Sentiment-Oriented Information Retrieval: Affective Analysis of Documents Based on the SenticNet Framework

Federica Bisio, Claudia Meda, Paolo Gastaldo, Rodolfo Zunino and Erik Cambria

Abstract Sentiment analysis research has acquired a growing importance due to its applications in several different fields. A large number of companies have included the analysis of opinions and sentiments of costumers as a part of their mission. Therefore, the analysis and automatic classification of large corpora of documents in natural language, based on the conveyed feelings and emotions, has become a crucial issue for text mining purposes. This chapter aims to relate the sentiment-based characterization inferred from books with the distribution of emotions within the same texts. The main result consists in a method to compare and classify texts based on the feelings expressed within the narrative trend.

Keywords Sentiment analysis · Text mining · Senticnet · SLAIR

1 Introduction

Sentiment analysis or opinion mining can be defined as a particular application of data mining, which aims to aggregate and extract emotions and feelings from different types of documents.

In recent years sentiment analysis has been applied especially to social networks, to analyze opinions on Twitter, Facebook or other digital communities in real time

F. Bisio (✉) · C. Meda · P. Gastaldo · R. Zunino
DITEN – University of Genoa, Via Opera Pia 11A, 16145 Genoa, Italy
e-mail: federica.bisio@edu.unige.it

C. Meda
e-mail: claudia.meda@edu.unige.it

P. Gastaldo
e-mail: paolo.gastaldo@unige.it

R. Zunino
e-mail: rodolfo.zunino@unige.it

E. Cambria
SCE – School of Computer Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore
e-mail: cambria@ntu.edu.sg

W. Pedrycz and S.-M. Chen (eds.), *Sentiment Analysis and Ontology Engineering*, Studies in Computational Intelligence 639, DOI 10.1007/978-3-319-30319-2_8

[21, 39, 50]. The emerging field of big (social) data analysis deals with information retrieval and knowledge discovery from natural language and social networks. Graph mining and natural language processing (NLP) techniques contribute to distill knowledge and opinions from the huge amount of information on the World Wide Web.

Sentiment analysis can enhance the capabilities of customer relationship management and recommendation systems by allowing, for example, to find out which features customers are particularly interested in or to exclude from ads or recommendations items that have received unfavorable feedbacks. Likewise, it can be used in social communication to enhance anti-spam systems.

Business intelligence can also benefit from sentiment analysis. Since predicting the attitude of the public towards a brand or a product has become of crucial importance for companies, an increasing amount of money is invested in marketing strategies involving opinion and sentiment mining.

That scenario led to sentic computing [10], which tackles those crucial issues by exploiting affective common-sense reasoning, i.e., the intrinsically human capacity to interpret the cognitive and affective information associated with natural language. In particular, sentic computing leverages on a common-sense knowledge base built through crowdsourcing [16, 19]. Common-sense is useful in many different computer-science applications including data visualization [11], text recognition [64], and affective computing [9]. In this context, common-sense is used to bridge the semantic gap between word-level natural language data and the concept-level opinions conveyed by these [15].

To perform affective common-sense reasoning [4], a knowledge database is required for storing and extracting the affective and emotional information associated with word and multi-word expressions. Graph-mining [13] and dimensionality-reduction [7] techniques have been employed on a knowledge base obtained by blending ConceptNet [56], a directed graph representation of common-sense knowledge, with WordNet-Affect (WNA) [58], a linguistic resource for the lexical representation of feelings. Unlike WNA, SenticNet [14] exploits an ensemble of common and common-sense knowledge to go beyond word-level opinion mining and, hence, to associate semantics and sentics to a set of natural language concepts.

The task of extracting emotions from a large corpus of natural language texts can be viewed from a different point of view: instead of extracting feelings from data in order to assign different sentiment categories to the analyzed corpuses (e.g., tweets, posts, messages), one may try and identify the similarities between any text based on their sentiment distribution. This work presents both a *semantic descriptor* using the SenticNet emotional database [14] and a *sentiment distance metric* which can characterize the emotional states of different documents (i.e., books), and classify them based on their sentiment distributions.

The experiments were performed implementing these functionalities within a text-mining tool called SLAIR [51] in order to automate the extraction of emotional distributions from different categories of books and classify them based on their feelings trend.

To test the proposed approach on a dataset of documents dense of sentimental and affective contents, the experiments involved a dataset of books selected from different literary genres. This approach provided a group of texts able to convey clear feelings and emotions, while offering at the same time a novel field of analysis.

In order to clarify the scope, the following example sentences convey different emotions:

(A)—"Pain and suffering are always inevitable for a large intelligence and a deep heart. The really great men must, I think, have great sadness on earth."

(B)—"Seldom, very seldom, does complete truth belong to any human disclosure; seldom can it happen that something is not a little disguised or a little mistaken."

(C)—"Never forget what you are, for surely the world will not. Make it your strength. Then it can never be your weakness. Armour yourself in it, and it will never be used to hurt you."

Reading these sentences, it is possible to notice how excerpts from different books can convey similar emotions, even if they were written by different authors in different works. For example, one notes an emotional alignment between sentence (A) and (B), intuitively suggested by the presence of deep thoughts about humanity. Indeed, sentences (A) and (B) were drawn from the same book genre (XIX century novels); sentence (A) was extracted from "Crime and Punishment" by Dostoevskij, whereas sentence (B) was extracted from "Emma" by Jane Austen. On the contrary, sentence (C) was extracted from "Game of Thrones" by G.R.R. Martin and clearly conveys a heartening and encouraging tone, although yet involving again a reflection on humans. Details on the distance metric results related to these sentences will be given later in the experimental results section.

The rest of this chapter is organized as follows: Sect. 2 introduces related work in the field of sentiment analysis research; Sect. 3 describes the affective resource SenticNet; Sect. 4 illustrates the techniques employed for the analysis of the feelings distribution; Sect. 5 explains the experimental setup and the metrics used; Sect. 6 illustrates experimental results; finally, Sect. 7 offers some concluding remarks.

2 Sentiment Analysis and Opinion Mining

The Social Web has provided people with new content-sharing services that allow to create and share personal contents, ideas and opinions, in a time- and cost-efficient way, with virtually millions of other people connected to the World Wide Web. Since this amount of information is mainly unstructured, research has so far focused on online information retrieval, aggregation, and processing. Moreover, when it comes to interpreting sentences and extracting meaningful information, these tasks become very critical. NLP requires high-level symbolic capabilities [25], including:

- creation and propagation of dynamic bindings;
- manipulation of recursive, constituent structures;
- acquisition and access of lexical, semantic, and episodic memories;
- representation of abstract concepts.

All these capabilities are required to shift from mere NLP to what is usually called natural language understanding [1].

Therefore, opinion mining and sentiment analysis have recently emerged as a challenging and active field of research, due to many open problems and a wide variety of practical applications. The potential applications of concept-level sentiment analysis are indeed countless and span interdisciplinary areas, such as stock market prediction, political forecasting, social network analysis, social stream mining, and man-machine interactions.

Today, most of the existing approaches still rely on word co-occurrence frequencies, i.e., the syntactic representation of text. Therefore, computational models aim to bridge the cognitive gap by emulating the way the human brain processes natural language. For instance, by leveraging on semantic features that are not explicitly expressed in text, one may accomplish complex NLP tasks such as word-sense disambiguation, textual entailment, and semantic role labeling. Computational models are useful for both scientific purposes (such as exploring the nature of linguistic communication), and practical purposes (such as enabling effective human-machine communication).

Most existing approaches to sentiment analysis rely on the extraction of a vector representing the most salient and important text features, which is later used for classification purposes. Commonly used features include term frequency, presence and the position of a token in a text. An extension of this feature is the presence of n-grams, typically bi-grams and tri-grams. Other methods rely on the distance between terms or on the Part-of-Speech (POS) information: for example, certain adjectives may be good indicators of sentiment orientation. A drawback of these approaches is the strict dependency on the considered domain of application and the related topics.

Sentiment analysis systems aim to classify entire documents by associating contents with some overall positive or negative polarity [44] or rating scores (e.g., 1–5 stars) [42]. These approaches are typically supervised and rely on manually labelled samples. We can distinguish between knowledge-based systems [17], based on approaches like keyword spotting and lexical affinity, and statistics-based systems [18]. At first, the identification of emotions and polarity was performed mainly by means of knowledge-based methods; recently, sentiment analysis researchers have been increasingly using statistics-based approaches, with a special focus on supervised statistical methods.

Keyword spotting is the most straightforward, and possibly also the most popular, approach thanks to its accessibility and economy. Text is classified into affect categories based on the presence of fairly unambiguous "affect words" like 'happy', 'sad', 'afraid', and 'bored'. Elliott's Affective Reasoner [26], for example, watches for 198 affect keywords plus affect intensity modifiers and a handful of cue phrases. Other popular sources of affect words are Ortony's Affective Lexicon [40], which groups terms into affective categories, and Wiebe's linguistic annotation scheme [66]. The crucial issue of this approaches lies in the ineffectiveness at handling negations and in the structure based on the presence of obvious affect words.

Rather than simply detecting affect words, lexical affinity assigns each word a probabilistic 'affinity' for a particular emotion. These probabilities are usually learnt from linguistic corpora [48, 55, 57]. Even if this method often outperforms pure keyword spotting, it still works at word level and can be easily tricked by negations and different senses of the same word. Besides, lexical affinity probabilities are often biased by the linguistic corpora adopted, which makes it difficult to develop a reusable, domain-independent model.

Statistical based approaches, such as Bayesian inference and support vector machines (SVM), have been used on several projects [24, 27, 44, 65]. By feeding a machine learning algorithm [63] a large training corpus of affectively annotated texts, it is possible not only to learn the affective valence of affect keywords (as in the keyword spotting approach), but the one of other arbitrary keywords (like lexical affinity), punctuation, and word co-occurrence frequencies. Anyway, it is worth noticing that statistical classifiers work well only when a sufficiently large text is given as input. This is due to the fact that, with the exception of affect keywords, other lexical or co-occurrence elements possess a little predictive value individually.

For example, Pang et al. [44] used the movie review dataset to compare the performance of different machine learning algorithms: in particular, they obtained 82.90 % of accuracy employing a large number of textual features. Socher et al. [54] proposed a recursive neural tensor network (RNTN) and improved the accuracy (85 %). Yu and Hatzivassiloglou [68] identified polarity at sentence level using semantic orientation of words. Melville et al. [36] developed a framework exploiting word-class association information for domain-dependent sentiment analysis. Reference [3] tackled a particular aspect of the sentiment classification problem: the ability of the framework itself to operate effectively in heterogeneous commercial domains. The approach adopts a distance-based predictive model to combine computational efficiency and modularity.

Some approaches exploit the following fact: many short n-grams are usually neutral while longer phrases are well distributed among positive and negative subjective sentence classes. Therefore, matrix representations for long phrases and matrix multiplication to model composition can also be used to evaluate sentiment. In such models, sentence composition is modeled using deep neural networks such as recursive auto-associated memories. Recursive neural networks (RNN) predict the sentiment class at each node in the parse tree and try to capture the negation and its scope in the entire sentence.

Several unsupervised learning approaches have also been proposed and rely on the creation of the lexicon via the unsupervised labeling of words or phrases with their sentiment polarity or subjectivity [43]. To this aim, early works were mainly based on linguistic heuristics. For example, Hatzivassiloglou and McKeown [28] built a system based on opposition constraints to help labeling decisions, in the case of polarity classification.

Other works exploited the seed words, defined as terms for which the polarity is known, and propagated them to terms that co-occur with them in general text, or in specific WordNet-defined relations. Popescu and Etzioni [46] proposed an algorithm that, starting from a global word label computed over a large collection of generic topic text, gradually tried to re-define such label to a more and more specific corpus, until the one that is specific to the particular context in which the word occurs. Snyder and Barzilay [53] also exploited the idea of utilizing discourse information to aid the inference of relationships between product attributes.

Regression techniques are often employed for the prediction of the degree of positivity in opinionated documents since they allow for modeling classes that correspond to points on a scale, such as the number of stars given by a reviewer [43]. Works like [41] attempted to address the problem via incorporating location information in the feature set and underlined the importance of the last sentences of a review in the calculation of the overall sentiment of a document.

One problem is represented by the fact that contrary attitudes can be present in the same document: therefore, a fine-grained level analysis can be useful to distinguish sentimental from non-sentimental sections, e.g., by using graph-based techniques for segmenting sections of a document on the basis of their subjectivity [41], or by performing a classification based on some fixed syntactic phrases that are likely to be used to express opinions [62], or by bootstrapping using a small set of seed opinion words and a knowledge base such as WordNet [31] or, finally, using unsupervised methodolgies [29].

In recent works, text analysis granularity has been taken down to sentence level, e.g., by using presence of opinion lexical items to detect subjective sentences [32, 49], or by using semantic frames defined in FrameNet [2, 33]. Since an author usually does not switch too frequently between adjacent sentences, a certain continuity level is preserved and some works propose a classification of the document based on assigning preferences for pairs of nearby sentences [42, 69].

Concept-based approaches [6, 30, 47, 61] focus on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the handling of conceptual and affective information associated with natural language opinions. By relying on large semantic knowledge bases, such approaches step away from blind use of keywords and word co-occurrence counts, but rather rely on the implicit meaning/features associated with natural language concepts. Unlike purely syntactical techniques, concept-based approaches are also able to detect sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but are implicitly linked to other concepts that do so.

More recent studies [20, 23, 34, 37] enhanced sentiment analysis of tweets by exploiting microblogging text or Twitter-specific features such as emoticons, hashtags, URLs, @symbols, capitalizations, and elongations. Tang et al. [59] developed a convolutional neural network based approach to obtain word embeddings for the words mostly used in tweets. These word vectors were then fed to a convolutional neural network for sentiment analysis. A deep convolutional neural network for sentiment detection in short text was also proposed by Santos et al. [52]. The approach based on Sentiment Specific Word Embeddings [60] considers word embeddings based on a sentiment corpora: this means including more affective clues than regular word vectors and producing a better result.

Finally recent researches faced the problem of identifying literary texts based on certain textual characteristics in common, for example [22], which limited the analysis to Dutch novels, or [38], that analyzed the narrative emotion related to the relationship of characters described into two different novels.

3 SenticNet

SenticNet[1] is a publicly available semantic and affective resource for concept-level sentiment analysis. The last release SenticNet 3 exploits 'energy flows' to connect different parts of both common and common-sense knowledge representations to one another, unlike standard graph-mining and dimensionality-reduction techniques. SenticNet 3 therefore models semantics and sentics (that is, the conceptual and affective information associated with multi-word natural language expressions).

To this aim, SenticNet 3 employs an energy-based knowledge representation [14] to provide the semantics and sentics associated with 30,000 concepts, thus enabling a fine-grained analysis of natural language opinions. SenticNet 3 contains both unambiguous adjectives as standalone entries (like 'good' and 'awful') and non-trivial multi-word expressions such as 'small room' and 'cold bed'. This is due to the fact that while unambiguous adjectives convey positive or negative polarities (whatever noun they are associated with), other adjectives are able to carry a specific polarity only when coupled with certain nouns.

SenticNet 3 focuses on the use of 'energy' or information flows to connect various parts of common and common-sense knowledge representations to one another. Each quantum of energy possesses a scalar magnitude, a valence (binary positive/negative), and an edge history, defined as a list of the edge labels that a particular quantum of energy has traversed in the past. Essentially, common and common-sense knowledge is broken down into 'atoms', thus allowing the fusing of data from different knowledge bases without requiring any ontology alignment.

3.1 Sources of Knowledge

SenticNet 3 embeds both common and common-sense knowledge, in order to boost sentiment analysis tasks such as feature spotting and polarity detection, respectively. In particular, it is generated by an ensemble of different methods. Regarding the common knowledge bases, the employed resources are either crafted by human experts or community efforts, such as DBPedia [5], a collection of 2.6 million entities extracted from Wikipedia, or automatically-built knowledge bases, such as Probase [67], Microsoft's probabilistic taxonomy counting about 12 million concepts learned iteratively from 1.68 billion web pages in Bing web repository.

Regarding common-sense knowledge, Open Mind Common Sense (OMCS) collects pieces of knowledge from volunteers on the Internet by enabling them to enter common-sense into the system with no special training or knowledge of computer science. OMCS has exploited these pieces of common-sense knowledge to build ConceptNet [56], a semantic network of 173,398 nodes.

[1] http://sentic.net/sentics.

3.2 Structure of SenticNet

The aggregation of common and common-sense knowledge bases is designed as a 2-stage process in which different pieces of knowledge are first translated into RDF triples and then inserted into a graph. Considering as an example 'Pablo Picasso is an artist', we obtain the RDF triple <Pablo Picasso-isA-artist> and, hence, the entry [artist—SUBSUME → Pablo Picasso]. In this way we obtain a shared representation for common and common-sense knowledge, thus performing a conceptual decomposition of relation types, i.e., the unfolding of relation types that are usually opaque in natural-language-based resources.

After low confidence score trimming and duplicates removal, the resulting semantic network (built out of about 25 million RDF statements) contains 2,693,200 nodes. Of these, 30,000 affect-driven concepts (that is, those concepts that are most highly linked to emotion nodes) have been selected for the construction of SenticNet 3.

SenticNet 3 conceptualizes the information as 'energy' and sets up pathways upon which this energy may flow between different semantic fragments. In this way, complex concepts can be built upon simpler pieces by connecting them together via energy flows. Once an element is reached by a certain quantum of energy flow, it is included in a wider concept representation, thus enabling simple elements to deeply affect larger conceptual connections. Such a representation is optimal for modeling domains characterized by nuanced, interconnected semantics and sentics (including most socially-oriented AI modeling domains).

Each quantum of energy possesses a scalar magnitude, a valence (binary positive/negative), and an edge history, defined as a list of the edge labels that a particular quantum of energy has traversed in the past. These three elements describe the semantics and sentics of the quantum of energy and they are extracted for each concept of the semantic network.

In particular, the extraction of semantics and sentics is achieved through multiple steps of spreading activation with respect to the nodes representing the activation levels of the Hourglass of Emotions [12], a brain-inspired model for the representation and the analysis of human emotions.

3.3 The Hourglass of Emotions

The Hourglass of Emotions is an affective categorization model developed starting from Plutchik's studies on human emotions [45]. The main advantage over other emotion categorization models is that it allows emotions to be deconstructed into independent but concomitant affective dimensions, whose different levels of activation make up the total emotional state of the mind. Such a modular approach to emotion categorization allows different factors (or energy flows) to be concomitantly taken into account for the generation of an affective state.

Table 1 The sentic levels of the Hourglass model

Interval	Pleasantness	Attention	Sensitivity	Aptitude
[G(1), G(2/3))	Ecstasy	Vigilance	Rage	Admiration
[G(2/3), G(1/3))	Joy	Anticipation	Anger	Trust
[G(1/3), G(0))	Serenity	Interest	Annoyance	Acceptance
(G(0), G(−1/3)]	Pensiveness	Distraction	Apprehension	Boredom
(G(−1/3), G(−2/3)]	Sadness	Surprise	Fear	Disgust
(G(−2/3), G(−1)]	Grief	Amazement	Terror	Loathing

The model can potentially synthesize the full range of emotional experiences in terms of four affective dimensions, Pleasantness, Attention, Sensitivity, and Aptitude, which determine the intensity of the expressed/perceived emotion as a $float \in [-1, +1]$. Each affective dimension is characterized by six levels of activation, termed 'sentic levels', which are also labeled as a set of 24 basic emotions (six for each affective dimension) (Table 1).

Previous works [8] already proved that a categorization model based on these four affective dimensions is effective in the design of an emotion categorization architecture.

The transition between different emotional states is modeled, within the same affective dimension, using the function $G(x) = -\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$, for its symmetric inverted bell curve shape that quickly rises up towards the unit value. In particular, the function models how valence or intensity of an affective dimension varies according to different values of arousal or activation, spanning from null value (emotional void) to the unit value (heightened emotionality). Mapping this space of possible emotions leads to a hourglass shape (Fig. 1).

Complex emotions can be synthesized by using different sentic levels, as shown in Table 2.

3.4 Semantics and Sentics Representation

The RDF triples are encoded in a XML format, in order to represent SenticNet 3 in a machine-processable way. For each concept, semantics and sentics are provided.

Given the concept 'celebrate special occasion', for example, SenticNet 3 provides a set of semantically related concepts, e.g., 'celebrate holiday', 'celebrate occasion' or 'celebrate birthday'. The resource also provides a sentic vector specifying Pleasantness, Attention, Sensitivity, and Aptitude associated with the concept (for tasks such as emotion recognition), a polarity value (for tasks such as polarity detection).

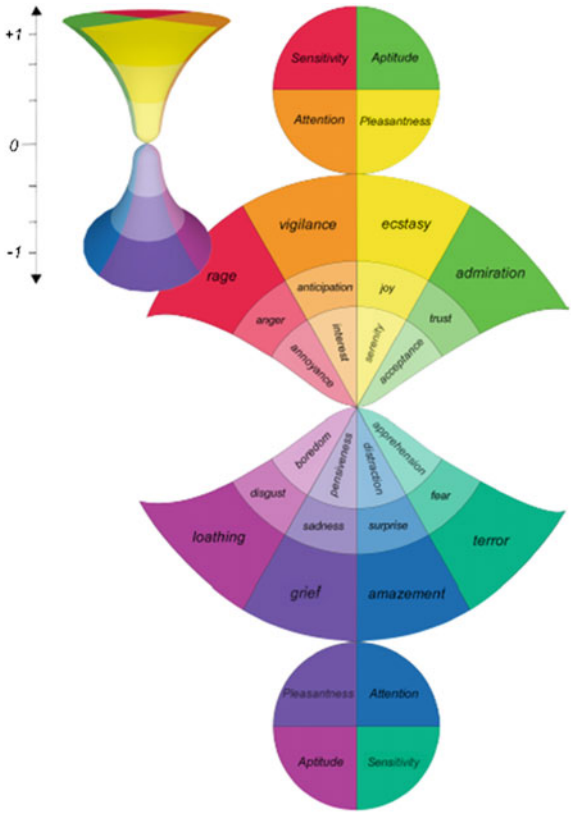

Fig. 1 The 3D model of the Hourglass of emotions

Table 2 The emotions generated by pairwise combination of the sentic levels of the Hourglass model

	Attention > 0	Attention < 0	Aptitude > 0	Aptitude < 0
Pleasantness > 0	Optimism	Frivolity	Love	Gloat
Pleasantness < 0	Frustration	Disapproval	Envy	Remorse
Sensitivity > 0	Aggressiveness	Rejection	Rivalry	Contempt
Sensitivity < 0	Anxiety	Awe	Submission	Coercion

4 Sentiment Distribution: A Different Point of View

From a sentiment analysis perspective, the present chapter aims to study and identify the best similarity metric able to describe the sentiment distribution of several types of books, establishing a different point of view on the interpretation of feeling extraction: the classification of documents based on an emotional distance (Fig. 2).

In order to implement this idea, an existing text miner application is employed: the SeaLab Advanced Information Retrieval—SLAIR—is a software, developed in

```
▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  ▼<rdf:Description rdf:about="http://sentic.net/api/en/concept/celebrate_special_occasion">
     <rdf:type rdf:resource="http://sentic.net/api/concept"/>
     <text xmlns="http://sentic.net/api">celebrate special occasion</text>
     <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_holiday"/>
     <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_occasion"/>
     <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_birthday"/>
     <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_wedding">
     <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/express_appreciation"/>
     <pleasantness xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.93</pleasantness>
     <attention xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.724</attention>
     <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</sensitivity>
     <aptitude xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</aptitude>
     <polarity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.551</polarity>
   </rdf:Description>
</rdf:RDF>
```


Fig. 2 A sample of SenticNet

C++ programming language, dedicated to the analysis of large amount of documents for clustering and classification purposes [51]. In this context the word '*document*' is used to denote any source of data which carries information, e.g., text written in natural language, web pages, images [35]. SLAIR normalizes input documents into an internal representation and applies several metrics to compute distances between pair of documents; then, the document-distance takes into account a conventional content-based similarity metric, a stylistic similarity criterion and a semantic representation of the documents, applying machine learning algorithms for both cluster and classification purposes. After a pre-processing phase, in which language identification, stemming and stopword removal steps are carried out, a text document becomes a '*docum object*', deprived of useless information (e.g., articles, prepositions, punctuation, special characters). At this level, different types of semantic descriptors can be applied to the new object, e.g., the SenticNet semantic descriptor explained in the present work.

4.1 SenticNet Semantic Descriptor

A document can hold sentiment information and one may be interested in analyzing its sentiment distribution, in order to study and compare different documents from a sentiment point of view. In particular, this chapter adopts the SenticNet framework, which allows one to retrieve four different sentiment experiences associated with a specific word, in order to develop a sentiment semantic descriptor made up of a vector of four affective dimensions, i.e., Pleasantness, Attention, Sensitivity and Aptitude. Hence, each document is described by a sentiment vector having four values in the range $[-1, +1]$.

The SenticNet network has been integrated into a MySQL relational database management system, called MariaDB, and SLAIR interacts with it through MySQL C++ APIs. After SLAIR '*text processing phase*', the SenticNet Semantic Descriptor extracts the list of words that compose the document and submits each single word to MariaDB through a proper query. If the word is present in SenticNet, the database returns the corresponding four dimensions, which are later saved into a vector. If the word is not included in the database, the sentiment descriptor assigns a default vector, that will not be considered during the calculation of the sentiment distribution. The entire SenticNet Semantic Descriptor Algorithm is described in Algorithm 1. After the semantic descriptor step, a new process starts, during which the distance between two document is calculated, using a specific metric.

Algorithm 1: SENTICNET SEMANTIC DESCRIPTOR

Input: A document $docum = \{(w)_i; i = 1, ..., N\}$,
where w is a *word* and N is the total number of words in *docum*.

Initialization: $sv[D]$, where sv is the sentiment vector that describes $docum$ and $D = 4$ is the number of affective dimensions [pleasantness, attention, sensitivity, aptitude]; $senticTemp[D]$ is a temporary vector used for calculation and wordCount = 0 is the number of valid SenticNet words.

for(i=0; i<N; i++): $SenticDB(w_i, senticTemp)$, where SenticDB is a function that makes a query to MariaDB and returns the filled senticTemp vector. If w_i is not present in MariaDB, the $senticTemp$ values are set to 2, which is a value out of the acceptable range [-1,+1], otherwise $wordCount++$.

    for(j=0; j<D; j++)
        if(senticTemp[j] != 2)
            $sv[j] = sv[j] + senticTemp[j]$

for(i=0; i<wordCount; i++)
    for(j=0; j<D; j++)
        $sv[j] = sv[j]/wordCount$

Output: $docum$ is described by four normalized affective dimensions, $Pleasantness$, $Attention$, $Sensitivity$, $Aptitude$.

5 Experimental Setup

SLAIR framework is provided with a module able to calculate the distance between two text documents, based on a specified metric. As shown in Algorithm 2, a *SenticNet Distance Metric Function* is implemented, and requires as arguments the two documents to be compared and the metric. First a Semantic SenticNet Descriptor is retrieved from each document. Once the two sentiment vectors *sv[D]* describing the two documents are obtained, a specific metric is applied to *sv[D]*. It is possible to choose one of the following different metrics:

- Manhattan norm $\|\mathbf{x}\|_1 :=$
$$\sum_{i=1}^{n} |x_i|$$
- Euclidean norm $\|\mathbf{x}\| :=$
$$\sqrt{x_1^2 + \cdots + x_n^2}$$

- Maximum norm $\|\mathbf{x}\|_\infty :=$

$$max(|x_1|, \ldots, |x_n|)$$

The *SenticNet Distance Metric Function* returns a distance which describes the closeness between the two documents; e.g., if the documents are very similar, the distance will be a positive number near zero, otherwise, if the documents are very different, the distance will be a number near one.

Algorithm 2: SENTICNET DISTANCE METRIC FUNCTION

Input: Two documents $documA$, $documB$, a $metric\ \Phi$ and a distance Δ.
1 **Get sv[D]:** extract $sv[D]_A$ vector from $documA$ and $sv[D]_B$ vector from $documB$.
2 **Calculate SenticNet Distance**: Δ = SenticDistMetric($sv[D]_A$, $sv[D]_B$, Φ) where $SenticDistMetric$ is a function that calculates the distance between $sv[D]_A$ and $sv[D]_B$ based on Φ.
3 **Return** Δ

6 Experimental Results

This section reports on the experimental setup adopted to demonstrate the validity of the approach employing the *SenticNet Semantic Descriptor* and the *Sentiment Distance Metric*, implemented in SLAIR.

In order to take advantage of the use of the four affective dimensions, the choice of the documents was guided by the necessity of having texts long enough to convey clear feelings and emotions: therefore, the experiments of the present work have been performed on a dataset of books, written or translated in English language. In fact, on one hand, books are full of sentimental and affective contents and, on the other hand, the analysis of sentiment orientation applied to this type of dataset can inspire a new method of classification. For example, from a commercial perspective, the analysis of the sentiment distribution of a book may be useful within a search engine which suggests the book most emotionally similar to another one.

The aim of the present work is to classify different literary genres; to this purpose, we selected the following five distinct categories:

- XIX Century Novel
- Fantasy
- Horror
- XXI Century Novel
- Greek Tragedy

Table 3 Books dataset description

ID	Category	Author	Title
1	XIX century novel	F. Dostoevskij	Crime and Punishment
2	XIX century novel	J. Austen	Emma
3	XIX century novel	L.M. Alcot	Little Women
4	XIX century novel	L. Tolstoj	War and Peace
5	Fantasy	G.R.R. Martin	A Game of Thrones
6	Fantasy	M. Weis, T. Hickman	Dragonlance Tales
7	Fantasy	T. Brooks	The Sword of Shannara
8	Fantasy	David Gemmell	Winter Warriors
9	Horror fiction	H.P. Lovecraft	The Beast in the Cave
10	Horror fiction	D. Sharen	Blood Beast
11	Horror fiction	Robin Becker	Brains: a Zombie Memoir
12	Horror fiction	E. A. Poe	The Masque Of The Red Death
13	XXI century novel	J. Egan	A Visit from the Goon Squad
14	XXI century novel	E. Gilbert	Eat Pray Love
15	XXI century novel	J. Martel	Life of Pi
16	XXI century novel	D. Nicholls	One Day
17	Greek tragedy	Sophocles	Antogone
18	Greek tragedy	Euripides	Hecuba
19	Greek tragedy	Sophocles	Oedipus the King
20	Greek tragedy	Euripides	The Suppliants

The dataset used is shown in Table 3: it is made up of 20 extracts of books, 20.000 characters each, selected from a random line of each book. Therefore, each category is represented by four extracts of books.

Three different distance metrics are applied to the extracts, namely the Manhattan, the Euclidean and the Maximum norm. The experimental results underline some differences between the distances resulted from the use of the three norms: Tables 4 and 5 respectively show the most similar and distant categories based on the minimum and maximum distances of each book with respect to the others. For example, it is possible to notice a similarity between the XIX century novels and the XXI century novels according to all the norms, while both genres result distant from the fantasy novels; this evidence is completely coherent from an affective point of view, because even though XIX and XXI century novels may be set in different environments, mentality and social constraints, they can still convey similar types of feelings.

Table 4 Most similar categories with different norms

ID	Manhattan norm	Euclidean norm	Maximum norm
1	XXI century novel	XXI century novel	XXI century novel
2	XXI century novel	XXI century novel	XXI century novel
3	XXI century novel	XIX century novel	XXI century novel
4	XXI century novel	XXI century novel	XXI century novel
5	Fantasy	Fantasy	Fantasy
6	Horror	Horror	Horror
7	Horror	Horror	Horror
8	Fantasy	Fantasy	Fantasy
9	Horror	Horror	Horror
10	Horror	Fantasy	Horror
11	Fantasy	Fantasy	Horror
12	Horror	Horror	Horror
13	XIX century novel	XIX century novel	XIX century novel
14	XXI century novel	XXI century novel	XXI century novel
15	XIX century novel	XIX century novel	XIX century novel
16	XIX century novel	XIX century novel	XIX century novel
17	Tragedy	Tragedy	Tragedy
18	Horror	Horror	Horror
19	XXI century novel	XXI century novel	XXI century novel
20	Tragedy	Tragedy	Tragedy

We chose the Euclidean norm to perform a deeper analysis, since it better underlines the similarities and dissimilarities between different book categories distributions. Table 6 reports on the results of the distances between the 20 extracts, employing the *Euclidean norm*: in particular, the table represents a symmetric matrix, in which each row indicates an extract of a book compared with the other extracts. The zero values on the diagonal hence indicate the difference between an extract and itself. For each book in the rows, the corresponding black cell represents the minimum distance, while the grey one represents the maximum distance. It is therefore straightforward to use this information to extract not only the affective affinity or difference between books, but also the possible influence of a previous book or author in the following ones. For example, it can be noticed that the XIX century novel extracts are all emotionally close to the extracts of the XXI century and different from the Fantasy genre. Also, Fantasy and Horror extracts are both extremely distant from the XXI century novel.

In order to better clarify the book extracts sentiment distributions, the following pictures are shown: each figure reports on the emotional trend of each category, with four plots for each book of the category. The figures prove the effectiveness of

Table 5 Most distant categories with different norms

ID	Manhattan norm	Euclidean norm	Maximum norm
1	XIX century novel	XIX century novel	XIX century novel
2	Fantasy	Fantasy	Fantasy
3	Fantasy	Fantasy	Fantasy
4	Fantasy	Fantasy	Horror
5	Fantasy	Fantasy	XIX century novel
6	XXI century novel	XIX century novel	XXI century novel
7	XXI century novel	XIX century novel	XXI century novel
8	XIX century novel	Fantasy	XXI century novel
9	XXI century novel	XIX century novel	XXI century novel
10	XXI century novel	XIX century novel	XXI century novel
11	XXI century novel	XIX century novel	XXI century novel
12	XXI century novel	XIX century novel	XXI century novel
13	XXI century novel	XIX century novel	XXI century novel
14	Fantasy	Fantasy	Horror
15	Fantasy	Fantasy	Horror
16	Fantasy	Fantasy	Fantasy
17	XXI century novel	Fantasy	XXI century novel
18	XXI century novel	XIX century novel	XXI century novel
19	Fantasy	Fantasy	Horror
20	Fantasy	Fantasy	Horror

the proposed approach: each different category is in fact represented by a particular distribution, shared by the books of the same category; moreover, it is possible to use local minima in order to find similarity points between a certain category and specific books. Another important result is achieved when noticing that different books from different categories may show a similar distribution; for example, Crime and Punishment by Dostoevskij shows a similar emotional distribution compared to two fantasy novels, namely The Game of Thrones and Winter Warriors: this means that, even if the genre is different, these books arouse similar feelings.

Eventually, Table 7 with the sentiment distances of the example sentences named in the Introduction section is reported, in order to show the similarity between the Dostoevskij's book and the Austen's one, differently from Martin's fantasy novel.

Table 6 Books sentiment distribution

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	3.373	0.865	1.255	0.26	0.839	2.698	0.323	1.195	2.022	1.422	1.504	0.19	2.823	1.669	0.644	1.198	2.229	1.303	1.259
2	3.373	0	0.878	0.601	3.073	3.691	7.954	2.947	4.654	5.742	5.667	7.35	2.359	0.561	0.546	1.198	2.405	5.172	1.279	1.928
3	0.865	0.878	0	0.156	0.732	1.155	4.163	0.684	1.825	2.678	2.354	3.302	0.381	1.081	0.406	0.092	0.808	2.547	0.503	0.621
4	1.255	0.601	0.156	0	1.227	1.924	5.482	1.091	2.697	3.921	3.397	4.505	0.835	0.458	0.09	0.136	1.108	3.301	0.252	0.786
5	0.26	3.073	0.732	1.227	0	0.59	3.261	0.126	1.378	2.121	1.233	1.721	0.308	2.83	1.631	0.589	1.237	2.389	1.206	1.137
6	0.839	3.691	1.155	1.924	0.59	0	1.598	0.354	0.295	0.75	0.222	0.835	0.554	4.228	2.742	1.122	0.535	0.724	1.78	0.703
7	2.698	7.954	4.163	5.482	3.261	1.598	0	2.943	0.617	0.331	0.976	0.51	2.396	8.958	6.819	4.273	2.477	1.107	5.659	3.354
8	0.323	2.947	0.684	1.091	0.126	0.354	2.943	0	1.008	1.925	0.941	1.596	0.329	2.746	1.595	0.46	0.665	1.631	0.849	0.591
9	1.195	4.654	1.825	2.697	1.378	0.295	0.617	1.008	0	0.281	0.255	0.499	0.869	5.339	3.709	1.824	0.68	0.313	2.655	1.128
10	2.022	5.742	2.678	3.921	2.121	0.75	0.331	1.925	0.281	0	0.499	0.536	1.48	7.004	5.076	2.933	1.621	0.845	4.227	2.241
11	1.422	5.667	2.354	3.397	1.233	0.222	0.976	0.941	0.255	0.499	0	0.34	1.245	6.314	4.462	2.275	1.14	0.623	3.108	1.509
12	1.504	7.35	3.302	4.505	1.721	0.835	0.51	1.596	0.499	0.536	0.34	0	1.588	7.714	5.627	3.189	2.097	1.238	4.375	2.69
13	0.19	2.359	0.381	0.835	0.308	0.554	2.396	0.329	0.869	1.48	1.245	1.588	0	2.396	1.294	0.383	0.757	1.781	1.11	0.83
14	2.823	0.561	1.081	0.458	2.83	4.228	8.958	2.746	5.339	7.004	6.314	7.714	2.396	0	0.176	1.021	2.804	6.02	0.815	2.198
15	1.669	0.546	0.406	0.09	1.631	2.742	6.819	1.595	3.709	5.076	4.462	5.627	1.294	0.176	0	0.373	1.813	4.466	0.418	1.364
16	0.644	1.198	0.092	0.136	0.589	1.122	4.273	0.46	1.824	2.933	2.275	3.189	0.383	1.021	0.373	0	0.7	2.448	0.227	0.479
17	1.198	2.405	0.808	1.108	1.237	0.535	2.477	0.665	0.68	1.621	1.14	2.097	0.757	2.804	1.813	0.7	0	0.621	0.882	0.079
18	2.229	5.172	2.547	3.301	2.389	0.724	1.107	1.631	0.313	0.845	0.623	1.238	1.781	6.02	4.466	2.448	0.621	0	2.908	1.072
19	1.303	1.279	0.503	0.252	1.206	1.78	5.659	0.849	2.655	4.227	3.108	4.375	1.11	0.815	0.418	0.227	0.882	2.908	0	0.502
20	1.259	1.928	0.621	0.786	1.137	0.703	3.354	0.591	1.128	2.241	1.509	2.69	0.83	2.198	1.364	0.479	0.079	1.072	0.502	0

Table 7 Sentiment distance metric example applied to sentences

Sentence ID	A	B	C
A	0	0.038827	0.11794
B	0.038827	0	0.035272
C	0.11794	0.035272	0

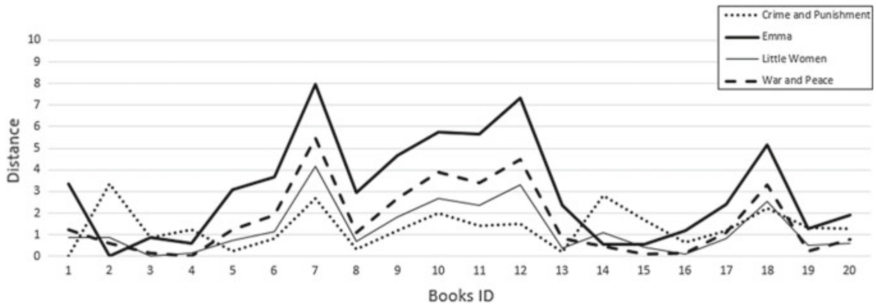

Fig. 3 XIX century novel books

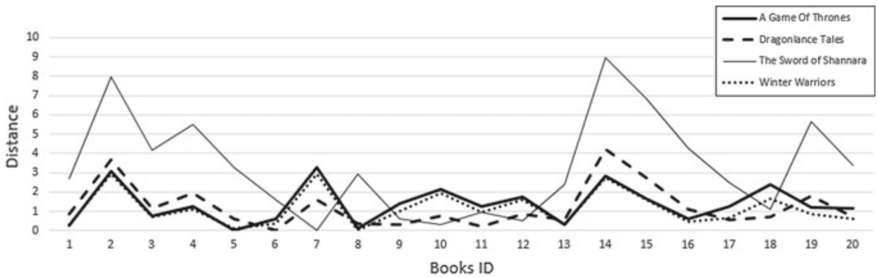

Fig. 4 Fantasy books

7 Conclusion

The present research has exploited a cognitive model for emotion recognition in natural language text. In particular, the aim of the proposed approach has involved the analysis of the sentimental and affective contents of some literary genres of books, in order to extract and study the distribution of each category and implement a method able to automatically detect similarities and dissimilarities between them.

In order to pursue this scope, SenticNet 3 has been used as a publicly available semantic and affective resource to obtain the values of Pleasantness, Attention, Sensitivity, and Aptitude of each book, which can potentially synthesize the full range of emotional experiences. The software application SLAIR has been employed to analyze the sentiment distribution of the books, in order to study and compare different documents from a sentiment point of view. In particular, a sentiment semantic

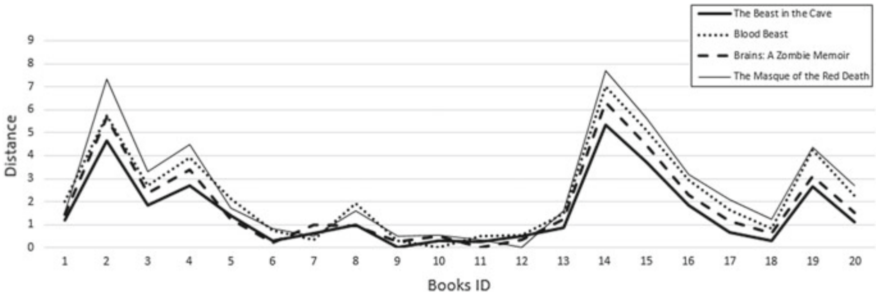

Fig. 5 Horror books

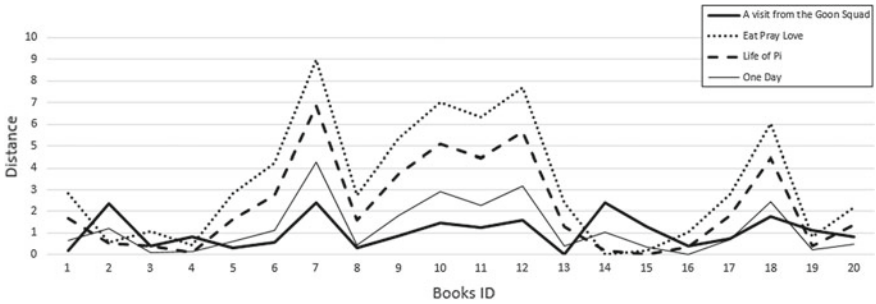

Fig. 6 XXI century novel books

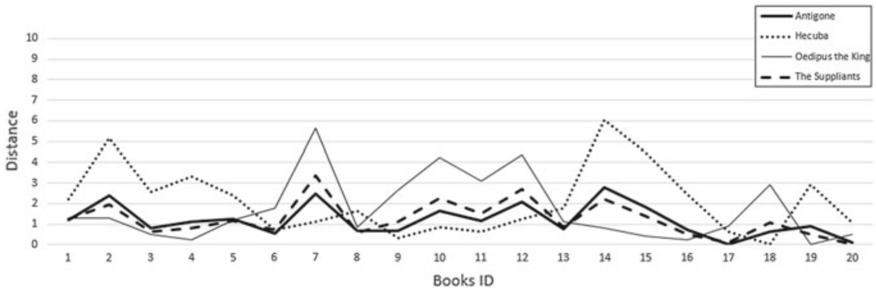

Fig. 7 Greek tragedy books

descriptor has been made up using a vector representing the four affective dimensions. Besides, different metrics have been proposed to calculate the distance between two documents (Figs. 3 and 4).

Experimental results have shown that the proposed approach is able to extract not only the affective affinity or difference between books, but also the possible influence of a previous book in the following ones. Moreover, we found out that each category is characterized by a certain distribution, and that local minima of this distribution can be exploited in order to find similarities with specific books (Figs. 5, 6 and 7).

References

1. Allen, J.: Natural Language Understanding. Benjamin/Cummings (1987)
2. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The berkeley framenet project. In: Proceedings of the 17th international conference on Computational linguistics, vol, 1, pp. 86–90. Association for Computational Linguistics (1998)
3. Bisio, F., Gastaldo, P., Peretti, C., Zunino, R., Cambria, E.: Data intensive review mining for sentiment classification across heterogeneous domains. In: Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on. pp. 1061–1067. IEEE (2013)
4. Bisio, F., Gastaldo, P., Zunino, R., Cambria, E.: A learning scheme based on similarity functions for affective common-sense reasoning. In: IJCNN. pp. 2476–2481 (2015)
5. Bizer, C., Jens, L., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia—a crystallization point for the web of data. Web Semant.: Sci. Serv. Agents World Wide Web **7**(3), 154–165 (2009)
6. Bosco, C., Patti, V., Bolioli, A.: Developing corpora for sentiment analysis and opinion mining: a survey and the Senti-TUT case study. IEEE Intell. Syst. **28**(2), 55–63 (2013)
7. Cambria, E., Fu, J., Bisio, F., Poria, S.: AffectiveSpace 2: enabling affective intuition for concept-level sentiment analysis. In: AAAI. pp. 508–514. Austin (2015)
8. Cambria, E., Gastaldo, P., Bisio, F., Zunino, R.: An ELM-based model for affective analogical reasoning. Neurocomputing **149**, 443–455 (2015)
9. Cambria, E.: Affective computing and sentiment analysis. IEEE Intelligent Systems **31**(2), (2016)
10. Cambria, E., Hussain, A.: Sentic computing: a common-sense-based framework for concept-level sentiment analysis. Springer, Cham, Switzerland (2015)
11. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: SenticSpace: visualizing opinions and sentiments in a multi-dimensional vector space. In: Setchi, R., Jordanov, I., Howlett, R., Jain, L. (eds.) Knowledge-Based and Intelligent Information and Engineering Systems. Lecture Notes in Artificial Intelligence, vol. 6279, pp. 385–393. Springer, Berlin (2010)
12. Cambria, E., Livingstone, A., Hussain, A.: The hourglass of emotions. In: Esposito, A., Vinciarelli, A., Hoffmann, R., Muller, V. (eds.) Cognitive Behavioral Systems. Lecture Notes in Computer Science, vol. 7403, pp. 144–157. Springer, Berlin Heidelberg (2012)
13. Cambria, E., Olsher, D., Kwok, K.: Sentic activation: a two-level affective common sense reasoning framework. In: AAAI. pp. 186–192. Toronto (2012)
14. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: AAAI. pp. 1515–1521. Quebec City (2014)
15. Cambria, E., Poria, S., Bisio, F., Bajpai, R., Chaturvedi, I.: The clsa model: a novel framework for concept-level sentiment analysis. In: Computational Linguistics and Intelligent Text Processing, pp. 3–22. Springer (2015)
16. Cambria, E., Rajagopal, D., Kwok, K., Sepulveda, J.: GECKA: game engine for commonsense knowledge acquisition. In: FLAIRS, pp. 282–287 (2015)
17. Cambria, E., Schuller, B., Liu, B., Wang, H., Havasi, C.: Knowledge-based approaches to concept-level sentiment analysis. IEEE Intell. Syst. **28**(2), 12–14 (2013)
18. Cambria, E., Schuller, B., Liu, B., Wang, H., Havasi, C.: Statistical approaches to concept-level sentiment analysis. IEEE Intell. Syst. **28**(3), 6–9 (2013)
19. Cambria, E., Xia, Y., Hussain, A.: Affective common sense knowledge acquisition for sentiment analysis. In: LREC, pp. 3580–3585. Istanbul (2012)
20. 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)
21. Chinthala, S., Mande, R., Manne, S., Vemuri, S.: Sentiment analysis on twitter streaming data. In: Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI), vol. 1, pp. 161–168. Springer (2015)

22. van Cranenburgh, A., Huygens, I., Koolen, C.: Identifying literary texts with bigrams. In: Computational Linguistics for Literature, p. 58 (2015)
23. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 241–249. Association for Computational Linguistics (2010)
24. Di Fabbrizio, G., Aker, A., Gaizauskas, R.: Summarizing on-line product and service reviews using aspect rating distributions and language modeling. IEEE Intell. Syst. **28**(3), 28–37 (2013)
25. Dyer, M.G.: Connectionist natural language processing: a status report. In: Computational Architectures Integrating Neural and Symbolic Processes, pp. 389–429. Springer (1995)
26. Elliott, C.D.: The affective reasoner: a process model of emotions in a multi-agent system. Northwestern University (1992)
27. García-Moya, L., Anaya-Sanchez, H., Berlanga-Llavori, R.: A language model approach for retrieving product features and opinions from customer reviews. IEEE Intell. Syst. **28**(3), 19–27 (2013)
28. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, pp. 174–181. Association for Computational Linguistics (1997)
29. Honkela, T., Korhonen, J., Lagus, K., Saarinen, E.: Five-dimensional sentiment analysis of corpora, documents and words. In: Advances in Self-Organizing Maps and Learning Vector Quantization, pp. 209–218. Springer (2014)
30. Hung, C., Lin, H.K.: Using objective words in SentiWordNet to improve sentiment classification for word of mouth. IEEE Intell. Syst. **28**(2), 47–54 (2013)
31. Kamps, J., Marx, M., Mokken, R.J., De Rijke, M.: Using wordnet to measure semantic orientations of adjectives. In: LREC. vol. 4, pp. 1115–1118. Citeseer (2004)
32. Kim, S.M., Hovy, E.: Automatic detection of opinion bearing words and sentences. In: Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), pp. 61–66 (2005)
33. Kim, S.M., Hovy, E.: Extracting opinions, opinion holders, and topics expressed in online news media text. In: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pp. 1–8. Association for Computational Linguistics (2006)
34. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: the good the bad and the omg!. Icwsm **11**, 538–541 (2011)
35. Meda, C., Bisio, F., Gastaldo, P., Zunino, R., Surlinelli, R., Scillia, E., Ottaviano, A.V.: Content-adaptive analysis and filtering of microblogs traffic for event-monitoring applications. In: Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, vol. 1, pp. 155–170. Springer (2015)
36. Melville, P., Gryc, W., Lawrence, R.D.: Sentiment analysis of blogs by combining lexical knowledge with text classification. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1275–1284. ACM (2009)
37. Mohammad, S.M., Kiritchenko, S., Zhu, X.: Nrc-canada: building the state-of-the-art in sentiment analysis of tweets. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), vol. 2, pp. 321–327 (2013)
38. Murtagh, F., Ganz, A.: Pattern recognition in narrative: analysis of narratives of emotion (2014). arXiv preprint arXiv:1405.3539
39. Ortigosa, A., Martín, J.M., Carro, R.M.: Sentiment analysis in facebook and its application to e-learning. Comput. Hum. Behav. **31**, 527–541 (2014)
40. Ortony, A., Clore, G., Collins, A.: Cogn. Struct. Emotions. Cambridge University Press, Cambridge (1988)
41. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p. 271. Association for Computational Linguistics (2004)
42. Pang, B., Lee, L.: Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 115–124. Association for Computational Linguistics (2005)

43. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
44. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: EMNLP, pp. 79–86. Philadelphia (2002)
45. Plutchik, R.: The nature of emotions. Am. Sci. **89**(4), 344–350 (2001)
46. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Natural Language Processing and Text Mining, pp. 9–28. Springer (2007)
47. Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., Bandyopadhyay, S.: Enhanced SenticNet with affective labels for concept-based opinion mining. Intell. Sys. IEEE **28**(2), 31–38 (2013)
48. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: EACL, pp. 675–682. Athens (2009)
49. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pp. 105–112. Association for Computational Linguistics (2003)
50. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S.M., Ritter, A., Stoyanov, V.: Semeval-2015 task 10: Sentiment analysis in twitter. In: Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval (2015)
51. Sangiacomo, F., Leoncini, A., Decherchi, S., Gastaldo, P., Zunino, R.: Sealab advanced information retrieval. In: IEEE Fourth International Conference on Semantic Computing (ICSC), pp. 444–445. IEEE (2010)
52. dos Santos, C.N., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland (2014)
53. Snyder, B., Barzilay, R.: Multiple aspect ranking using the good grief algorithm. In: HLT-NAACL, pp. 300–307 (2007)
54. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), vol. 1631, p. 1642. Citeseer (2013)
55. Somasundaran, S., Wiebe, J., Ruppenhofer, J.: Discourse level opinion interpretation. In: COLING, pp. 801–808. Manchester (2008)
56. Speer, R., Havasi, C.: ConceptNet 5: a large semantic network for relational knowledge. In: Hovy, E., Johnson, M., Hirst, G. (eds.) Theory and Applications of Natural Language Processing, chap. 6. Springer (2012)
57. Stevenson, R., Mikels, J., James, T.: Characterization of the affective norms for english words by discrete emotional categories. Behav. Res. Methods **39**, 1020–1024 (2007)
58. Strapparava, C., Valitutti, A.: WordNet-Affect: An affective extension of WordNet. In: LREC, pp. 1083–1086. Lisbon (2004)
59. Tang, D., Wei, F., Qin, B., Liu, T., Zhou, M.: Coooolll: a deep learning system for twitter sentiment classification. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 208–212 (2014)
60. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 1555–1565 (2014)
61. Tsai, A., Tsai, R., Hsu, J.: Building a concept-level sentiment dictionary based on commonsense knowledge. IEEE Intell. Syst. **28**(2), 22–30 (2013)
62. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational linguistics, pp. 417–424. Association for Computational Linguistics (2002)
63. Vahdat, M., Oneto, L., Anguita, D., Funk, M., Rauterberg, M.: Can machine learning explain human learning? Neurocomputing (In Press)
64. Wang, Q., Cambria, E., Liu, C., Hussain, A.: Common sense knowledge for handwritten chinese recognition. Cogn. Comput. **5**(2), 234–242 (2013)

65. Weichselbraun, A., Gindl, S., Scharl, A.: Extracting and grounding context-aware sentiment lexicons. IEEE Intell. Syst. **28**(2), 39–46 (2013)
66. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Lang. Resour. Eval. **39**(2), 165–210 (2005)
67. Wu, W., Li, H., Wang, H., Zhu, K.: Probase: a probabilistic taxonomy for text understanding. In: SIGMOD, pp. 481–492. Scottsdale (2012)
68. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pp. 129–136. Association for Computational Linguistics (2003)
69. Zirn, C., Niepert, M., Stuckenschmidt, H., Strube, M.: Fine-grained sentiment analysis with structural features. In: IJCNLP, pp. 336–344 (2011)