experiments, we performed the proposed method for polarity classification on the whole evaluation dataset (i.e., relation-bearing sentences and phrases). We also automatically extracted RDF-triplets and polar concepts using the proposed relation extraction algorithm and linguistic patterns, respectively. Table 10 illustrates the performance of the proposed method for polarity classification of relation-bearing sentences and phrases. Comparison of Table 10 with Table 9 shows that using the proposed relation extraction algorithm caused 5.53 and 4.28 percentage points decline in *F*-measure of the WB and BB configurations for polarity classification of relation-bearing sentences, respectively.

We employed a dictionary-based method and a distant-supervision approach as baseline methods. For the dictionary-based approach, we chose SentiWordNet that is a well-known and popular resource for opinion mining. In order to calculate the overall sentiment of a text, we first extracted the polarity value of each word using SentiWordNet, and then, we aggregated these values using two methods: majority voting that counts the number of positive and negative words of the text and selects the majority number, and sum of predictions in which the text polarity value is computed as the sum of polarity values of its words.

Distant supervision aims at using a large set of weakly labeled data to train a supervised classifier [50]. Previous works in the field of opinion mining exploited various features such as emoticons, overall rating of the reviews and the review structure (i.e., pros and cons) for obtaining noisy training data [51,52]. We used the structure of review sites to obtain such data. Most of review sites ask the reviewer to describe pros and cons separately. We assumed that sentences in the 'pros' field of a review are positive, and sentences in the 'cons' field are negative. We used the obtained dataset and a set of features exploited by Pang et al. [20], i.e., unigrams

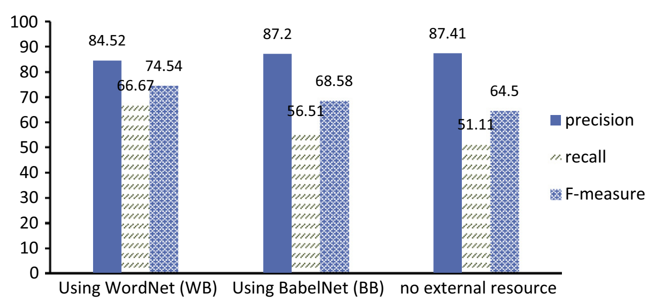

Fig. 5. Results of polarity classification with and without using an external source of knowledge.

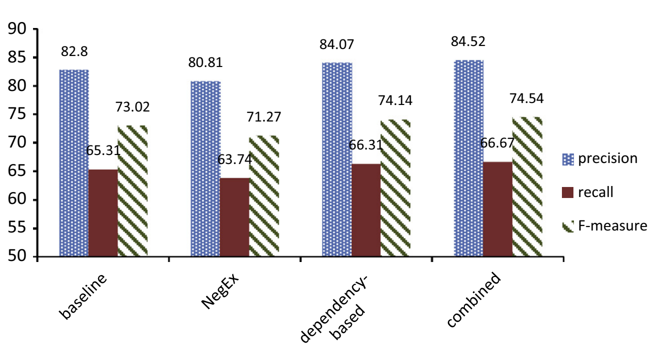

Fig. 6. Results of polarity classification with different methods of valence shifter detection.

Table 10
Performance of the proposed approach for polarity classification.

	Precision (%)		Recall (%)		*F*-measure (%)	
	WB	BB	WB	BB	WB	BB
Phrases	82.57		64.13		72.19	
Relation-bearing sentences	75.49	78.48	64.01	54.91	69.28	64.61
Total	77.87	79.78	64.60	59.33	70.62	68.05

Table 9
Effectiveness of using BabelNet for determining semantic types of untagged concepts.

	WB		BB	
	Only MetaMap	MetaMap and BabelNet	Only MetaMap	MetaMap and BabelNet
Precision	84.52	84.53	87.20	87.19
Recall	66.67	**67.09**	56.51	**56.94**
F-measure	74.54	74.81	68.58	68.89

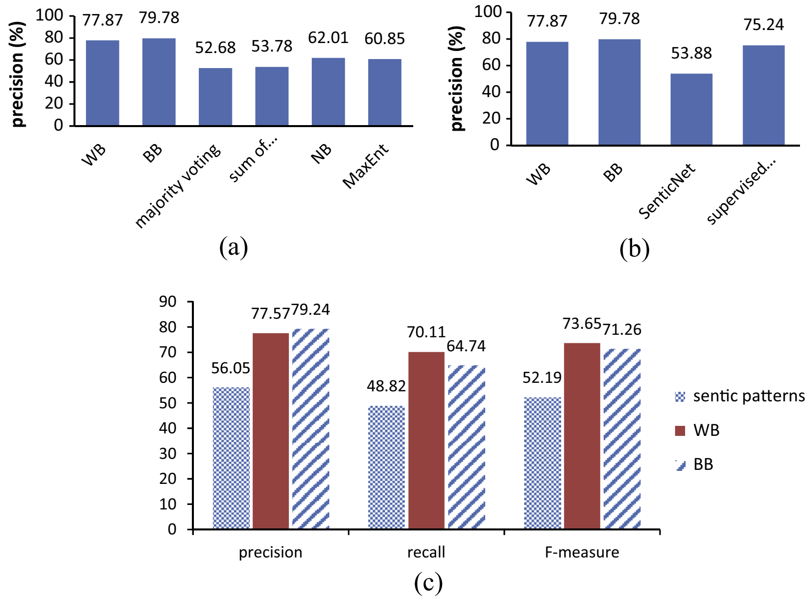

Fig. 7. Comparison of the proposed approach for polarity classification with (a) the baseline methods, (b) other state-of-the-art methods, and (c) sentic patterns.

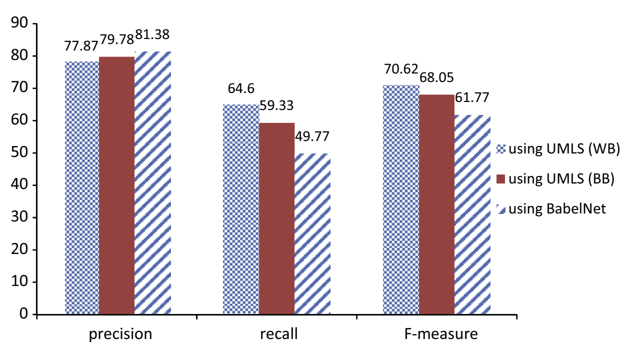

Fig. 8. Comparison of using UMLS vs. using BabelNet for polarity classification.

(presence of a certain word, frequencies of words), bigrams, POS tags, and adjectives for training the Naïve Bayes (NB) and Maximum Entropy (MaxEnt) classifiers which are widely used in opinion mining. In our experiments, using unigrams alone yielded the best result.

Fig. 7a compares the precision of the proposed approach for polarity classification with the baseline methods. As can be seen from Fig. 7a, both configurations of the proposed approach (i.e., WB and BB) significantly outperform the baseline methods.

Fig. 7b compares the proposed approach with two state-of-the-art methods. The first approach was proposed by Cambria and Hussain [53] and was exploited effectively for measuring health-care quality [54]. This approach first deconstructs the given text into concepts using a semantic parser, and then, uses SenticNet, a rich concept-level knowledge base for opinion mining, to associate polarity values to these concepts. Finally, it computes the overall polarity of the text by averaging such values. The second one is a supervised method proposed by Habernal et al. [21]. They evaluated different pre-processing techniques, various features and feature selection algorithms as well as different classifiers for supervised sentiment analysis. Employing this approach, the best performance was achieved by using the following setting. For the pre-processing, the pipeline of tokenization, stemming and lower-casing was employed. The feature set included unigrams, bigrams, and various POS features. We ignored the emoticons feature since they rarely occur in drug reviews. Among various feature selection methods, mutual information yielded the best result, and improved the overall precision slightly. The weighting scheme based on TF-IDF[9] did not improve the performance and was ignored. In addition, the MaxEnt classifier led to better result than that of the Naïve Bayes classifier. Since there was no publicly available dataset of patients' experiences to use as training set, we used 10-fold cross-validation on the evaluation dataset. As depicted in Fig. 7b, the proposed approach achieves better results in comparison with other methods.

We also compared the proposed approach with sentic patterns [55], a concept-level approach that merges linguistics, common sense computing, and machine learning for polarity detection. To produce results using sentic patterns, we used sentic demo[10]. Since sentic patterns is a sentence-level method and there are some phrases in our test set, it could not tag some instances of the test set. Thus, for a fair comparison, we computed the performance of the proposed approach on the portion of the test set that was tagged by sentic patterns. Fig. 7c illustrates the performance of the proposed method and compares it with the sentic patterns performance. As can be seen in Fig. 7c, the proposed method outperforms the sentic patterns method. The main drawback of sentic patterns is that it does not contain technical concepts which occur frequently in the drug domain.

In the previous experiments, we performed named entity recognition and semantic pattern extraction based on UMLS, a domain-specific Metathesaurus. In the last experiment, we employed BabelNet, a wide-coverage multilingual knowledge base of concepts and semantic relations. To this end, we used Babelfy to named entity disambiguation and WiBi taxonomy [56] to extract semantic patterns. WiBi is a bitaxonomy, i.e., a taxonomy of Wikipedia pages and categories that according to the conducted experiments in [56] outperforms available knowledge resources including MENTA, DBpedia, YAGO, WikiNet and WikiTaxonomy. WiBi is integrated into BabelNet 3.0. We assessed different strategies to extract semantic classes of the entities; using Wikipedia

[9] Term frequency-inverse document frequency.
[10] *Sentic.net/demo*, Access at January, 2015.

categories, using Wikipedia page taxonomy and using both of them. We also exploited direct hypernyms and two levels of hypernyms as semantic classes for each entity. According to our experiments, using two levels of Wikipedia pages and categories outperforms other strategies in terms of *F*-measure. This strategy achieves 81.38 percent precision and 49.77 percent recall (Fig. 8). As can be seen in Fig. 8, using BabelNet achieves lower recall in comparison with using UMLS. The main reason is that some concepts (especially technical concepts in medical domain) do not get semantic class using BabelNet. Another reason is that BabelNet categories are more specific than those of UMLS. For this reason, using BabelNet leads to the sparse but accurate set of patterns. However, this experiment indicates that BabelNet is an appropriate resource, especially for resource-lean domains.

5.4. Discussion

Experimental results (Fig. 7) shows that FactNet is an appropriate resource for polarity classification of patients' experiences. However, there is some room for improving the performance of polarity classification.

From the error analysis we performed, we have found errors were mainly caused by the following reasons: (1) misspelling words which caused some entities not to be recognized by MetaMap; (2) words and phrases which got wrong tags by MetaMap and caused errors in entity recognition; (3) errors in dependency and parse trees which led to errors in relation extraction between entities; (4) unusual abbreviations which caused errors in named entity recognition by MetaMap; (5) errors of the valence shifter detection method; (6) missing entities in the corpus; and (7) lack of synonyms or abuse of them, especially when we used WordNet without word sense disambiguation. These errors indicate how to improve the quality of our proposed approach in the future. To this end, we can use spell checking, abbreviation recognition, alternative tools for parsing, alternative methods for relation extraction, or alternative approaches for named entity recognition and categorization.

6. Conclusion

This paper proposed an automatic method for polarity classification of patients' experiences of drugs expressed in unlabeled textual user reviews. The main contributions of this paper are as follows. First, we proposed a combination of lexico-syntactic patterns and a rule-based method for relation extraction to extract polar facts from existing knowledge in LOD, in the structured form of RDF triplets. Different from existing approaches for polar fact classification, which mainly depend upon the availability of annotated corpora to train a classifier, our proposed approach exploits the existing knowledge in LOD. Second, a generalization method was presented to extract generalized patterns of polar facts and organize them into a hierarchy. Generalization aims to overcome the missing knowledge issue and improve the recall of the polarity classification method. Using the extracted knowledge, we built FactNet, a knowledge base of polar fact instances and their generalized patterns. Finally, we proposed a method for polarity classification of patients' experiences based on FactNet.

Some experiments were designed to evaluate the quality of FactNet. We also compared the performance of the proposed method for polarity classification on a dataset of drug reviews to that of some baseline methods and state-of-the-art approaches for sentiment analysis and opinion mining. The results indicate that the proposed method outperforms these approaches.

We concluded that our approach is appropriate for polarity classification of drug reviews, although extra knowledge is required to increase the recall of classification. Therefore, future work aims at extending FactNet by assessing more datasets of LOD and adding new knowledge from other sources such as biomedical ontologies.

Although we focused on the polarity classification task, the same approach could be applied to other tasks such as drug-to-drug interaction and side effects detection. In addition to its applications for polarity classification in drug domain, the extracted generalized patterns may also be used for polarity classification in other related domains. Thus, in future, we intend to exploit the generalized patterns in other domains and for other tasks.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] L. Eaton, Europeans and Americans turn to internet for health information, BMJ: Brit. Med. J. 325 (2002) 989.
[2] K. Zickuhr, Generations 2010, Pew Int. Am. Life Project (2010) 16.
[3] C.A. Brownstein, J.S. Brownstein, D.S. Williams, P. Wicks, J.A. Heywood, The power of social networking in medicine, Nat. Biotechnol. 27 (2009) 888–890.
[4] S. Abe, K. Inui, K. Hara, H. Morita, C. Sao, M. Eguchi, et al., Mining personal experiences and opinions from Web documents, Web Intel. Agent Syst. 9 (2011) 109–121.
[5] A. Reyes, P. Rosso, On the difficulty of automatically detecting irony: beyond a simple case of negation, Knowl. Inf. Syst. (2013) 1–20.
[6] T. Wilson, Annotating Subjective Content in Meetings, in: LREC, Citeseer, 2008.
[7] K. Inui, S. Abe, K. Hara, H. Morita, C. Sao, M. Eguchi, et al., Experience mining: building a large-scale database of personal experiences and opinions from web documents, in: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, IEEE Computer Society, 2008, pp. 314–321.
[8] R. Feldman, Techniques and applications for sentiment analysis, Commun. ACM 56 (2013) 82–89.
[9] C. Bizer, T. Heath, T. Berners-Lee, Linked data-the story so far, Int. J. Semantic Web Inform. Syst. 5 (2009) 1–22.
[10] F. Ciravegna, A.L. Gentile, Z. Zhang, LODIE: linked open data for web-scale information extraction, SWAIE 925 (2012) 11–22.
[11] D. Damljanovic, K. Bontcheva, Named entity disambiguation using linked data, in: Proceedings of the 9th Extended Semantic Web Conference, 2012.
[12] A. Chen, Patient Experience in Online Support Forums: Modeling Interpersonal Interactions and Medication Use, ACL (Student Research Workshop), 2013, pp. 16–22.
[13] V. Jijkoun, M. de Rijke, W. Weerkamp, P. Ackermans, G. Geleijnse, Mining user experiences from online forums: an exploration, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media: Association for Computational Linguistics, 2010, pp. 17–18.
[14] W. Murphy, Using Supervised Learning to Identify Descriptions of Personal Experiences Related to Chronic Disease on Social Media, The Pennsylvania State University, 2014.
[15] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Ret. 2 (2008) 1–135.
[16] B. Liu, Sentiment analysis and opinion mining, Synth. Lect. Human Lang. Technol. 5 (2012) 1–167.
[17] E. Cambria, A. Hussain, Sentic Computing: Techniques, Tools, and Applications, Springer, Dordrecht, Netherlands, 2012.
[18] A. Esuli, F. Sebastiani, Sentiwordnet: a publicly available lexical resource for opinion mining, in: Proceedings of LREC, 2006, pp. 417–422.
[19] E. Cambria, R. Speer, C. Havasi, A. Hussain, SenticNet: a publicly available semantic resource for opinion mining, in: AAAI Fall Symposium: Commonsense Knowledge, 2010, p. 02.
[20] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10: Association for Computational Linguistics, 2002, pp. 79–86.
[21] I. Habernal, T. Ptáček, J. Steinberger, Supervised sentiment analysis in Czech social media, Inf. Process. Manage. 50 (2014) 693–707.
[22] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (2011) 1138–1152.
[23] C. Toprak, N. Jakob, I. Gurevych, Sentence and expression level annotation of opinions in user-generated discourse, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics: Association for Computational Linguistics, 2010, pp. 575–584.
[24] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, L. Donaldson, Machine learning and sentiment analysis of unstructured free-text information about patient experience online, Lancet 380 (2012) S10.
[25] A. Tanvir, M. Sokolova, D. Schramm, D. Inkpen, Opinion Learning from Medical Forums, RANLP, 2013, pp. 18–24.

[26] M. Sokolova, V. Bobicev, Sentiments and Opinions in Health-related Web messages, in: RANLP, Citeseer, 2011, pp. 132–139.
[27] W. Wang, L. Chen, M. Tan, S. Wang, A.P. Sheth, Discovering Fine-grained sentiment in suicide notes, Biomed. Inform. Insights 5 (2012) 137.
[28] J. Parker, Y. Wei, A. Yates, O. Frieder, N. Goharian, A framework for detecting public health trends with Twitter, in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining: ACM, 2013, pp. 556–563.
[29] J.-C. Na, W.Y.M. Kyaing, C.S. Khoo, S. Foo, Y.-K. Chang, Y.-L. Theng, Sentiment Classification of Drug Reviews Using a Rule-Based Linguistic Approach. The Outreach of Digital Libraries: A Globalized Resource Network, Springer, 2012, pp. 189–198.
[30] D. Yalamanchi, Sideffective-System to Mine Patient Reviews: Sentiment Analysis, Rutgers University-Graduate School-New Brunswick, 2011.
[31] L. Goeuriot, J.-C. Na, W.Y. Min Kyaing, C. Khoo, Y.-K. Chang, Y.-L. Theng, et al., Sentiment lexicons for health-related opinion mining, in: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, ACM, 2012, pp. 219–226.
[32] M.T. Wiley, C. Jin, V. Hristidis, K.M. Esterling, Pharmaceutical drugs chatter on Online Social Networks, J. Biomed. Inform. (2014).
[33] S.M. Mohammad, P.D. Turney, Crowdsourcing a word–emotion association lexicon, Comput. Intel. 29 (2013) 436–465.
[34] H. Yang, R. Swaminathan, A. Sharma, V. Ketkar, D.S. Jason, Mining Biomedical Text towards Building a Quantitative Food-Disease-Gene Network. Learning Structure and Schemas from Documents, Springer, 2011, pp. 205–225.
[35] Q. Miao, S. Zhang, Y. Meng, Y. Fu, H. Yu, Healthy or harmful? Polarity analysis applied to biomedical entity relationships, in: PRICAI 2012: Trends in Artificial Intelligence, Springer, 2012, pp. 777–782.
[36] M. Samwald, A. Jentzsch, C. Bouton, C.S. Kallesøe, E. Willighagen, J. Hajagos, et al., Linked open drug data for pharmaceutical research and development, J. Cheminform. 3 (2011) 19.
[37] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web. Sci. Am. 284 (2001) 28–37.
[38] M. Gupta, J. Han, Applications of pattern discovery using sequential data mining, in: Pattern Discovery Using Sequence Data Mining: Applications and Studies, 2012, pp. 1–23.
[39] R. Srikant, R. Agrawal, Mining Sequential Patterns: Generalizations and Performance Improvements, Springer, 1996.
[40] A. Sharma, R. Swaminathan, H. Yang, A verb-centric approach for relationship extraction in biomedical text, in: Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on: IEEE, 2010, pp. 377–385.
[41] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, in: Proceedings of the AMIA Symposium: American Medical Informatics Association, 2001, p. 17.
[42] A. Moro, A. Raganato, R. Navigli, Entity linking meets word sense disambiguation: a unified approach, Trans. Assoc. Comput. Linguist. 2 (2014).
[43] O. Bodenreider, A.T. McCray, Exploring semantic groups through visual approaches, J. Biomed. Inform. 36 (2003) 414–432.
[44] C. Barrikre, From a Children's First Dictionary to a Lexical Knowledge Base of Conceptual Graphs, Simon Fraser University, 1997.
[45] C. Fellbaum, WordNet, Wiley Online Library, 1998.
[46] R. Navigli, S.P. Ponzetto, BabelNet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artif. Intell. 193 (2012) 217–250.
[47] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, J. Biomed. Inform. 34 (2001) 301–310.
[48] Y. Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3–Volume 3, Association for Computational Linguistics, 2009, pp. 1533–1541.
[49] J.L. Fleiss, B. Levin, M.C. Paik, Statistical Methods for Rates and Proportions, John Wiley & Sons, 2013.
[50] M. Mintz, S. Bills, R. Snow, D. Jurafsky, 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: Volume 2–Volume 2, Association for Computational Linguistics, 2009, pp. 1003–1011.
[51] K.-L. Liu, W.-J. Li, M. Guo, Emoticon Smoothed Language Models for Twitter Sentiment Analysis, AAAI, 2012.
[52] J. Broß, Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques, Freie Universität Berlin, 2013.
[53] E. Cambria, A. Hussain, Sentic Computing, Springer, 2012.
[54] E. Cambria, T. Benson, C. Eckl, A. Hussain, Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality, Expert Syst. Appl. 39 (2012) 10533–10543.
[55] S. Poria, E. Cambria, G. Winterstein, G.-B. Huang, Sentic Patterns: Dependency-based Rules for Concept-Level Sentiment Analysis, Knowledge-Based Systems, 2014.
[56] T. Flati, D. Vannella, T. Pasini, R. Navigli, Two is Bigger (and better) than one: the Wikipedia Bitaxonomy Project, Proc of ACL, 2014, pp. 945–955.