Sentic API

A Common-Sense Based API for Concept-Level Sentiment Analysis

http://sentic.net/api

Erik Cambria
Temasek Laboratories
National University of Singapore
cambria@nus.edu.sg

Soujanya Poria
School of Electrical & Electronic Engineering
Nanyang Technological University
sporia@ntu.edu.sg

Alexander Gelbukh
Center for Computing Research
National Polytechnic Institute of Mexico
gelbukh@cic.ipn.mx

Kenneth Kwok
Temasek Laboratories
National University of Singapore
kenkwok@nus.edu.sg

ABSTRACT

The bag-of-concepts model can represent semantics associated with natural language text much better than bags-of-words. In the bag-of-words model, in fact, a concept such as `cloud_computing` would be split into two separate words, disrupting the semantics of the input sentence. Working at concept-level is important for tasks such as opinion mining, especially in the case of microblogging analysis. In this work, we present Sentic API, a common-sense based application programming interface for concept-level sentiment analysis, which provides semantics and sentics (that is, denotative and connotative information) associated with 15,000 natural language concepts.

Categories and Subject Descriptors

H.3.1 [**Information Systems Applications**]: Linguistic Processing; I.2.7 [**Natural Language Processing**]: Language parsing and understanding

General Terms

Algorithms

Keywords

Natural language processing; Sentiment analysis

1. INTRODUCTION

Hitherto, online information retrieval, aggregation, and processing have mainly been based on algorithms relying on the textual representation of webpages. Such algorithms are very good at retrieving texts, splitting them into parts, checking the spelling and counting the number of words.

Published as part of the #Microposts2014 Workshop proceedings, available online as CEUR Vol-1141 (http://ceur-ws.org/Vol-1141)

#Microposts2014, April 7th, 2014, Seoul, Korea.

But when it comes to interpreting sentences and extracting meaningful information, their capabilities are known to be very limited. Machine-learning algorithms, in fact, are limited by the fact that they can process only the information that they can 'see'. As human text processors, we do not have such limitations as every word we see activates a cascade of semantically related concepts, relevant episodes, and sensory experiences, all of which enable the completion of complex tasks – such as word-sense disambiguation, textual entailment, and semantic role labeling – in a quick and effortless way. Machine learning techniques, moreover, are intrinsically meant for chunking numerical data. Through escamotages such as word frequency counting, it is indeed possible to apply such techniques also in the context of natural language processing (NLP), but it would be no different from trying to understand an image by solely looking at bits per pixel information.

Concept-level sentiment analysis, instead, focuses on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation 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 count, but rather rely on the implicit features associated with natural language concepts. Unlike purely syntactical techniques, concept-based approaches are able to detect also sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but which are implicitly linked to other concepts that do so. The bag-of-concepts model can represent semantics associated with natural language much better than bags-of-words. In the bag-of-words model, in fact, a concept such as `cloud_computing` would be split into two separate words, disrupting the semantics of the input sentence (in which, for example, the word `cloud` could wrongly activate concepts related to `weather`).

By allowing for the inference of semantics and sentics, the analysis at concept-level enables a comparative fine-grained feature-based sentiment analysis. Rather than gathering isolated opinions about a whole item (e.g., iPhone 5S or Galaxy S5), users are generally more interested in comparing different products according to their specific features (e.g., iPhone 5S's vs Galaxy S5's touchscreen), or sub-features (e.g., fragility of iPhone 5S's vs Galaxy S5's touchscreen). In this context, the construction of comprehensive common and common-sense knowledge bases is key for feature-spotting and polarity detection, respectively.

Common-sense, in particular, is necessary to properly deconstruct natural language text into sentiments– for example, to appraise the concept `small_room` as negative for a hotel review and `small_queue` as positive for a post office, or the concept `go_read_the_book` as positive for a book review but negative for a movie review.

The rest of the paper is organized as follows: Section 2 presents available resources for concept-level sentiment analysis; Section 3 illustrates the techniques exploited to build the Sentic API; Section 4 describes in detail how the API is developed and how it can be used; Section 5 proposes an evaluation of the API; finally, Section 6 concludes the paper and suggests further research directions.

2. RELATED WORK

Commonly used resources for concept-level sentiment analysis include ANEW [3], WordNet-Affect (WNA) [21], ISEAR [1], SentiWordNet [9], and SenticNet [7]. In [22], for example, a concept-level sentiment dictionary is built through a two-step method combining iterative regression and random walk with in-link normalization. ANEW and SenticNet are exploited for propagating sentiment values based on the assumption that semantically related concepts share common sentiment. Moreover, polarity accuracy, Kendall distance, and average-maximum ratio are used, in stead of mean error, to better evaluate sentiment dictionaries.

A similar approach is adopted in [19], which presents a methodology for enriching SenticNet concepts with affective information by assigning an emotion label to them. Authors use various features extracted from ISEAR, as well as similarity measures that rely on the polarity data provided in SenticNet (those based on WNA) and ISEAR distance-based measures, including point-wise mutual information, and emotional affinity. Another recent work that builds upon an existing affective knowledge base is [14], which proposes the re-evaluation of objective words in SentiWordNet by assessing the sentimental relevance of such words and their associated sentiment sentences. Two sampling strategies are proposed and integrated with support vector machines for sentiment classification. According to the experiments, the proposed approach significantly outperforms the traditional sentiment mining approach, which ignores the importance of objective words in SentiWordNet. In [2], the main issues related to the development of a corpus for opinion mining and sentiment analysis are discussed both by surveying the existing work in this area and presenting, as a case study, an ongoing project for Italian, called Senti-TUT, where a corpus for the investigation of irony about politics in social media is developed.

Other work explores the ensemble application of knowledge bases and statistical methods. In [24], for example, a hybrid approach to combine lexical analysis and machine learning is proposed in order to cope with ambiguity and integrate the context of sentiment terms. The context-aware method identifies ambiguous terms that vary in polarity depending on the context and stores them in contextualized sentiment lexicons. In conjunction with semantic knowledge bases, these lexicons help ground ambiguous sentiment terms to concepts that correspond to their polarity.

More machine-learning based works include [10], which introduces a new methodology for the retrieval of product features and opinions from a collection of free-text customer reviews about a product or service. Such a methodology relies on a language modeling framework that can be applied to reviews in any domain and language provided with a seed set of opinion words. The methodology combines both a kernel-based model of opinion words (learned from the seed set of opinion words) and a statistical mapping between words to approximate a model of product features from which the retrieval is carried out.

3. TECHNIQUES ADOPTED

In this work, we exploit the ensemble application of spectral association [12], an approximation of many steps of spreading activation, and CF-IOF (concept frequency - inverse opinion frequency), an approach similar to TF-IDF weighting, to extract semantics from ConceptNet [20], a semantic network of common-sense knowledge. The extraction of sentics, in turn, is performed through the combined use of AffectiveSpace [4], a multi-dimensional vector space representation of affective common-sense knowledge, and the Hourglass of Emotions [6], a brain-inspired emotion categorization model.

3.1 Spectral Association

Spectral association is a technique that involves assigning activations to 'seed concepts' and applying an operation that spreads their values across the graph structure of ConceptNet. This operation transfers the most activation to concepts that are connected to the key concepts by short paths or many different paths in common-sense knowledge.

In particular, we build a matrix C that relates concepts to other concepts, instead of their features, and add up the scores over all relations that relate one concept to another, disregarding direction. Applying C to a vector containing a single concept spreads that concept's value to its connected concepts. Applying C^2 spreads that value to concepts connected by two links (including back to the concept itself). As we aim to spread the activation through any number of links, with diminishing returns, the operator we want is:

$$1 + C + \frac{C^2}{2!} + \frac{C^3}{3!} + ... = e^C$$

We can calculate this odd operator, e^C, because we can factor C. C is already symmetric, so instead of applying Lanczos' method [15] to CC^T and getting the singular value decomposition (SVD), we can apply it directly to C and get the spectral decomposition $C = V\Lambda V^T$. As before, we can raise this expression to any power and cancel everything but the power of Λ. Therefore, $e^C = Ve^{\Lambda}V^T$. This simple twist on the SVD lets us calculate spreading activation over the whole matrix instantly. We can truncate this matrix to k axes and therefore save space while generalizing from similar concepts. We can also rescale the matrix, so that activation values have a maximum of 1 and do not tend to collect in highly-connected concepts, by normalizing the truncated rows of $Ve^{\Lambda/2}$ to unit vectors, and multiplying that matrix by its transpose to get a rescaled version of $Ve^{\Lambda}V^T$.

3.2 CF-IOF Weighting

CF-IOF is a technique that identifies common topic-dependent semantics in order to evaluate how important a concept is to a set of opinions concerning the same topic. It is hereby used to feed spectral association with 'seed concepts'. Firstly, the frequency of a concept c for a given domain d is calculated by counting the occurrences of the concept c in the set of available d-tagged opinions and dividing the result by the sum of number of occurrences of all concepts in the set of opinions concerning d. This frequency is then multiplied by the logarithm of the inverse frequency of the concept in the whole collection of opinions, that is:

$$CF\text{-}IOF_{c,d} = \frac{n_{c,d}}{\sum_k n_{k,d}} \log \sum_k \frac{n_k}{n_c}$$

where $n_{c,d}$ is the number of occurrences of concept c in the set of opinions tagged as d, n_k is the total number of concept occurrences and n_c is the number of occurrences of c in the whole set of opinions.

A high weight in CF-IOF is reached by a high concept frequency (in the given opinions) and a low opinion frequency of the concept in the whole collection of opinions. Therefore, thanks to CF-IOF weights, it is possible to filter out common concepts and detect relevant topic-dependent semantics.

3.3 AffectiveSpace

To extract sentics from natural language text, we use AffectiveSpace, a multi-dimensional vector space built upon ConceptNet and WNA. The alignment operation operated over these two knowledge bases yields a matrix, A, in which common-sense and affective knowledge coexist, i.e., a matrix 15,000 $\times$ 118,000 whose rows are concepts (e.g., `dog` or `bake_cake`), whose columns are either common-sense and affective features (e.g., *isA-pet* or *hasEmotion-joy*), and whose values indicate truth values of assertions.

Therefore, in A, each concept is represented by a vector in the space of possible features whose values are positive for features that produce an assertion of positive valence (e.g., 'a penguin is a bird'), negative for features that produce an assertion of negative valence (e.g., 'a penguin cannot fly') and zero when nothing is known about the assertion. The degree of similarity between two concepts, then, is the dot product between their rows in A. The value of such a dot product increases whenever two concepts are described with the same feature and decreases when they are described by features that are negations of each other. In particular, we use truncated SVD [23] in order to obtain a new matrix containing both hierarchical affective knowledge and common-sense.

The resulting matrix has the form $\tilde{A} = U_k \, \Sigma_k \, V_k^T$ and is a low-rank approximation of A, the original data. This approximation is based on minimizing the Frobenius norm [13] of the difference between A and $\tilde{A}$ under the constraint $rank(\tilde{A}) = k$. For the Eckart–Young theorem [8] it represents the best approximation of A in the least-square sense, in fact:

$$\min_{\tilde{A}|rank(\tilde{A})=k} |A - \tilde{A}| = \min_{\tilde{A}|rank(\tilde{A})=k} |\Sigma - U^* \tilde{A} V| = \min_{\tilde{A}|rank(\tilde{A})=k} |\Sigma - S|$$

assuming that $\tilde{A}$ has the form $\tilde{A} = USV^*$, where S is diagonal. From the rank constraint, i.e., S has k non-zero diagonal entries, the minimum of the above statement is obtained as follows:

$$\min_{\tilde{A}|rank(\tilde{A})=k} |\Sigma - S| = \min_{s_i} \sqrt{\sum_{i=1}^{n} (\sigma_i - s_i)^2} =$$

$$= \min_{s_i} \sqrt{\sum_{i=1}^{k} (\sigma_i - s_i)^2 + \sum_{i=k+1}^{n} \sigma_i^2} = \sqrt{\sum_{i=k+1}^{n} \sigma_i^2}$$

Therefore, $\tilde{A}$ of rank k is the best approximation of A in the Frobenius norm sense when $\sigma_i = s_i$ $(i = 1, ..., k)$ and the corresponding singular vectors are the same as those of A. If we choose to discard all but the first k principal components, common-sense concepts and emotions are represented by vectors of k coordinates: these coordinates can be seen as describing concepts in terms of 'eigenmoods' that form the axes of AffectiveSpace, i.e., the basis e_0,...,e_{k-1} of the vector space. For example, the most significant eigenmood, e_0, represents concepts with positive affective valence. That is, the larger a concept's component in the e_0 direction is, the more affectively positive it is likely to be. Thus, by exploiting the information sharing property of truncated SVD, concepts with the same affective valence are likely to have similar features – that is, concepts conveying the same emotion tend to fall near each other in AffectiveSpace.

Concept similarity does not depend on their absolute positions in the vector space, but rather on the angle they make with the origin. For example we can find concepts such as `beautiful_day`, `birthday_party`, `laugh` and `make_person_happy` very close in direction in the vector space, while concepts like `sick`, `feel_guilty`, `be_laid_off` and `shed_tear` are found in a completely different direction (nearly opposite with respect to the centre of the space).

3.4 The Hourglass of Emotions

To reason on the disposition of concepts in AffectiveSpace, we use the Hourglass of Emotions (Figure 1), an affective categorization model developed starting from Plutchik's studies on human emotions [18]. In the model, sentiments are reorganized around four independent dimensions whose different levels of activation make up the total emotional state of the mind. The Hourglass of Emotions, in fact, is based on the idea that the mind is made of different independent resources and that emotional states result from turning some set of these resources on and turning another set of them off [16].

The primary quantity we can measure about an emotion we feel is its strength. But when we feel a strong emotion it is because we feel a very specific emotion. And, conversely, we cannot feel a specific emotion like 'fear' or 'amazement' without that emotion being reasonably strong. Mapping this space of possible emotions leads to an hourglass shape.

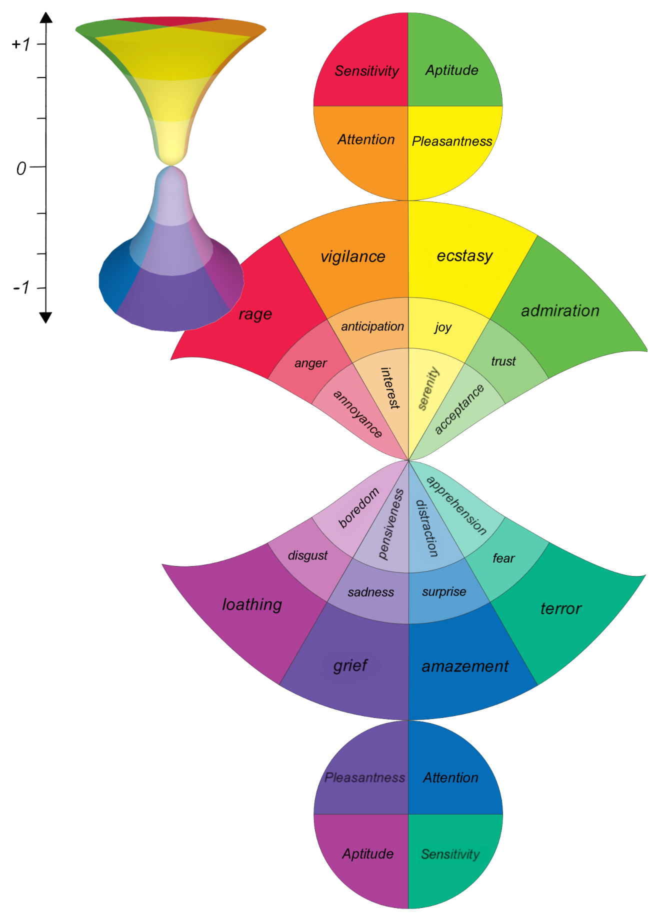

Figure 1: The Hourglass model

In the model, affective states are not classified, as often happens in the field of emotion analysis, into basic emotional categories, but rather into four concomitant but independent dimensions, characterized by six levels of activation, which determine the intensity of the expressed/perceived emotion as a $float \in$ [-1,+1]. Such levels are also labeled as a set of 24 basic emotions (six for each of the affective dimensions) in a way that allows the model to specify the affective information associated with text both in a dimensional and in a discrete form.

4. BUILDING AND USING THE API

Currently available lexical resources for opinion polarity and affect recognition such as SentiWordNet or WNA are known to be pretty noisy and limited. These resources, in fact, mainly provide opinion polarity and affective information at syntactical level, leaving out polarity and affective information for common-sense knowledge concepts like `celebrate_special_occasion`, `accomplish_goal`, `bad_feeling`, `be_on_cloud_nine`, or `lose_temper`, which are usually found in natural language text to express viewpoints and affect.

In order to build a comprehensive resource for opinion mining and sentiment analysis, we use the techniques described in Section 3 to extract both cognitive and affective information from natural language text in a way that it is possible to map it into a fixed structure. In particular, we propose to bridge the cognitive and affective gap between word-level natural language data and their relative concept-level opinions and sentiments, by building semantics and sentics on top of them (Figure 2). To this end, the Sentic API provides polarity (a float number between -1 and +1 that indicates whether a concept is positive or negative), semantics (a set of five semantically-related concepts) and sentics (affective information in terms of the Hourglass affective dimensions) associated with 15,000 natural language concepts. This information is encoded in RDF/XML using the descriptors defined by Human Emotion Ontology (HEO) [11].

4.1 Extracting Semantics

The extraction of semantics associated with common-sense knowledge concepts is performed through the ensemble application of spectral association and CF-IOF on the graph structure of ConceptNet. In particular, we apply CF-IOF on a set of 10,000 topic-tagged posts extracted from LiveJournal[1], a virtual community of more than 23 million who are allowed to label their posts not only with a topic tag but also with a mood label, by choosing from more than 130 predefined moods or by creating custom mood themes.

[1] `http://livejournal.com`

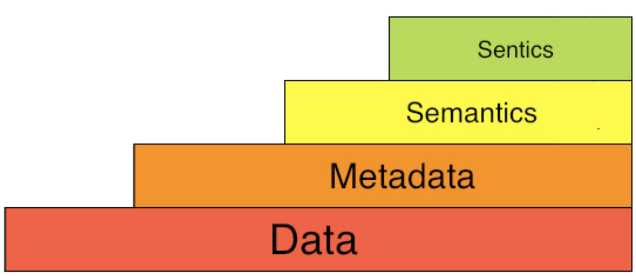

Figure 2: The semantics and sentics stack


```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about="http://sentic.net/api/concept/celebrate_special_occasion">
        <rdf:type rdf:resource="http://sentic.net/api/concept/semantics"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/celebrate_holiday"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/celebrate_occasion"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/celebrate_birthday"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/celebrate_wed"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/express_appreciation"/>
    </rdf:Description>
</rdf:RDF>
```


Figure 3: XML file resulting from querying about the semantics of `celebrate_special_occasion`

Thanks to CF-IOF weights, it is possible to filter out common concepts and detect domain-dependent concepts that individualize topics typically found in online opinions such as art, food, music, politics, family, entertainment, photography, travel, and technology. These concepts represent seed concepts for spectral association, which spreads their values across the ConceptNet graph. In particular, in order to accordingly limit the spreading activation of ConceptNet nodes, the rest of the concepts detected via CF-IOF are given as negative inputs to spectral association so that just domain-specific concepts are selected.

4.2 Extracting Sentics

The extraction of sentics associated with common-sense knowledge concepts is performed through the combined use of AffectiveSpace and the Hourglass model. In particular, we discard all but the first 100 singular values of the SVD and organize the resulting vector space using a k-medoids clustering approach [17], with respect to the Hourglass of Emotions (i.e., by using the model's labels as 'centroid concepts').

By calculating the relative distances (dot product) of each concept from the different centroids, it is possible to calculate its affective valence in terms of Pleasantness, Attention, Sensitivity and Aptitude, which is stored in the form of a four-dimensional vector, called sentic vector.

4.3 Encoding Semantics and Sentics

In order to represent the Sentic API in a machine-accessible and machine-processable way, results are encoded in RDF triples using a XML syntax (Figure 3). In particular, concepts are identified using the ConceptNet Web API and statements are encoded in RDF/XML format on the base of HEO. Statements have forms such as *concept – hasPlesantness – pleasantnessValue, concept – hasPolarity – polarityValue, and concept – isSemanticallyRelatedTo – concept*.

Given the concept `celebrate_special_occasion`, for example, the Sentic API provides a set of semantically related concepts, e.g., `celebrate_birthday`, and a sentic vector specifying Pleasantness, Attention, Sensitivity and Aptitude associated with the concept (which can be decoded into the emotions of *ecstasy* and *anticipation* and from which a positive polarity value can be inferred).

Encoding semantics and sentics in RDF/XML using the descriptors defined by HEO allows cognitive and affective information to be stored in a Sesame triple-store, a purpose-built database for the storage and retrieval of RDF metadata. Sesame can be embedded in applications and used to conduct a wide range of inferences on the information stored, based on RDFS and OWL type relations between data. In addition, it can also be used in a standalone server mode, much like a traditional database with multiple applications connecting to it.

4.4 Exploiting Semantics and Sentics

Thanks to its Semantic Web aware format, the Sentic API is very easy to interface with any real-world application that needs to extract semantics and sentics from natural language. This cognitive and affective information is supplied both at category-level (through domain and sentic labels) and dimensional-level (through polarity values and sentic vectors).

Sentic labels, in particular, are useful in case we deal with real-time adaptive applications (in which, for example, the style of an interface or the expression of an avatar has to quickly change according to labels such as 'excitement' or 'frustration' detected from user input). Polarity values and sentic vectors, in turn, are useful for tasks such as information retrieval and polarity detection (in which it is needed to process batches of documents and, hence, perform calculations, such as addition, subtraction, and average, on both conceptual and affective information).

Averaging results obtained at category-level is also possible by using a continuous 2D space whose dimensions are evaluation and activation, but the best strategy is usually to consider the opinionated document as composed of small bags of concepts (SBoCs) and feed these into the Sentic API to perform statistical analysis of the resulting sentic vectors.

To this end, we use a pre-processing module that interprets all the affective valence indicators usually contained in text such as special punctuation, complete upper-case words, onomatopoeic repetitions, exclamation words, negations, degree adverbs and emoticons, and eventually lemmatizes text.

A semantic parser then deconstructs text into concepts using a lexicon based on 'sentic n-grams', i.e., sequences of lexemes which represent multiple-word common-sense and affective concepts extracted from ConceptNet, WNA and other linguistic resources. We then use the resulting SBoC as input for the Sentic API and look up into it in order to obtain the relative sentic vectors, which we average in order to detect primary and secondary moods conveyed by the analyzed text and/or its polarity, given by the formula [6]:

$$p = \sum_{i=1}^{N} \frac{Plsnt(c_i) + |Attnt(c_i)| - |Snst(c_i)| + Aptit(c_i)}{3N}$$

where N is the size of the SBoC. As an example of how the Sentic API can be exploited for microblogging analysis, intermediate and final outputs obtained when a natural language opinion is given as input to the system can be examined. The tweet "I think iPhone4 is the top of the heap! OK, the speaker is not the best i hv ever seen bt touchscreen really puts me on cloud 9... camera looks pretty good too!" is selected. After the pre-processing and semantic parsing operations, the following SBoCs are obtained:

SBoC#1:
<*Concept: 'think'*>
<*Concept: 'iphone4'*>
<*Concept: 'top heap'*>
SBoC#2:
<*Concept: 'ok'*>
<*Concept: 'speaker'*>
<*Concept: !'good'++*>
<*Concept: 'see'*>
SBoC#3:
<*Concept: 'touchscreen'*>
<*Concept: 'put cloud nine'++*>
SBoC#4:
<*Concept: 'camera'*>
<*Concept: 'look good'−−*>

Table 1: Structured output example

Opinion Target	Category	Moods	Polarity
'iphone4'	'phones', 'electronics'	'ecstasy', 'interest'	+0.71
'speaker'	'electronics', 'music'	'annoyance'	-0.34
'touchscreen'	'electronics'	'ecstasy', 'anticipation'	+0.82
'camera'	'photography', 'electronics'	'acceptance'	+0.56

After feeding the extracted concepts to the Sentic API, we can exploit semantics and sentics to detect opinion targets and obtain, for each of these, the relative affective information both in a discrete way (with one or more emotional labels) and in a dimensional way (with a polarity value $\in$ [-1,+1]) as shown in Table 1.

5. EVALUATION

As a use case evaluation of the proposed API, we select the problem of crowd validation of the UK national health service (NHS) [5], that is, the exploitation of the wisdom of patients to adequately validate the official hospital ratings made available by UK health-care providers and NHS Choices[2]. To validate such data, we exploit patient stories extracted from PatientOpinion[3], a social enterprise providing an online feedback service for users of the UK NHS. The problem is that this social information is often stored in natural language text and, hence, intrinsically unstructured, which makes comparison with the structured information supplied by health-care providers very difficult. To bridge the gap between such data (which are different at structure-level yet similar at concept-level), we exploit the Sentic API to marshal PatientOpinion's social information in a machine-accessible and machine-processable format and, hence, compare it with the official hospital ratings provided by NHS Choices and each NHS trust.

In particular, we use Sentic API's inferred ratings to validate the information declared by the relevant health-care providers, crawled separately from each NHS trust website, and the official NHS ranks, extracted using the NHS Choices API[4]. This kind of data usually consists of ratings that associate a polarity value to specific features of health-care providers such as 'communication', 'food', 'parking', 'service', 'staff', and 'timeliness'. The polarity can be either a number in a fixed range or simply a flag (positive/negative).

Since each patient opinion can regard more than one topic and the polarity values associated with each topic are often independent from each other, we need to extract, from each opinion, a set of topics and then, from each topic detected, the polarity associated with it. Thus, after deconstructing each opinion into a set of SBoCs, we analyze these through Sentic API in order to tag each SBoC with one of the relevant topics (if any) and calculate a polarity value. We ran this process on a set of 857 topic- and polarity-tagged short stories extracted from PatientOpinion database and computed recall and precision rates as evaluation metrics.

As for the SBoC categorization, results showed that the Sentic API can detect topics in patient stories with satisfactory accuracy. In particular, the classification of stories about 'food' and 'communication' was performed with a precision of 80.2% and 73.4% and recall rates of 69.8% and 61.4%, for a total F-measure of 74.6% and 66.8%, respectively.

[2] `http://nhs.uk`
[3] `http://patientopinion.org.uk`
[4] `http://data.gov.uk/data`

Table 2: Comparative evaluation against WNA and SenticNet

Category	WNA	SenticNet	Sentic API
clinical service	59.12%	69.52%	78.06%
communication	66.81%	76.35%	80.12%
food	67.95%	83.61%	85.94%
parking	63.02%	75.09%	79.42%
staff	58.37%	67.90%	76.19%
timeliness	57.98%	66.00%	75.98%

As for the polarity detection, in turn, positivity and negativity of patient opinions were identified with particularly high precision (91.4% and 86.9%, respectively) and good recall rates (81.2% and 74.3%), for a total F-measure of 85.9% and 80.1%, respectively. More detailed comparative statistics are listed in Table 2, where the Sentic API is compared against WNA and SenticNet with respect to the polarity detection F-measures obtained on the 857 short stories.

6. CONCLUSION

Today user-generated contents are perfectly suitable for human consumption, but they remain hardly accessible to machines. Currently available information retrieval tools still have to face a lot of limitations. To bridge the conceptual and affective gap between word-level natural language data and the concept-level opinions and sentiments conveyed by them, we developed Sentic API, a common-sense based application programming interface that provides semantics and sentics associated with 15,000 natural language concepts.

We showed how Sentic API can easily be embedded in real-world NLP applications, specifically in the field of microblogging analysis, where statistical methods usually fail as syntax-based text processing works well only on formal-English documents and after training on big text corpora. We are keeping on developing the resource in a way that it can be continuously enhanced with more concepts from the always-growing Open Mind corpus and other publicly available common and common-sense knowledge bases. We are also developing novel techniques and tools to allow the Sentic API to be more easily merged with external domain-dependent knowledge bases, in order to improve the extraction of semantics and sentics from many different types of media and contexts.

7. REFERENCES

[1] C. Bazzanella. Emotions, language and context. In E. Weigand, editor, *Emotion in dialogic interaction. Advances in the complex*, pages 59–76. Benjamins, Amsterdam, 2004.

[2] C. Bosco, V. Patti, and A. Bolioli. Developing corpora for sentiment analysis and opinion mining: A survey and the Senti-TUT case study. *IEEE Intelligent Systems*, 28(2):55–63, 2013.

[3] M. Bradley and P. Lang. Affective norms for english words (ANEW): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida., 1999.

[4] E. Cambria and A. Hussain. *Sentic Computing: Techniques, Tools, and Applications*. Springer, Dordrecht, Netherlands, 2012.

[5] E. Cambria, A. Hussain, C. Havasi, C. Eckl, and J. Munro. Towards crowd validation of the UK national health service. In *WebSci*, Raleigh, 2010.

[6] E. Cambria, A. Livingstone, and A. Hussain. The hourglass of emotions. In A. Esposito, A. Vinciarelli, R. Hoffmann, and V. Muller, editors, *Cognitive Behavioral Systems*, volume 7403 of *Lecture Notes in Computer Science*, pages 144–157. Springer, Berlin Heidelberg, 2012.

[7] E. Cambria, R. Speer, C. Havasi, and A. Hussain. SenticNet: A publicly available semantic resource for opinion mining. In *AAAI CSK*, pages 14–18, Arlington, 2010.

[8] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[9] A. Esuli and F. Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *LREC*, 2006.

[10] L. García-Moya, H. Anaya-Sanchez, and R. Berlanga-Llavori. A language model approach for retrieving product features and opinions from customer reviews. *IEEE Intelligent Systems*, 28(3):19–27, 2013.

[11] M. Grassi. Developing HEO human emotions ontology. volume 5707 of *Lecture Notes in Computer Science*, pages 244–251. Springer, Berlin Heidelberg, 2009.

[12] C. Havasi, R. Speer, and J. Holmgren. Automated color selection using semantic knowledge. In *AAAI CSK*, Arlington, 2010.

[13] R. Horn and C. Johnson. Norms for vectors and matrices. In *Matrix Analysis*, chapter 5. Cambridge University Press, 1990.

[14] C. Hung and H.-K. Lin. Using objective words in SentiWordNet to improve sentiment classification for word of mouth. *IEEE Intelligent Systems*, 28(2):47–54, 2013.

[15] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of The National Bureau of Standards*, 45(4):255–282, 1950.

[16] M. Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, New York, 2006.

[17] H. Park and C. Jun. A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2):3336–3341, 2009.

[18] R. Plutchik. The nature of emotions. *American Scientist*, 89(4):344–350, 2001.

[19] S. Poria, A. Gelbukh, A. Hussain, D. Das, and S. Bandyopadhyay. Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2):31–30, 2013.

[20] R. Speer and C. Havasi. ConceptNet 5: A large semantic network for relational knowledge. In E. Hovy, M. Johnson, and G. Hirst, editors, *Theory and Applications of Natural Language Processing*, chapter 6. Springer, 2012.

[21] C. Strapparava and A. Valitutti. WordNet-Affect: An affective extension of WordNet. In *LREC*, pages 1083–1086, Lisbon, 2004.

[22] A. Tsai, R. Tsai, and J. Hsu. Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intelligent Systems*, 28(2):22–30, 2013.

[23] M. Wall, A. Rechtsteiner, and L. Rocha. Singular value decomposition and principal component analysis. In D. Berrar, W. Dubitzky, and M. Granzow, editors, *A Practical Approach to Microarray Data Analysis*, pages 91–109. Springer, 2003.

[24] A. Weichselbraun, S. Gindl, and A. Scharl. Extracting and grounding context-aware sentiment lexicons. *IEEE Intelligent Systems*, 28(2):39–46, 2013.