CrossMark

Unsupervised Commonsense Knowledge Enrichment for Domain-Specific Sentiment Analysis

Nir Ofek[1] · Soujanya Poria[2] · Lior Rokach[1] · Erik Cambria[3] · Amir Hussain[4] ·
Asaf Shabtai[1]

Abstract Sentiment analysis in natural language text is a challenging task involving a deep understanding of both syntax and semantics. Leveraging the polarity of multiword expressions—or concepts—rather than single words can mitigate the difficulty of such a task as these expressions carry more contextual information than isolated words. Such contextual information is the key to understanding both the syntactic and semantic structure of natural language text and hence is useful in tasks such as sentiment analysis. In this work, we propose a new method to enrich SenticNet (a publicly available knowledge base for concept-level sentiment analysis) with domain-level concepts composed of aspects and sentiment word pairs, along with a measure of their polarity. We process a set of unlabeled texts and, by considering the statistical co-occurrence information, generate a direct acyclic graph (DAG) of concepts. The polarity score of known concepts is propagated and used to compute polarity scores of new concepts. By designing and implementing our exhaustive algorithm, we are able to use a seed set containing only two sentiment words (*good* and *bad*). In our evaluation conducted on a dataset of hotel reviews, SenticNet was enriched by a factor of three (from 30,000 to nearly 90,000 concepts). The experiments demonstrate the merit of the concepts discovered by our method at improving sentence-level and aspect-level sentiment analysis tasks. Results of the two-factor ANOVA statistical test showed a confidence level of 95 %, verifying that the improvements are statistically significant.

Keywords Sentiment analysis · Sentiment lexicon · SenticNet · Sentic patterns

Introduction

Sentiment analysis is a task conducted to predict sentiment orientation by analyzing sentiment terms in subjective texts.

Distilling sentiment from text is a complex text-mining challenge, which, unlike simple text categorization, relies heavily on an understanding of context, such as negation. In addition, some words do not carry any specific polarity of their own but acquire it in context; for example, the adjective *cold* is positive in the context of beer and negative in the context of hotel's room.

An emerging approach for sentiment analysis suggests deconstructing text into small meaning units, i.e., concepts—semantic forms of human language—and assigning emotions to such concepts. Using polarity of natural language concepts can mitigate the complexity of the task, since understanding concepts does not require a great deal of familiarity with the language [5]. Concepts may consist of a product's feature described by an opinion word (*small room*) or an expression (*keep alive*). Concept-level sentiment analysis aims to infer the semantics associated with natural language opinions, thereby facilitating comparative

✉ Nir Ofek
nirofek@bgu.ac.il

Soujanya Poria
soujanya.poria@gmail.com

[1] Department of Information Systems Engineering, Ben Gurion University, Beersheba, Israel

[2] Temasek Laboratories, Nanyang Technological University, Singapore, Singapore

[3] School of Computer Engineering, Nanyang Technological University, Singapore, Singapore

[4] Division of Computing Science and Maths, School of Natural Science, University of Stirling, Scotland, UK

fine-grained feature-based sentiment analysis [4]. Instead of extracting opinions about an entire product (e.g., iPhone), users are generally more interested in comparing specific aspects of different products (e.g., iPhone's touch screen versus that of the Samsung).

SenticNet 3 [4] is a common and commonsense knowledge base of concepts along with their polarity scores. It is useful for sentiment analysis tasks such as feature spotting and polarity detection. It provides the semantics and polarity scores associated with 30,000 multiword expressions (concepts) and enables a deeper and multifaceted analysis of natural language opinions. The majority of these concept sentiments are unambiguous; therefore, SenticNet 3 does not contain domain-specific knowledge. In this work, we construct concepts based on paired *aspects* and *sentiment words*, in which the aspect determines the context and is therefore used to disambiguate the polarity of the sentiment word. For example, the concept <aspect:*beer*, word:*cold*>, carries a positive polarity, while the concept <aspect:*room*, word:-*cold*> carries a negative polarity. This is because the polarity of the word *cold* is ambiguous and dependent on the context. Such paired aspects and sentiment words are the types of concepts we construct.

The suggested method uses a set of unlabeled, opinion-based texts. However, in some domains and Web sites, more information can be found, such as user ratings of reviews. Weakly supervised sentiment analysis works use the overall ratings of reviews in training machine learning algorithms [27, 30, 31]. However, they are limited to specific domains that offer a rating in addition to the text itself. The problem with using unlabeled data to generate domain-specific sentiment information is addressed by several studies [12, 24, 29, 31]. However, they all depend upon a large set of seed words (from hundreds to thousands) to determine the polarity of new words. This requires manual work performed by experts[1] and is not easy to adapt to other domains and languages. We have designed an algorithm that simplifies the process of assigning polarity to concept and interactively expands the concept graph. This is accomplished by: (1) using only two sentiment seed words: *good* and *bad*; these two words, which have unambiguous polarity, are intuitive for use in sentiment analysis [15] and can easily be obtained based on the concept's more common usage and a degree in any language, and (2) information that was learned in previous iterations to propagate through the graph to compute the polarity of new concepts; thus, the polarity is of sentiment, which can be computed as opposed to a coarse binary score.

The proposed method starts by constructing a multiword concept in the form of an aspect-sentiment word pair. This involves employing a rule-based approach for aspect detection which uses a dependency parser. Since in addition to constructing concepts, we wish to compute their polarity, the next step involves assigning polarity to concepts, and it is this step that represents the novelty of our method. We generate a direct acyclic graph (DAG) of concepts for each aspect separately, in which edges represent constrained co-occurrence information between concepts. The idea is to infer the polarity of new concepts based on sentiment information of a known concept. The inference is subjected to a rigid set of constraints, which aims to improve the accuracy of a concept's polarity. Co-occurrences that do not comply with our set of constraints do not contribute to the inference process. To compensate this information loss, the process is iterative in which repeatedly new knowledge is extracted from the complete set of texts until exhaustiveness. Designing an exhaustive process has been made possible by the advent of distributed frameworks, which enable processing data and spanning information on multiple machines.

SenticNet 3 was designed to boost sentiment analysis tasks such as feature spotting and polarity detection. Therefore, in order to evaluate the contribution of the enriched knowledge base with respect to the original, we employ a state-of-the-art method that utilizes SenticNet knowledge for sentiment analysis.

To our knowledge, this is the first unsupervised method to employ this approach for domain-specific sentiment knowledge enrichment. Experimental results statistically demonstrate the merit of the generated concepts in improving sentence-level and aspect-level sentiment analysis tasks.

Related Work

Traditionally, sentiment analysis techniques were applied at the document level [8]. Recently, it has been recognized that even if a document bears a negative classification, it can contain some positive indicators. For this reason, researchers have developed an increasing interest in applying opinion mining techniques at a more granular level, specifically at the phrase or sentence level [30], and at the aspect level [17, 18, 27]. A number of studies utilize other information in addition to text (e.g., user ratings) for domain-specific sentiment tasks [9, 18, 27, 28]. The last two utilize overall ratings in computing the sentiment of specific aspects, based on the assumption that the overall rating of a review is correlated with the sentiment of its aspects. Since they use the overall rating to address a more granular problem, they are referred to as weakly supervised

---

[1] Huang et al. [12] suggest automatically mining seed terms from lists of pros and cons; however, this technique depends on the availability of such information.

learning and are bounded to text sources that provide the overall rating. We review only corpus-based approaches that use unlabeled opinionated texts for domain-specific sentiment analysis.

Wu and Wen [29] determine the polarity of sentiment terms based on linguistic patterns. The pattern, such as '<object> is a little too <attribute>,' is defined manually in Chinese and is used to disambiguate the polarity of 14 adjectives (such as *large* and *small*) that frequently comply with these patterns. Zhao et al. [31] use existing sentiment lexicons to infer the polarity of unknown words. They use a set of 3730 positive and 3116 negative Chinese words as a seed. However, it is unclear how sensitive their method is to various sizes of seed words. Kanayama and Nasukawa [16] construct lexicons using context coherency, both the intra- and inter-sentential co-occurrences, to learn the polarity of terms. Huang et al. [12] addressed the problem of obtaining seed words. They suggest automatically mining terms related to pros and cons from the Web to be used as seed words. This approach, however, relies on the availability of this information in the target domain. Brody and Elhadad [3] use the label propagation algorithm [32] to infer the polarity of sentiment words by using a large general-purpose lexicon as a seed set. The inference is done by using conjunction patterns, but they ignore reverting patterns, which limits their method's coverage. Our work is related to the research of Qiu et al. [24], which exploits statistical co-occurrence information found in a large domain corpus. Their work employs a bootstrapping process, which begins with a seed lexicon of 1752 sentiment terms taken from Hu and Liu [11]. Their expansion process involves utilizing syntactic relations that link sentiment words and target aspects. They formulate eight rules that are based on a syntactic parser output. However, apart from using a large seed set, they do not construct lexicons and concepts, which cannot be composed. Additionally, they rely upon homogeneous assumption, which may include inaccurate information in the inference process, while we use more rigid constraints. Overall, most discussed methods require a large set of seed words that pose difficulties in adapting them to other domains and languages. The suggested method utilizes only two trivial sentiment seed words (good, bad) and two extraction patterns (and, but); hence, they can be easily adapted to other domains and languages.

Typically, Web ontologies and semantic networks are used in concept-based approaches of sentiment analysis. This facilitates the aggregation of conceptual and affective information associated with natural language opinions. The reliance on large semantic knowledge bases diverges from the blind use of lexical and co-occurrence counts and relies upon the informative features associated with natural language concepts.

The main approaches for concept-level sentiment analysis leverage existing affective knowledge bases, including SenticNet [4], SentiWordNet [10], WordNet-Affect [25], and ANEW [2]. Tsai et al. [26] construct a concept-level dictionary through a two-step method combining iterative regression and random walk with in-link normalization. SenticNet assumes that semantically related concepts share relatively similar sentiment in propagating sentiment values.

Concept Construction and Polarity Assignment

Our method begins by extracting concepts by using a dependency parser and a set of constraints. This generates a multiword concept in the form of an aspect–sentiment word pair. Since we want to create concepts and also compute their polarity, the second step involves assigning polarity to concepts, and it is this step that represents the novelty of our method: Separately, for each aspect, we generate a direct acyclic graph (DAG) of concepts in which existing concepts are used to infer the polarity of new concepts based on constrained co-occurrence information.

Concept Construction

Definition Concept: In standard human to human communication people usually rely on the presumption that facts or definitions are known and proceed to build upon that. These known facts and definitions are called commonsense knowledge. A concept is an entity that defines the commonsense knowledge. This commonsense knowledge is often taken for granted by humans. In this study, we construct a concept in this way—a *concept $<a, jj>$* is constructed for a pair comprised of aspect a and adjective jj, which co-occur in the same sentence and comply with three constraints: (1) the aspect and adjective are interdependent, and the dependency type is one of the following {'*amod*', '*nsubj*', '*dep*'}, where '*amod*' captures adjectival modifier, '*nsubj*' captures a noun phrase, which is the syntactic subject of a clause, and '*dep*' is an unlabeled dependency. (2) The dependency *governor* is the aspect a, and its POS is either NN or NNS (singular or plural noun). (3) The dependency *dependent* is the adjective jj, and its POS is JJ, JJR (adjectives with the comparative ending), or JJS (adjectives with the superlative ending).

The process of concept construction begins with employing a rule-based approach for aspect detection, which uses a dependency parser. It follows by identifying connected sentiment words for each aspect.

Aspect Detection

We employ the unsupervised rule-based approach developed by Poria et al. [22] to extract aspects from unlabeled opinion-rich text. At a preliminary stage, we obtain the dependency parse tree of each sentence using the Stanford Parser.[2] In this work, we only consider nouns and noun phrases as aspects. Hence, we exclude the rules that extract implicit aspects, e.g., expensive and heavy. The rules for aspect extraction are categorized into two categories: The rules for sentences that have a 'subject–noun' relation in the dependency parse tree and other types of rules. In the following, we show when the rules are triggered and the corresponding behavior.

Trigger: when the active token is found to be the syntactic subject of a token.

Behavior: if an active token h is in a subject–noun relationship with a word t, then:

- Rule 1—if t has any adverbial or adjective modifier and the modifier exists in SenticNet, then we extract t as an aspect.
- Rule 2—if the sentence does not have auxiliary verb, i.e., *is, was, would, should, could*, etc., then:
  - Rule 2.1—if the verb t is modified by an adjective or adverb or is in adverbial clause modifier relation with another token, then h is extracted as an aspect. Notice the excerpt 'the battery lasts little.' *Battery* is in a subject relation with *lasts* and *lasts* is modified by an adjective modifier *little*, so *battery* is extracted as an aspect.
  - Rule 2.2—if t has any direct object relation with a token n and the part of speech of the token is noun and n is not in SenticNet, then n is extracted as an aspect.
    - 'I like the lens of this camera.' Here, *lens* is extracted as an aspect.
- Rule 3—a copula is the relation between the complement of a copular verb and the copular verb. If the token t is in copula relation with a copular verb and the part of speech of h is noun, then h is extracted as an explicit aspect. In the sentence, ' the camera is nice' camera is extracted as an aspect.

The Importance of Adjectives to Sentiment Analysis

People use adjectives to express subjective information, as indicated by Hatzivassiloglou and Wiebe [14]. Since the

---

function of adjectives is to characterize nouns, using them in sentiment analysis is considered a logical choice [15]. Therefore in this work we consider sentiment words to be adjectives, which is in line with previous studies on aspect-level sentiment analysis [3, 24, 29]. Even without this consideration, Blair-Goldensohn et al. [1] report that adjectives comprise 90 % of their sentiment lexicon; in the following excerpt an adjective (spacious) conveys the sentiment on the noun (lobby): "the lobby is spacious but the room is small." Once aspects are detected (as previously described), we aim at constructing concepts by pairing aspects with their modifying adjectives.

Connecting Aspects with Adjectives

For each aspect a, given a sentence s, we seek to identify the set of adjectives in s which are semantically related to a and associate them with the aspect a.

Because natural languages are versatile and do not always comply with simple rules, connecting aspects to their modifying adjectives is not trivial. In the previous example, applying a rule that connects adjectives with their nearest noun will construct the concept <room, spacious>. Thus, connecting adjectives with corresponding nouns becomes crucial for concept-level sentiment analysis.

In this work, use the Stanford Parser output in order to identify relevant information. This includes sentence segmentation, dependency parsing, word tokenization, and part-of-speech (POS) tagging.

Polarity Assignment Using Direct Acyclic Graph (DAG)

Assigning polarity to concepts is an iterative process, which generates a direct acyclic graph of concepts according to contextual evidence. To our knowledge, this is the first unsupervised method to employ this approach for domain-specific sentiment knowledge enrichment. In each iteration, known concepts are used to compute the polarity of new concepts. The assumption is that concepts involving the same aspect in the same sentence have a semantic relationship. This process is employed separately for every identified aspect.

Definition Similar Polarity Relation (SPR): A *similar polarity relation* between two concepts means that their adjectives are either: (1) conjunct by '*and*' conjunction term, and none or both of them is negated, or (2) conjunct by '*but*' conjunction term, and one of them is negated.

Definition Inverse Polarity Relation (IPR): An *inverse polarity relation* between two concepts means that their adjectives are either: (1) conjunct by the conjunction term

'*but*,' and none or both of them is negated, or (2) conjunct by '*and*,' with one of them negated.

We chose these two conjunction terms that are in line with previous research [13], since they can be easily obtained in any language and the parser identifies them with the dependency label *conj_and* and *conj_but*. To detect negated adjectives, we consider negation relations, which are identified by the parser and known prefixes (*mis, un, dis, im*).

Definition Interaction: Two occurrences of concepts *interact* with each other if: (1) they both occur in the same sentence, (2) they both involve the same occurrence of aspect, and (3) they comply with one of the two relation definitions: *SPR* or *IPR*.

The exhaustive algorithm generates the concepts graph for aspect *a*, *DAG(a)*, and computes the polarity for each concept. Algorithm 1 provides a detailed description for constructing a lexicon from which concepts are derived. The algorithm's main steps are described below:

Step 1: Construct all concepts *C* in corpus *D* involving aspect *a*. At the starting point (iteration 0), add the two concepts: *concept* <*a,good*> that is assigned with polarity = 1 (represents a positive polarity) and *concept* <*a,bad*> that is assigned with the polarity = 0 (represents a negative polarity) to the graph.

Step 2: For every *concept* <*a,jj₁*> ∈*C* that does not exist in *DAG(a)*, iterate through all of its occurrences in the text and count the number of *SPR interactions* and the number of *IPR interactions* with each *concept* <*a,jj₂*> ∈*C* that exists in *DAG(a)*. Add each *concept* <*a,jj₁*> to the graph with its incoming edges from *interacting* concepts and their *interaction* counters. Do not add concepts that have no

interaction with any existing concept in *DAG(a)*. Figure 1 illustrates the graph at the end of iteration *i* = 2. It can be seen that the concept <*pool,crowded*> has 16 relations in the corpus with the concept <*pool,wide*>. In 14 out of 16 times the relation type is *inconsistent*.

Step 3: Compute the polarity of new concepts that were added to the graph in the previous step by averaging the polarities of the concepts with which they *interact*. The average is weighted according to the number of *interactions*. In the case of *IPR*s, consider the inverse polarity of the concept (i.e., the absolute value of 1-polarity).

Steps 2 and 3 repeat until no new concept is added. In each iteration, only the concepts that were discovered in the previous iteration are involved with adding new concepts to the graph. This is because if they had *interacted* with concepts in a previous iteration, they would have already been added. The final polarity of each concept is in the range of [0:1], since the initial seed concept's assignment is '0' or '1.' Consider Fig. 1; the two seed concepts <*pool,good*> and <*pool,bad*> were used in expanding the graph to include the concept <*pool,wide*> since they are source nodes of its incoming edges (according to Step 2). Thus, according to the current step, the polarity of the node <*pool,wide*> is the weighted average of the two seed concepts:

$$\text{pol} <pool,wide> = [24 * \text{pol} <pool,good> + 2 \\ * \text{pol} <pool,bad> + 26 \\ * inverse_{pol} <pool,]/(24 + 2 + 26)$$

Since the polarity scores of <*pool,good*> and <*pool,bad*> are '1' and '0' respectively (according to Step 1):

Fig. 1 Expanding the concepts graph for the aspect pool. Labels above each edge represent consistent (*left*) and inconsistent (*right*) relation counters
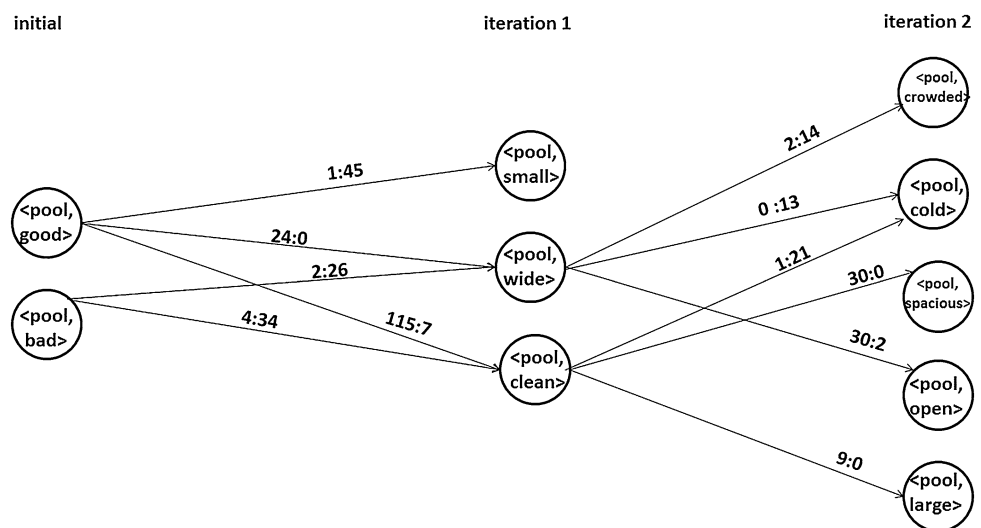
$$\text{pol} <pool, wide> = [24*1 + 2*0 + 26*(1 - 0)]/(24 + 2 + 26)$$
$$\approx 0.96$$

The polarity of the concept $<pool,wide>$ is strongly positive since the score is close to '1'. This node is used in computing the polarity of the node $<pool,crowded>$ according to the graph in the figure:

$$\text{pol} <pool, crowded> = [2*\text{pol} <pool, wide> + 14 * \text{inverse}_{pol} <pool, wide>]/(14 + 2)$$
$$= [2*0.96 + 14*(1 - 0.96)]$$

In computing the polarity of the concept $<pool, crowded>$ we can see how sentiment information propagates through the graph. The seed adjectives are used in computing the polarity of the node $<pool,wide>$ in the first iteration; in the second iteration this node is used in computing the polarity of the node $<pool,crowded>$ which eventually is found to be negative (relatively close to '0'.)

Evaluation and Discussions

In this section, we wish to evaluate the merit of the enriched SenticNet. Additionally, we evaluate the generated concepts by performing the aspect-level sentiment analysis task, independently from SenticNet; the polarity score that was computed for each concept is manually evaluated in a qualitative process.

Our empirical experiments were conducted on a benchmark dataset of nearly 200,000 hotel reviews complied by Wang et al. [27]. Each review contains ratings ranging from 1 to 5 stars for seven aspects: value, room, location, cleanliness, check-in/front desk, service, and business services, in addition to an overall rating. This information is not used in our learning process. The learning phase starts by parsing all of the reviews using the Stanford Parser, which includes sentence segmentation, dependency parsing, and POS tagging. Following this, it then extracts all aspects as described and generates a DAG for each aspect in order to construct concepts and to compute their polarity scores. When this process has been completed, we can use all concepts in the graph and their polarity. As mentioned before, we use the two sentiment words *good* and *bad* to compose seed concepts. Table 1 demonstrates concepts involving various aspects and the adjective *big*, where the polarity score is in the scale of '1' (very positive) to '0' (very negative). Next, we conduct several evaluations to investigate the merit of the constructed concepts.

Table 1 Examples of new generated concepts, compromising the adjective 'big'

Concept	Polarity
Big balcony	1
Big lounge	1
Big buffet	0.75
Big walk	0.23

Table 2 Precision of the top X positive concepts

Aspect	Precision at top			
	30	50	70	100
Room	0.97	0.9	0.91	0.91
Staff	1	1	0.99	0.99
Area	1	1	1	0.97

Table 3 Precision of the top X negative concepts

Aspect	Precision at top			
	30	50	70	100
Room	0.93	0.94	0.89	0.98
Staff	1	1	1	0.95
Area	1	1	0.99	0.99

Algorithm 1 Constructing Lexicon for aspect a

Input: SL – seed lexicon $\{c: polarity \mid \forall c, c.polarity = \begin{cases}1, if\ positive \\ 0, if\ negative\end{cases}\}$
Input: D – a corpus of reviews, segmented to sentences $\{s_1, ..., s_n\}$
Input: P – a set of patterns

**Output:** $lexicon(a)$ – the extended lexicon for aspect a

1. $M \leftarrow \emptyset$;
2. **foreach** $s_i \in D$ **do**
3.   Fill M with all *interactions* of concept pairs which are *connected* with aspect a;
4. **end for**;
5. $itr \leftarrow 0$;
6. Initialize $DAG(a)$ with SL;
7. **repeat**
8. $itr \leftarrow itr + 1$;
9.   Let c_x^i be concept x, where i represents the iteration (itr) in which a concept was added to $DAG(a)$
10.   **foreach** concept $c_x^i \in DAG(a)$ **do**
11.   **foreach** *interaction* in M involving c_x^i and c_y^j **where** ($c_y^j \notin DAG(a)$ **or** $j == itr$) **do**
12.   **if** $c_y^j \notin DAG(a)$
13. j itr;
14.   Add c_y^j to $DAG(a)$;
15.   **end if**;
16. $neg(c_x^i) \leftarrow$ detect negation of adjective in c_x^i; $neg(c_y^j) \leftarrow$ detect negation of adjective in c_y^j;
17. $p_k \leftarrow$ extract pattern in between c_x^i and c_y^j;
18.   Detect relation type using $neg(c_x^i), neg(c_y^j), p_k$;
19.   Update edge from c_x^i to c_y^j in $DAG(a)$ with counters correspond to the relation type;
20.   **end for**;
21.   **end for**;
22.   **foreach** c_y^j in $DAG(a)$, **where** $j == itr$ **do**
23. $c_x^i.polarity \leftarrow$ compute polarity using incoming edges;
24.   **end for**;
25.   **if** at least one concept was added to $DAG(a)$ in current iteration (itr) go to Step 7, **else** stop;
26. **end loop**;
27. $lexicon(a) \leftarrow \emptyset$;
28. **foreach** c_x^i in $DAG(a)$ **do**
29.   Add c_x^i to $lexicon(a)$;
30. **end for**;
31. **return** $lexicon(a)$

Evaluation of Lexicon Quality

There is no existing dataset available to evaluate the polarity scores of aspect-adjective concepts. In this section, we describe how we create a gold standard by using human

judges to annotate the lexicons generated based on a set of hotel reviews from TripAdvisor.

Three common aspects were sampled. As previously described, for each aspect our method constructs all concepts and computed their polarity scores. All concepts were ordered by their polarity score. The top X concepts are regarded as positive concepts (close to 1). Those falling at the bottom of that list are regarded as negative concepts, and they comprise the top X negative set (close to 0). Two human judges annotated the lists in the following way: (1) each concept was annotated twice as either being positive or negative, and (2) the annotators discussed any case of disagreement to resolve it unanimously (this occurred in
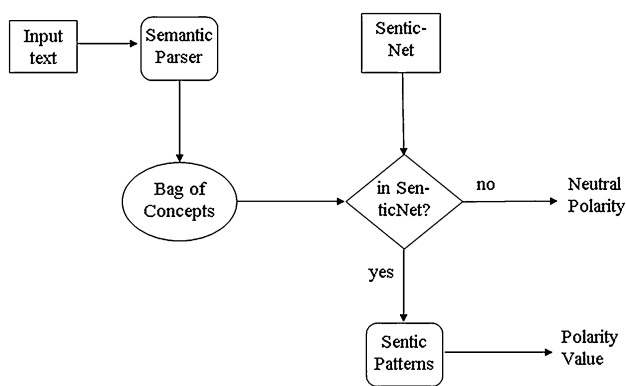
Fig. 2 Illustration of the classification procedure

Table 4 Concepts extracted using the original and enriched versions of SenticNet

Sentence	Concepts detected	
	SenticNet	Enriched SenticNet
1. Service was unbeatable, my executive room was first rate and the pool, sauna and other such-like facilities were excellent	Unbeatable Excellent	Unbeatable service Executive room Excellent facility
2. The cold towels and cool drinks after travelling for 16 h was a greatly appreciated touch	Appreciate Towel Travel	Cold towel Appreciate Travel
3. Hotel was immaculate	–	Immaculate hotel
4. The door to the room was warped and the room and hallway stunk big time	Stink	Stink Big time
5. The hotel was clean, the internet free, the gym functional and it came with a good breakfast	Clean Free Functional Breakfast Good	Clean hotel Free internet Functional Good breakfast

less than 5 % of the adjectives). The precision is presented in Tables 2 and 3.

Evaluation of Enriched SenticNet

Each of the learned concepts is first added to SenticNet 3, enriching the system's knowledge base. Our goal is to evaluate whether the enriched outperforms the existing SenticNet in predicting sentiment. Although the enriched knowledge base is useful in various sentiment analysis tasks, we evaluated its usefulness in sentence-level sentiment analysis. This task does not rely on additional procedures such as aspect segmentation. We randomly sampled 500 reviews from TripAdvisor from which (after removing non-English sentences and sentences containing a single word) we sampled 1000 sentences. The ground-truth scores were obtained by an annotating process involved two human judges where: (1) each sentence was annotated twice as either positive or negative, omitting sentences that contain no opinion, and (2) the annotators discussed any case of disagreement to resolve it unanimously (disagreements were found in less than 7 % of the sentences). After removing the neutral sentences from the dataset, 834 sentences remained; only 184 sentences were ultimately annotated as negative. Since negative sentences were underrepresented, we aimed to obtain a more balanced set, and we sampled an additional number of sentences and added 266 negative sentences to the set. The obtained dataset consists of 450 negative and 650 positive sentences. SenticNet 3 was designed to boost sentiment analysis tasks such as feature spotting and polarity detection. As a sentiment analysis framework, we used the method proposed by Poria et al. [21] which is a state-of-the-art method for sentence-level sentiment analysis using SenticNet concepts. This framework is a hybrid engine that consists of dependency-based patterns and an ELM (Extreme Learning Machine) supervised classifier for sentiment classification.

First, concepts are extracted from each sentence. Next, the model determines whether any of the concepts are present in SenticNet. Finally, dependency-based patterns are used for sentiment classification or alternatively the ELM classifier is used as a fallback method. In our evaluation, the ELM classifier is excluded from the framework, so we only use dependency based sentiment patterns. With this, we reduce external effects on the quality of the results, such as the quality of the ELM training data, as well as the sentences for which no concepts are found in either SenticNet. Figure 2 illustrates our procedure.

The polarity score of a sentence is a function of the polarity scores associated with its sub-constituents. In order to calculate those polarities, sentic patterns consider each of the sentence's tokens by following their linear order and

Table 5 Results of the two methods in sentence level sentiment classification per class

Method	Positive class			Negative class		
	Precision	Recall	F-measure	Precision	Recall	F-measure
SenticNet	86.36	93.54	89.81	89.39	78.67	83.69
Enriche SenticNet	90.93	95.69	93.25	93.27	86.22	89.61

Table 6 Examples of concepts extracted using the original and enriched versions of SenticNet

Aspect	Concepts*	LARA	LARA + Concepts**
Value	1.96 (42 %)	3.04	2.70
Room	1.44 (82 %)	2.90	1.80
Check-in/front desk	1.73 (54 %)	2.99	2.55
Service	1.89 (73 %)	2.59	2.18
Location	1.30 (59 %)	1.71	1.61
Cleanliness	1.79 (13 %)	3.70	3.60
Business service	2.88 (11 %)	3.31	3.16

* Concepts has an average coverage of nearly 60 %

** Statistically significant with respect to LARA

look at the dependency relations they entertain with other elements. A dependency relation is a binary relation characterized by the following features:

- The type of the relation that specifies the nature of the (syntactic) link between the two elements in the relation.

- The head of the relation: this is the element that is the pivot of the relation. Core syntactic and semantics properties (e.g., agreement) are inherited from the head.

- The dependent is the element that depends on the head and which usually inherits some of its characteristics (e.g., number, gender in the case of agreement).

Most of the time, the active token is considered in a relation if it acts as the head of the relation, although some rules are an exception. Once the active token has been identified as the trigger for a rule, there are several ways to compute its contribution depending on how the token is found on SenticNet. The preferred way is to consider the contribution not of the token alone, but in combination with the other element in the dependency relation. This crucially exploits the fact that SenticNet is not just a polarity dictionary, but also encodes the polarity of complex concepts. For example, in the sentence "The breakfasts were repetitive," the contribution of the noun *breakfast* will preferably be computed by considering the complex concept *repetitive breakfast* rather than the isolated concepts *breakfast* and *repetitive*.

If SenticNet has no entry for the multi-word concept formed by the active token and the element related to it, then the way individual contributions are taken into account depends on the type of the dependency relation. For example in the last example, *repetitive breakfast* is not found SenticNet but *repetitive* exists in SenticNet, so the polarity of *repetitive* is used to infer the polarity of the complex concept *repetitive breakfast*.

We found that in 673 sentences more concepts were extracted by using the enriched SenticNet comparing with the use of the original SenticNet. Example sentences and their detected concepts are given by Table 4.

For example, from the negative sentence 'the room was tiny and i mean tiny,' three concepts are extracted: mean room, room, and tiny room. None of them appear in SenticNet, while the concept tiny room appears in the enriched SenticNet, with a negative polarity score (close to 0).

The results of the sentence-level sentiment classification are presented in Table 5.

In the original SenticNet, concepts are found in 1058 (96 %) sentences, as compared to 1083 sentences (98 %) extracted concepts found in the enriched SenticNet.

Multiword concepts are detected in 723 sentences in the original SenticNet, whereas the enriched SenticNet detects multiword concepts in 942 sentences.

Evaluation of Generated Concepts

The following set of experiments aims to evaluate the merit of the generated concepts independently from SenticNet. This is performed by aspect-level sentiment analysis.

The sentiment of each aspect is computed based on the polarity of the concepts that include them. The goal is to predict user rating for any aspect in a given review (the ground truth (GT)). Our method of aspect-level sentiment analysis (denoted as Concepts) is as follows: Given an input text d and a target aspect a, we first extract all concepts involving a (as explained in "Polarity Assignment using Direct Acyclic Graph (DAG)" section) in d. To predict the GT score for aspect a in review r, we average the sentiment of all concept occurrences involving a in r. Negation is considered as explained in "Polarity

Assignment using Direct Acyclic Graph (DAG)" section. In the excerpt, 'the room was good but small,' two concepts are identified: <room,good> (polarity = 1), <room,small> (polarity = 0.13), and therefore, the sentiment score of room in the review is 0.56.

The coverage of Concepts system may be limited since it includes only aspect-adjective concepts, attempting to predict more accurate values at the expense of coverage. To investigate the impact of Concepts when combined with a more complex method, we combine it with the Latent Aspect Rating Analysis (LARA) system [27], a state-of-the-art method to predict the sentiment rating of aspects. LARA utilizes the overall rating of each review to predict the latent rating of aspects in the review, and therefore, it is limited to opinioned text, which is associated with overall user rating. The principal of LARA is to train a generative Latent Rating Regression (LRR) model aimed at predicting aspect ratings based on the review text and the associated overall rating; therefore, it is considered to be weakly supervised. LRR assumes that the overall rating is generated based on a weighted combination of the latent ratings over all the aspects, where the weights constitute the relative emphasis that the reviewer has placed on each aspect when giving the overall rating. A sentence is assigned to the aspect that shares the maximum term overlapping with the sentence.

LARA uses most words (not only adjectives) to convey sentiment and does not model the relation between the target aspect and the term, i.e., every term that co-occurs with the target aspect in the same sentence is considered to affect the sentiment. Hence, LARA has a greater degree of coverage. We employ a seamless integration model to combine the two: Concepts is first used to predict sentiment; in cases in which this returns no results, LARA is used as a fallback. This cascading approach allows us to assess the extent to which our method is able to increase accuracy without loss of coverage.

Since TripAdvisor provides only seven user-rated aspects, some aspects can be associated with the rated ones. Both LARA and Concepts are using the same procedure to map aspects to the rated aspects, employing a bootstrapping process to identify the major topical terms that correspond with each aspect, i.e., for each rated aspect whose sentiment is to be computed, a few seed terms are used to expand the terms set [27]. For example, the terms room, suite, view, and bed are specified to describe the aspect room. Therefore, the sentiment of room in review r will be the average of the sentiment of these aspects that appear in r.

The ground-truth rating provided by TripAdvisor is in the range of [1:5]; therefore, to evaluate the method, the sentiment score should be mapped to the same range (our original scores range from 0 to 1). Individually, for each rated aspect, we split the system's range into five sections, each corresponding to a single value in the range of [1:5] in accordance with the distribution of the GT ratings across the [1:5] range. For each rated aspect a, we order all of concepts involving a by their polarity (ascending). The first section of scores is mapped to the rating '1' (out of 5) in a threshold point p where the percentage of a's concepts in the range $[0:p]$ equals the percentage of the actual rating ($GT_a = 1$) in the dataset.

Table 6 shows the mean squared error (MSE) values for each aspect separately. The combined approach where LARA is used as a fallback system for our method (Concepts) outperforms the LARA system. That is true for all aspects. On its own, our method provides the most accurate MSE values; however, its coverage is more limited than that of LARA (around 60 % or 293,707 instances).

The two-factor ANOVA test with a confidence level of 95 % verifies that the differences in MSE between the combined method (Concepts + LARA) compared to LARA are statistically significant. The null hypothesis, which both methods perform the same and the observed differences are merely random, was rejected with $F(6,1) = 13.93236$ and p value = 0.0027 <5 %.

Table 5 indicates the coverage of Concepts comparing to LARA. Our method's coverage is on average 60 %, i.e., this rate reflects the percentage of instances that Concepts was able to predict sentiment to, out of all instances that LARA provided sentiment to (in some cases, LARA was not provide sentiment at all). This may be due to the fact that with our method (1) no adjective is connected to the aspect, or the adjective that is connected does not pertain to the corresponding set of concepts, (2) in some cases the aspect that appeared in the text which is associated with the rated aspect is not a noun or is a noun which rarely appears in the training corpus. For example, our method's coverage for the rated aspect cleanliness is relatively low. This aspect is represented by the following set of terms: clean, dirty, nonsmoking, valet, smoke, smell, tidy, maintain, smoker, resort, linen, cleanliness, musty, cigarette, spotlessly. Out of these, only seven are nouns, and some, such as cleanliness and linen, are rarely seen in the corpus. Moreover, some nouns, such as mosquitoes, can convey sentiment only by their presence or absence ('there are mosquitoes'), which our concepts do not cover.

Conclusions

In this paper, we propose a commonsense knowledge enrichment method for domain-specific sentiment analysis. The generated concepts are comprised of aspects and adjectives; based on the context aspect, the polarity of the concept is disambiguated. To our knowledge, this is the

first unsupervised method to employ this approach for domain-specific sentiment knowledge enrichment.

The generated concets can be utilized in many forms to perform sentiment analysis tasks. The merit of the enriched knowledge base demonstrated in preforming sentence level sentiment analysis. Comparing with the original knowledge base, the enriched SenticNet outperformed by all measures as Table 5 demonstrates.

Recall that in more than half of the evaluated sentences more concepts were extracted when using the enriched SenticNet. Such concepts carry a major role in classifying the polarity more accurately. Furthermore, it helps to explain the results by providing more context to sentiment words, as demonstrated by example five in Table 4. The merit of concepts generated by the presented method is well demonstrated when there are no concepts conveying polarity in the original knowledge base. Note the negative sentence "the room was tiny and i mean tiny." Three concepts are extracted: mean room, room, tiny room. None of them appear in SenticNet, while the concept "tiny room" appears in the enriched SenticNet, with a negative polarity score.

Apart from concepts, SenticNet 3 holds unambiguous words, mainly adjectives, such as great, excellent, bad, horrible, and good. In this study, the procedure used for sentiment analysis is capable of using the polarity of unambiguous adjectives, in determining the polarity of concepts that include them. Note the excerpt "the room was great and big." The concept extractor detects the concept "great room" which does not pertain to the original SenticNet. The polarity of the unambiguous adjective great is used on-the-fly to infer the polarity of the concept "great room". This can explain the relatively high coverage of the original SenticNet (96 %), comparing with the enriched knowledge base (98 %). Therefore, in our evaluation, generated concepts containing unambiguous adjectives have no impact on the results. Moreover, since the polarity of these adjectives is set as a-prior convention, they can constitute the seed concept set of the proposed methodology.

Since the polarity of many adjectives is ambiguous, the merit of the proposed methodology is in disambiguating their polarity by using domain-specific knowledge. For example, the polarity of the concept "big room" could only be determined by utilizing the proposed method, since the polarity of the adjective big depends on the modifying noun, hence it is not part of the original SenticNet. This is observed in our evaluation. Multiword concepts are detected in 723 and 942 sentences by the original and the enriched SenticNet respectively; using these multiword concepts facilitate the task of sentiment analysis by providing more multiword concepts with their polarity score. Table 1 shows examples of several concepts along their

polarity score, involving the adjective big. By encoding the context of the ambiguous adjective big into the concept, the polarity of the composed concepts can be inferred. For example, "big lounge" carries a positive polarity, while "big walk" carries a relatively negative polarity.

The usefulness of the generated concepts is demonstrated by a set of experiments, and the results statistically demonstrate the merit of the generated concepts in improving sentence-level and aspect-level sentiment analysis tasks. The proposed methodology is applicable to many domains, because it does not rely upon labeled data.

The proposed approach will also be enhanced by using a more advanced sentiment analysis algorithm [6, 7, 20]. We also plan to research the use of this approach on various text analysis applications e.g., personality detection [22], textual entailment [19].

References

1. Blair-Goldensohn, S, Hannan K, McDonald R, Neylon T, Reis GA, Reynar J. Building a sentiment summarizer for local service reviews. In: WWW workshop on NLP in the information explosion era (NLPIX), New York, NY, USA, 2008. ACM.
2. Bradley M, Lang P. Affective norms for English words (ANEW): stimuli, instruction manual and affective ratings. Technical report, the center for research in psychophysiology, University of Florida, 1999.
3. Brody S, Elhadad N. An unsupervised aspect-sentiment model for online reviews. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics pp 804–812. Association for Computational Linguistics 2010.
4. Cambria E, Olsher D, Rajagopal D. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. Quebec City: AAAI; 2014. p. 1515–21.
5. Cambria E, White B. Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag. 2014;9(2):48–57.
6. Chikersal P, Poria S, Cambria E. SeNTU: sentiment analysis of tweets by combining a rule based classifier with supervised learning. In: Proceedings of the International Workshop on Semantic Evaluation (SemEval 2015), 2015.

7. Chikersal P, Poria S, Cambria E, Gelbukh A, Siong CE. Modelling public sentiment in twitter: using linguistic patterns to enhance supervised learning. In: Computational linguistics and intelligent text processing, Springer International Publishing, pp 49–65.

8. Dave K, Lawrence S, Pennock D. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceeding of WWW, Budapest, Hungary, 20–24 May 2003. pp 519–528.

9. Demiroz G, Yanikoglu B, Tapucu D, Saygin Y. Learning domain-specific polarity lexicons. In: Data mining workshops (ICDMW), 2012 IEEE 12th international conference on 2012. pp 674–679.

10. Esuli A, Sebastiani F. SentiWordNet: a publicly available lexical resource for opinion mining. In: LREC, 2006.

11. Hu M, Liu B. Mining and summarizing cus-tomer reviews. In: SIGKDD 2004. pp 168–177.

12. Huang S, Niu Z, Shi C. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. Knowl Based Syst. 2014;56:191–200.

13. Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. In: Proceedings of ACL, 1997. pp 174–181.

14. Hatzivassiloglou V, Wiebe JM. Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the 18th conference on computational linguistics, vol 1, 2000. Association for Computational Linguistics.

15. Kamps J, Marx M, Mokken RJ, de Rijke M. Using word net to measure semantic orientation of adjectives. In: Proceedings of LREC'04, 2004. pp 1115–1118.

16. Kanayama H, Nasukawa T. Fully automatic lexicon expansion for domain oriented sentiment analysis. In: Proceedings of EMNLP, 2006. pp. 355–363.

17. Lu B, Ott M, Cardie C, Tsou BK. Multi-aspect sentiment analysis with topic models. In: Data mining workshops (ICDMW), 2011.

18. Lu Y, Castellanos M, Dayal U, Zhai C. Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: Proceedings of WWW, 2011. ACM pp. 347–356.

19. Pakray P, Neogi S, Bhaskar P, Poria S, Bandyopadhyay S, Gelbukh A. A textual entailment system using anaphora resolution. In: System Report. Text analysis conference recognizing textual entailment track notebook, November 2011.

20. Poria S, Cambria E, Gelbukh A, Bisio F, Hussain A. Sentiment data flow analysis by means of dynamic linguistic patterns. IEEE Comput Intell Mag. 2015;10(4):26–36.

21. Poria S, Cambria E, Ku LW, Gui C, Gelbukh A. A rule-based approach to aspect extraction from product reviews. SocialNLP 2014. p 28.

22. Poria S, Gelbukh A, Agarwal B, Cambria E, Howard N. Common sense knowledge based personality recognition from text. In: Advances in soft computing and its applications, Springer Berlin Heidelberg, 2013. pp 484–496.

23. Poria S, Gelbukh A, Hussain A, Howard N, Das D, Bandyopadhyay S. Enhanced SenticNet with affective labels for concept-based opinion mining. IEEE Intell Syst. 2013;28(2):31–8.

24. Qiu G, Liu B, Bu J, Chen C. Expanding domain sentiment lexicon through double propagation. In: IJCAI, vol. 9, pp 1199–1204, 2009.

25. Strapparava C, Valitutti A. WordNet-affect: an affective extension of WordNet. In: LREC, Lisbon, 2004. pp 1083–1086.

26. Tsai A, Tsai R, Hsu J. Building a concept-level sentiment dictionary based on commonsense knowledge. IEEE Intell Syst. 2013;28(2):22–30.

27. Wang H, Lu Y, Zhai C. Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedins of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. pp 783–792. ACM.

28. Weichselbraun A, Gindl S, Scharl A. Extracting and grounding context-aware sentiment lexicons. IEEE Intell Syst. 2013;28(2):39–46.

29. Wu Y, Wen M. Disambiguating dynamic sentiment ambiguous adjectives. In: Proceedings of COLING, 2010. pp 1191–1199.

30. Xu R, Wong KF, Lu Q, Xia Y, Li W. Learning knowledge from relevant webpage for opinion analysis. In: Proceedings of the 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, Sydney, Australia, 9–12 Dec 2008. pp 307–313.

31. Zhao Y, Qin B, Liu T. Collocation polarity disambiguation using web-based pseudo contexts. In: Proceedings of EMNLP/CoNLL, 2012. pp 160–170.

32. Zhu X, Ghahramani Z. Learning from labeled and unlabeled data with label propagation. Technical report CMU-CALD-02-107, Carnegie Mellon University, 2002.