SenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis

Erik Cambria
School of Computer Engineering
Nanyang Technological University
cambria@ntu.edu.sg

Daniel Olsher
The Robotics Institute
Carnegie Mellon University
dan@intmind.com

Dheeraj Rajagopal
Temasek Laboratories
National University of Singapore
dheeraj@nus.edu.sg

Abstract

SenticNet is a publicly available semantic and affective resource for concept-level sentiment analysis. Rather than using graph-mining and dimensionality-reduction techniques, SenticNet 3 makes use of 'energy flows' to connect various parts of extended common and common-sense knowledge representations to one another. SenticNet 3 models nuanced semantics and sentics (that is, the conceptual and affective information associated with multi-word natural language expressions), representing information with a symbolic opacity of an intermediate nature between that of neural networks and typical symbolic systems.

Introduction

As society evolves at the same pace as the Web, online social data is becoming increasingly important for both individuals and businesses. The Web 2.0, however, has unleashed an era of online participation that is causing user-generated content (UGC) to grow exponentially and, hence, to become continuously larger and more complex. In order to truly achieve *collective* – rather than merely *collected* – intelligence (Gruber 2007), a shift from lexical semantics to compositional semantics is required.

In recent years, sentiment analysis research has gradually been developing into a field itself that lies in between natural language processing (NLP) and natural language understanding. Unlike standard syntactical NLP tasks such as summarization and auto-categorization, opinion mining mainly deals with the inference of the semantics and sentics (denotative and connotative information) associated with natural language concepts, without strictly requiring a deep understanding of the given text (Cambria and White 2014).

In order to infer the polarity of a sentence, in fact, an opinion mining engine only needs to extract the features or aspects of the discussed service or product, e.g., size or weight of a phone, and the sentiments associated with each of these, e.g., positive or negative, so that the output of a sentence such as "I love the phone's touchscreen but its battery life is too short" would be something like <touchscreen: +> and <battery: –>.

Because of the ambiguity and the complexity of natural language, however, the correct inference of such semantics and sentics is not trivial and requires an analysis of source texts that is content-, concept-, and context-based. Content-level analysis is needed for the collection of opinions over the Web, while filtering out non-opinionated UGC (subjectivity detection), and for accordingly balancing the trustworthiness of such opinions with respect to their source.

Context-level analysis, in turn, ensures that all gathered opinions are relevant for the specific user. In the era of social context (where intelligent systems have access to a great deal of personal identities and social relationships), opinion mining will be tailored to each user's preferences and intent. Irrelevant opinions will be accordingly filtered with respect to their source (e.g., a relevant circle of friends or users with similar interests) and intent (e.g., selection of camera for trekking, rather than for night shooting).

Finally, concept-level analysis aims to infer the semantics and sentics associated with natural language opinions and, hence, to enable a comparative fine-grained feature-based sentiment analysis. Rather than gathering isolated opinions about a whole item (e.g., iPhone5), users are generally more interested in comparing different products according to their specific features (e.g., iPhone5's versus Galaxy S4's touchscreen), or even sub-features (e.g., fragility of iPhone5's versus Galaxy S4's touchscreen).

In this context, SenticNet 3 exploits an energy-based knowledge representation (EBKR) formalism (Olsher 2014) to provide the semantics and sentics associated with 30,000 multi-word expressions and, hence, enables a deeper and more multi-faceted analysis of natural language opinions. In particular, non-trivial multi-word associations such as 'low price' or 'long queue' are preferred over concepts such as 'good restaurant' or 'awful service'.

Adjectives such as 'good' and 'awful' are unambiguous from a polarity point of view, i.e., they convey a positive and negative polarity respectively, no matter which noun they are associated with. Adjectives such as 'long' and 'high', instead, do not carry any specific polarity on their own, but rather assume one depending on which noun they are coupled with. Hence, SenticNet 3 contains unambiguous adjectives as standalone entries plus non-trivial multi-word expressions such as 'small room' or 'cold bed'.

SenticNet 3 focuses on the use of 'energy' or information flows to connect various parts of extended 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. Unlike state-of-the-art techniques, such a framework enables the representation of information with a symbolic opacity between that of artificial neural networks and of typical symbolic systems. In essence, pieces of common and common-sense knowledge are broken down into 'atoms', which allow the fusing of data coming from other knowledge bases without requiring complex inter-source ontology alignment and the aggregation arising from multiple sources during reasoning.

The rest of this paper is organized as follows: the first section is a brief overview of main approaches to opinion mining; followed by a section describing types and sources of knowledge exploited in this work; the next section explains how SenticNet 3 is built from such sources; then, a section explaining how SenticNet 3 can be used for concept-level sentiment analysis is presented; afterwhich a section focused on evaluation is presented; finally, some concluding remarks and future work recommendations are made.

Related Work

Due to many challenging research problems and a wide variety of practical applications, opinion mining and sentiment analysis have become very active research areas in the last decade. Common opinion mining tasks include product feature retrieval (García-Moya, Anaya-Sanchez, and Berlanga-Llavori 2013), opinion holder detection (Gangemi, Presutti, and Reforgiato 2014), opinion summarization (Di Fabbrizio, Aker, and Gaizauskas 2013), domain adaptation (Xia et al. 2013), and cyber-issue detection (Lau, Xia, and Ye 2014).

Existing approaches to the task of polarity classification can be grouped into four main categories: keyword spotting, in which text is classified into categories based on the presence of fairly unambiguous affect words (Elliott 1992; Wiebe, Wilson, and Cardie 2005); lexical affinity, which assigns arbitrary words a probabilistic affinity for a particular topic or emotion (Rao and Ravichandran 2009; Weichselbraun, Gindl, and Scharl 2013); statistical methods, which calculate the valence of word co-occurrence frequencies on the base of a large training corpus (Poria et al. 2013); and concept-level approaches, which make use of semantic networks to infer conceptual and affective information conveyed by natural language concepts (Cambria and Hussain 2012; Tsai, Tsai, and Hsu 2013; Olsher 2014).

Unlike other sentiment analysis resources such as WordNet-Affect (WNA) (Strapparava and Valitutti 2004), SenticNet 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. SenticNet 3, in particular, uses a novel EBKR formalism and, besides common and common-sense knowledge, also exploits affective knowledge modeled by a biologically-inspired emotion categorization model.

Knowledge Sources

In standard human-to-human communication, people usually refer to existing facts and circumstances and build new useful, funny, or interesting information on the top of those. This common knowledge includes information usually found in news, articles, debates, lectures, etc. (factual knowledge), but also principles and definitions that can be found in collective intelligence projects such as Wikipedia (vocabulary knowledge).

Moreover, when people communicate with each other, they rely on similar background knowledge, e.g., the way objects relate to each other in the world, people's goals in their daily lives, and the emotional content of events or situations. This taken-for-granted information is what is termed common-sense – obvious things people normally know and usually leave unstated.

The difference between common and common-sense knowledge can be expressed as the difference between knowing the name of something and truly understanding something. For example, you can know the name of all the different kinds or brands of 'pipe', but you understand nothing about a pipe until you get to know how to grab it, pack it, light it, and smoke it. In other words, a 'pipe' is not a pipe unless it can be used.

Common Knowledge Sources

Attempts to build a common knowledge base are countless and include both resources crafted by human experts or community efforts, such as DBPedia (Bizer et al. 2009), a collection of 2.6 million entities extracted from Wikipedia, and Freebase (Bollacker et al. 2008), a social database of 1,450 concepts, and automatically-built knowledge bases, such as YAGO (Suchanek, Kasneci, and Weikum 2007), a semantic knowledge base of 149,162 instances derived from Wikipedia Infoboxes and WordNet, NELL (Carlson et al. 2010), with 242,000 beliefs mined from the Web, and Probase (Wu et al. 2012), Microsoft's probabilistic taxonomy counting about 12 million concepts learned iteratively from 1.68 billion web pages in Bing web repository.

Common-Sense Knowledge Sources

One of the biggest projects aiming to build a comprehensive common-sense knowledge base is Cyc (Lenat and Guha 1989). Cyc, however, requires the involvement of experts working on a specific programming language, which makes knowledge engineering labor-intensive and time-consuming. A more recent and scalable project is Open Mind Common Sense (OMCS), which is collecting 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 exploits these pieces of common-sense knowledge to automatically build ConceptNet (Speer and Havasi 2012), a semantic network of 173,398 nodes. Other projects that fall under this umbrella include WordNet, with its 25,000 synsets, and derivative resources such as WNA.

Building SenticNet 3

Unlike previous versions (which focused only on common-sense knowledge), SenticNet 3 contains both common and common-sense knowledge, in order to boost sentiment analysis tasks such as feature spotting and polarity detection, respectively. In particular, an ensemble of all the above-mentioned resources was created, with the exception of Freebase and NELL (as the knowledge they contain is rather noisy and mostly deals with geolocation data) and YAGO (which is mostly derived from DBPedia and WordNet, already embedded in SenticNet 3).

Knowledge Integration

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 through the EBKR formalism. For example, a piece of knowledge such as "Pablo Picasso is an artist" is automatically translated into the RDF triple <*Pablo Picasso*-isA-*artist*> and, hence, into the EBKR entry [*Pablo Picasso* –PARTICIP-CATEGORY–> *artist*]. The purpose of such an integration process is two-fold: firstly, it provides a shared representation for common and common-sense knowledge to be efficiently stored and, hence, used for reasoning; secondly, it performs 'conceptual decomposition' of opaque relation types. The EBKR formalism enables the representation of pieces of knowledge in a common framework, which allows the fusing of data from different sources without requiring ontology alignment and to combine data arising from multiple knowledge bases during reasoning (Kuo and Hsu 2012). Conceptual decomposition allows the unfolding of relation types that are usually opaque in natural-language-based resources, in order to aid common-sense reasoning.

The paradigm of cognition-driven reasoning is ideology: abstract ideological schemas are the prototype of efficiency in information processing, supplying a fairly general but densely elaborated set of constructs in terms of which information can be effectively organized, stored, and retrieved (Sniderman, Brody, and Tetlock 1991).

For example, the ConceptNet relation type *CausesDesire* does not really mean much to the machine unless we unpack such a 'suitcase word' into a more semantic-preserving fragment that specifies how subject and object are affected by the predicate (Figure 1). This way, rather than simply having an opaque tag on a semantic network edge, a substructure defining how such a relationship might change the goal of a person highly enhances reasoning and decision-making processes. Such substructures are termed 'concept fields' and represent delineated subregions of the semantic network that have a certain semantic coherence sufficient (and often necessary) to define them as 'meaningfully' connected fields.

After low confidence score trimming and duplicate 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.

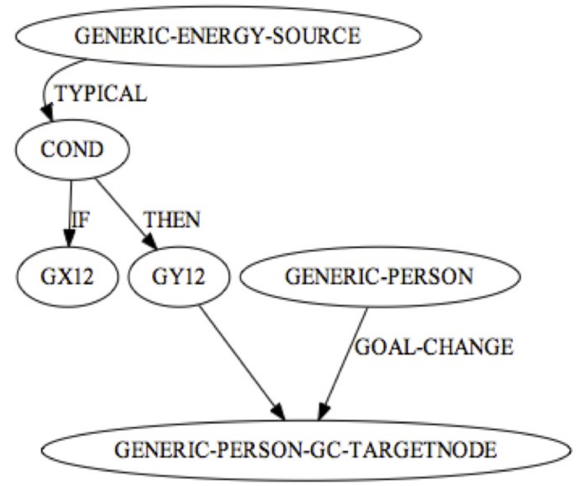

Figure 1: Semantic fragment of the EBKR formalism showing how the ConceptNet relation *CausesDesire* is translated into a more meaningful structure.

Traversing the Knowledge Base

The way semantics and sentics are defined in SenticNet 3 is inspired by neuroscience and cognitive psychology theories on associative memory. According to Westen, for example, associative memory involves the unconscious activation of networks of association – thoughts, feelings, wishes, fears, and perceptions that are connected, so that activation of one node in the network leads to activation of the others (Westen 2002). Memory is not a 'thing' that is stored somewhere in a mental warehouse and can be pulled out and brought to the fore. Rather, it is a potential for reactivation of a set of concepts that together constitute a particular meaning.

In this context, EBKR represents complex concepts by setting up pathways upon which information (conceptualized as 'energy') may flow between various semantic fragments. Rather than using symbolic representations, the key idea is that complex representations can be built up from simpler pieces by connecting them together via energy flows. Each element reached by a certain quantum of energy flow participates in and becomes part of a wider concept representation. Through this mechanism, conceptual connections between simple elements deeply affect the modeling of larger systems. 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, taken together, describe the semantics and sentics indicated by that quantum of energy. Such conceptual and affective information is extracted, for each concept of the semantic network, by analyzing how energy flows serve the function of quickly spreading an associative relevancy measure over declarative memory (Anderson and Pirolli 1984).

Concepts have long-term affective information attached to them in memory and such information exerts differential, generally significant, effects upon cognition in different contexts (Lodge and Taber 2005). Thus, the extraction of semantics and sentics is achieved through multiple steps of spreading activation (Cambria, Olsher, and Kwok 2012) with respect to the nodes representing the activation levels of the Hourglass of Emotions (Cambria and Hussain 2012), a brain-inspired model for the representation and the analysis of human emotions.

The main advantage of the Hourglass of Emotions over other emotion categorization models is that it allows emotions to be deconstructed into independent but concomitant affective dimensions. 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 and, hence, work with emotions both in a categorical way and in a dimensional format.

Besides exploiting the semantic connections built by means of common and common-sense knowledge integration, such a process heavily exploits also the links established through the integration of WNA, which helps to enhance the affective connections between standard nodes and the seed nodes representing the activation levels of the Hourglass model, e.g., 'joy', 'anger', or 'surprise'.

Encoding Semantics and Sentics

In order to represent SenticNet 3 in a machine-accessible and machine-processable way, results are encoded in RDF using a XML syntax. In particular, concepts are identified using the SenticNet API[1] and, for each of them, semantics and sentics (e.g., category-tags and mood-tags) are provided in RDF/XML format.

Given the concept 'birthday party', for example, SenticNet 3 provides 'event' as high-level domain of pertinence (which can be useful for tasks such as gisting or document auto-categorization) and a set of semantically related concepts, e.g., 'cake', 'surprise friend' or 'gift' (which can be exploited as extra/contextual information for improving search results). 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), a primary and secondary mood (for tasks such as HCI), and a set of affectively related concepts, e.g., 'celebration' or 'special occasion' (for tasks such as opinion classification).

The encoding of semantics and sentics in RDF/XML format is mainly driven by the need of exportability and interoperability, but also to allow conceptual and affective information to be stored in a triplestore, a purpose-built database for the storage and retrieval of RDF metadata, which can be used to conduct a wide range of inferences based on RDFS and OWL type relations between data.

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

Working with SenticNet 3

SenticNet 3 can be either downloaded as a standalone resource[2] or accessed online either through an API or as a Python web service[3]. Thanks to its Semantic Web aware format, it is very easy to interface the resource with any real-world application that needs to extract semantics and sentics from natural language. This conceptual and affective information is supplied both at category-level (through domain and sentic labels) and dimensional-level (through polarity values and sentic vectors). 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 user's 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 can be done using a continuous evaluation-activation space, e.g., Whissell space, but the best strategy is usually to consider the opinionated document as composed of small bags of concepts (SBoCs) and feed these into SenticNet 3 to perform statistical analysis of the resulting sentic vectors. SenticNet 3, however, only provides semantics and sentics at concept-level.

Hence, to build a comprehensive cognition-driven opinion-mining engine, it is necessary to couple the resource with a pre-processing module, a semantic parser, and an opinion target detector. After such modules deconstruct natural language text into concepts and extract opinion targets, the concepts associated with each detected target are given as input to SenticNet 3 to look up semantics and sentics and, hence, calculate polarity[4].

Pre-Processing

The pre-processing module exploits linguistic dictionaries to interpret all the affective valence indicators usually contained in opinionated text, e.g., special punctuation, complete upper-case words, cross-linguistic onomatopoeias, exclamation words, degree adverbs, and emoticons.

Moreover, the module detects negation and spreads it in a way that it can be accordingly associated to concepts during the parsing phase. Handling negation is an important concern in sentiment analysis, as it can reverse the meaning of a statement. Such a task, however, is not trivial as not all appearances of explicit negation terms reverse the polarity of the enclosing sentence and negation can often be expressed in rather subtle ways, e.g., sarcasm and irony, which are quite difficult to detect. Lastly, the module converts text to lower-case and, after lemmatizing it, splits the opinion into single clauses (SBoCs) according to grammatical conjunctions and punctuation.

[2]http://sentic.net/downloads
[3]http://pypi.python.org/pypi/senticnet
[4]http://sentic.net/demo

Semantic Parsing

Semantic parsing is performed through a graph-based approach to common-sense concept extraction, which breaks sentences into chunks first and then extracts concepts by selecting the best match from a parse graph that maps all the multi-word expressions contained in SenticNet 3. Each verb and its associated noun phrase are considered in turn, and one or more concepts is extracted from these. As an example, the clause "I went for a walk in the park", would contain the concepts *go walk* and *go park*.

The Stanford Chunker (Manning 2011) is used to chunk the input text. A general assumption during clause separation is that, if a piece of text contains a preposition or subordinating conjunction, the words preceding these function words are interpreted not as events but as objects. Next, clauses are normalized in two stages. First, each *verb* chunk is normalized using the Lancaster stemming algorithm (Paice 1990). Second, each potential *noun* chunk associated with individual verb chunks is paired with the stemmed verb in order to detect multi-word expressions of the form 'verb plus object'.

The POS-based bigram algorithm extracts concepts such as *market*, *some fruits*, *fruits*, and *vegetables*. In order to capture event concepts, matches between the object concepts and the normalized verb chunks are searched. This is done by exploiting a parse graph that maps all the multi-word expressions contained in SenticNet 3. Such an unweighted directed graph helps to quickly detect multi-word concepts, without performing an exhaustive search throughout all the possible word combinations that can form a common-sense concept. Single-word concepts, e.g., *house*, that already appear in the clause as a multi-word concept, e.g., *beautiful house*, in fact, are pleonastic and are discarded.

Opinion Target Detection

Opinion targets exhibit semantic coherence in that they tend to generate lexical items and phrases with related semantics. Most words related to the same target tend to share some semantic characteristics. Our common-sense-based approach is similar to the process undertaken by humans when finding similar items - we look at what the *meanings* of the items have in common.

Thus, under our model, opinion targets are not discovered merely based on document co-occurrence, but rather by considering the definitive semantic character of constituent concepts. In SenticNet 3, concepts inter-define one another, with directed edges indicating semantic dependencies between concepts. In the present algorithm, the features for any particular concept C are defined as the set of concepts reachable via outbound edges from C. Put differently, for each input concept we retrieve those other concepts that, collectively, generate the core semantics of the input concept.

Our algorithm uses clustering to generate opinion targets from semantic features. Based on experiments with various clustering algorithms, e.g., k-means (Hartigan and Wong 1979) and expectation-maximization (EM) clustering (Dempster, Laird, and Rubin 1977), we determined that group average agglomerative clustering (GAAC) provides the highest accuracy.

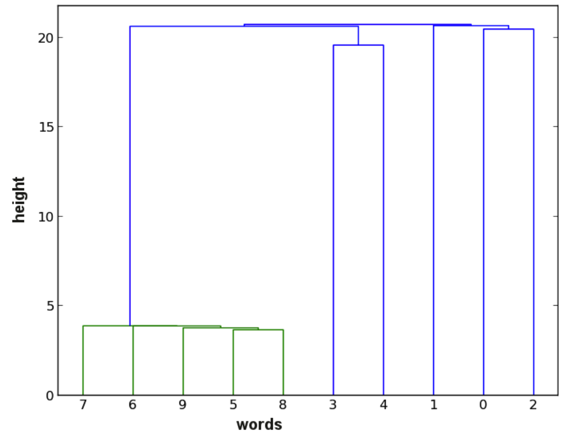

Figure 2: A sample dendrogram resulting from hierarchical clustering.

GAAC partitions data into trees (Berkhin 2006) containing *child* and *sibling* clusters. It generates dendrograms specifying nested groupings of data at various levels (Jain and Dubes 1988). During clustering, documents are represented as vectors of common-sense concepts. For each concept, the corresponding features are extracted from Senticnet 3. The proximity matrix is constructed with concepts as rows and features as columns. If a feature is an outbound link of a concept, the corresponding entry in the matrix is 1, while in other situations it is 0. Cosine distance is used as the distance metric.

Agglomerative algorithms are bottom-up in nature. GAAC consists of the following steps:

1. Compute proximity matrix. Each data item is an initial cluster.
2. From the proximity matrix, form pair of clusters by merging. Update proximity matrix to reflect merges.
3. Repeat until all clusters are merged.

A sample dendrogram is shown in Figure 2. The dendogram is pruned at a height depending on the number of desired clusters. The *group average* between the clusters is given by the average similarity distance between the groups. Distances between two clusters and similarity measures are given by the equations below:

$$X_{sum} = \sum_{d_m \in \omega_i v \omega_i} \sum_{d_n \in \omega_i v \omega_j d_n \neq d_m} \overrightarrow{d_n}.\overrightarrow{d_m} \quad (1)$$

$$sim\left(\omega_i, \omega_j\right) = \frac{1}{\left(N_i + N_j\right)\left(N_i + N_j - 1\right)} X_{sum} \quad (2)$$

where $\overrightarrow{d}$ is the vector of document of length d. Vector entries are boolean, 1 if the feature is present, 0 otherwise. N_i, N_j is the number of features in ω_i and ω_j respectively, which denote clusters.

The main drawback of the hierarchical clustering algorithm is the running complexity (Berkhin 2006), which averages $\theta(N^2 log\, N)$.

"horse"	"stationery"	"food"	"party"
horse	paper	apple	dance
eye	paint	fish	protest
farm	plate	bread	music
	card	cake	party
	metal		door
			sound

Table 1: Example of feature-based clustering

We choose average link clustering because our clustering is connectivity-based. The concept proximity matrix consists of features from SenticNet 3 and 'good' connections occur when two concepts share multiple features. After clustering, the number of clusters is determined and the dendrogram is pruned accordingly. The output of this process is the set of opinion targets present in the document. Table 1 provides an example of the results of feature-based clustering.

Use Case Evaluation

As a use case evaluation of the system, we select the problem of crowd validation of the UK national health service (NHS), 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[5].

To validate such data, we exploit patient stories extracted from PatientOpinion[6], 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 SenticNet 3 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 SenticNet 3 inferred polarity values to assess the official NHS ranks (which we extracted using the NHS Choices API[7]) and the ratings of relevant health-care providers (which we crawled from each NHS trust website individually). 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 be regarding more than one topic and the polarity values associated with each topic are often independent from each other, in order to efficiently perform the mapping, we need to extract (from each opinion) a set of topics and then (from each topic detected) the polarity associated with it.

[5]http://nhs.uk

[6]http://patientopinion.org.uk

[7]http://data.gov.uk/data-requests/nhs-choices-api

Thus, after deconstructing each opinion into a set of SBoCs, we analyze these through SenticNet 3 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 2000 topic- and polarity-tagged stories extracted from PatientOpinion database and computed recall and precision rates as evaluation metrics. On average, each post contained around 140 words, from which about 12 affective valence indicators and 60 concepts were extracted.

As for the SBoC categorization, results showed that SenticNet 3 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 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 SenticNet 3 is compared with its former versions, namely SenticNet and SenticNet 2.

Category	**SenticNet**	**SenticNet 2**	**SenticNet 3**
clinical service	59.12%	69.52%	78.06%
communi-cation	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%

Table 2: F-measures relative to patient opinions' polarity detection using the different versions of SenticNet.

Conclusion and Future Efforts

Today UGCs are perfectly suitable for human consumption, but they remain hardly accessible by machines. Currently available information retrieval tools still face many 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 SenticNet 3, a publicly available resource for concept-level sentiment analysis that associates semantics and sentics to 30,000 common and common-sense concepts.

We showed how SenticNet 3 can easily be embedded in real-world applications, specifically in the field of social data mining, in order to effectively combine and compare structured and unstructured information. We are continually developing the resource such 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 SenticNet 3 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.

References

Anderson, J., and Pirolli, P. 1984. Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:791–799.

Berkhin, P. 2006. A survey of clustering data mining techniques. In *Grouping multidimensional data*. Springer. 25–71.

Bizer, C.; Jens, L.; Kobilarov, G.; Auer, S.; Becker, C.; Cyganiak, R.; and Hellmann, S. 2009. Dbpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web* 7(3):154–165.

Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *SIGMOD*, 1247–1250.

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

Cambria, E., and White, B. 2014. Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine* 9(2):48–57.

Cambria, E.; Olsher, D.; and Kwok, K. 2012. Sentic activation: A two-level affective common sense reasoning framework. In *AAAI*, 186–192.

Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka, E.; and Mitchell, T. 2010. Toward an architecture for never-ending language learning. In *AAAI*, 1306–1313.

Dempster, A.; Laird, N.; and Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society, Series B* 39(1):1–38.

Di Fabbrizio, G.; Aker, A.; and Gaizauskas, R. 2013. Summarizing on-line product and service reviews using aspect rating distributions and language modeling. *IEEE Intelligent Systems* 28(3):28–37.

Elliott, C. D. 1992. *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System*. Ph.D. Dissertation, Northwestern University, Evanston.

Gangemi, A.; Presutti, V.; and Reforgiato, D. 2014. Frame-based detection of opinion holders and topics: a model and a tool. *IEEE Computational Intelligence Magazine* 9(1):20–30.

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

Gruber, T. 2007. Collective knowledge systems: The social web meets the semantic web. *Web Semantics* 6(1):4–13.

Hartigan, J., and Wong, M. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society* 28(1):100–108.

Jain, A., and Dubes, R. 1988. *Algorithms for clustering data*. Prentice-Hall, Inc.

Kuo, Y., and Hsu, J. 2012. Planning for reasoning with multiple common sense knowledge bases. *ACM Transactions on Interactive Intelligent Systems* 2(3):1–24.

Lau, R.; Xia, Y.; and Ye, Y. 2014. A probabilistic generative model for mining cybercriminal networks from online social media. *IEEE Computational Intelligence Magazine* 9(1):31–43.

Lenat, D., and Guha, R. 1989. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Boston: Addison-Wesley.

Lodge, M., and Taber, C. 2005. The automaticity of affect for political leaders, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology* 26(3):455–482.

Manning, C. 2011. Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In Gelbukh, A., ed., *Computational Linguistics and Intelligent Text Processing*, volume 6608 of *Lecture Notes in Computer Science*. Springer. 171–189.

Olsher, D. 2014. Semantically-based priors and nuanced knowledge core for big data, social AI, and language understanding. *Neural Networks*.

Paice, C. 1990. Another stemmer. *SIGIR Forum* 24:56–61.

Poria, S.; Gelbukh, A.; Hussain, A.; Howard, N.; Das, D.; and Bandyopadhyay, S. 2013. Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems* 28(2):31–38.

Rao, D., and Ravichandran, D. 2009. Semi-supervised polarity lexicon induction. In *EACL*, 675–682.

Sniderman, P.; Brody, R.; and Tetlock, P. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge University Press.

Speer, R., and Havasi, C. 2012. ConceptNet 5: A large semantic network for relational knowledge. In *Theory and Applications of Natural Language Processing*. Springer.

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

Suchanek, F.; Kasneci, G.; and Weikum, G. 2007. Yago: a core of semantic knowledge. In *WWW*, 697–706.

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

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

Westen, D. 2002. Implications of developments in cognitive neuroscience for psychoanalytic psychotherapy. *Harv Rev Psychiatry* 10(6):369–73.

Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39(2):165–210.

Wu, W.; Li, H.; Wang, H.; and Zhu, K. 2012. Probase: A probabilistic taxonomy for text understanding. In *SIGMOD*, 481–492.

Xia, R.; Zong, C.; Hu, X.; and Cambria, E. 2013. Feature ensemble plus sample selection: A comprehensive approach to domain adaptation for sentiment classification. *IEEE Intelligent Systems* 28(3):10–18.