# SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives

**Erik Cambria, Soujanya Poria, Rajiv Bajpai**
Nanyang Technological University
50 Nanyang Ave, Singapore 639798
{cambria,sporia
rbajpai}@ntu.edu.sg

**Björn Schuller**
Imperial College London
180 Queens' Gate, Huxley Bldg.,
London SW7 2AZ, UK
bjoern.schuller@imperial.ac.uk

## Abstract

An important difference between traditional AI systems and human intelligence is the human ability to harness commonsense knowledge gleaned from a lifetime of learning and experience to make informed decisions. This allows humans to adapt easily to novel situations where AI fails catastrophically due to a lack of situation-specific rules and generalization capabilities. Commonsense knowledge also provides background information that enables humans to successfully operate in social situations where such knowledge is typically assumed. Since commonsense consists of information that humans take for granted, gathering it is an extremely difficult task. Previous versions of SenticNet were focused on collecting this kind of knowledge for sentiment analysis but they were heavily limited by their inability to generalize. SenticNet 4 overcomes such limitations by leveraging on conceptual primitives automatically generated by means of hierarchical clustering and dimensionality reduction.

## 1 Introduction

The opportunity to capture the opinion of the general public has raised growing interest within both the scientific community as well as the business world, due to the remarkable benefits to be had from marketing and financial prediction, which has led to many exciting open challenges (Pang and Lee, 2008; Liu, 2012). Mining opinions and sentiments from natural language, however, is an extremely difficult task as it requires a deep understanding of most of the explicit and implicit, regular and irregular, syntactic and semantic rules of a language. Existing approaches to sentiment analysis mainly rely on parts of text in which opinions are explicitly expressed such as polarity terms, affect words, and their co-occurrence frequencies. However, opinions and sentiments are often conveyed implicitly through latent semantics, which make purely syntactic approaches ineffective.

SenticNet (Cambria et al., 2014) captures such latent information in terms of *semantics* and *sentics*, i.e., the denotative and connotative information commonly associated with real-world objects, actions, events, and people. SenticNet steps away from blindly using keywords and word co-occurrence counts, and instead relies on the implicit meaning associated with commonsense concepts. Superior to purely syntactic techniques, SenticNet can detect subtly expressed sentiments by enabling the analysis of multiword expressions that do not explicitly convey emotion, but are instead related to concepts that do so. The main limitation of SenticNet is that it is unable to generalize instances of concepts, e.g., `eat_pasta` or `slurp_noodles`: unless there is an exact match, SenticNet 3 raises a not-found error.

In SenticNet 4, however, both verb and noun concepts are linked to primitives so that, for example, concepts such as `eat_pasta` or `slurp_noodles` are generalized as `INGEST_FOOD`. In this way, most concept inflections can be captured by the knowledge base: verb concepts like `eat`, `slurp`, `munch` are all represented by their conceptual primitive `INGEST` while noun concepts like `pasta`, `noodles`, `steak` are replaced with their ontological parent `FOOD`.





*Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2666–2677, Osaka, Japan, December 11-17 2016.

The idea behind this generalization is that there is a finite set of mental primitives for affect-bearing concepts and a finite set of principles of mental combination governing their interaction. Conceptual primitives are automatically discovered in SenticNet through the ensemble application of hierarchical clustering and dimensionality reduction.

The rest of the paper is organized as follows: Section 2 presents related work in the field of sentiment analysis; Section 3 proposes an excursus on conceptual primitives; Sections 4 and 5 describe in detail how noun concepts and verb concepts are generalized, respectively; Section 6 proposes experimental results on two different state-of-the-art datasets; finally, Section 7 provides concluding remarks.

## 2 Related Work

Sentiment analysis systems can be broadly categorized into knowledge-based or statistics-based systems (Cambria, 2016). While the use of knowledge bases was initially more popular for the identification of sentiment polarity in text, recently sentiment analysis researchers have been increasingly using statistics-based approaches, with a special focus on supervised statistical methods. Pang et al. (Pang et al., 2002) pioneered this trend by comparing the performance of different machine learning algorithms on a movie review dataset and obtained a 82% accuracy for polarity detection. A recent approach by Socher et al. (Socher et al., 2013) obtained a 85% accuracy on the same dataset using a recursive neural tensor network. Yu and Hatzivassiloglou (Yu and Hatzivassiloglou, 2003) used semantic orientation of words to identify polarity at sentence level. Melville et al. (Melville et al., 2009) developed a framework that exploits word-class association information for domain-dependent sentiment analysis.

More recent studies exploit microblogging text or Twitter-specific features such as emoticons, hashtags, URLs, @symbols, capitalizations, and elongations to enhance sentiment analysis of tweets. Tang et al. (Tang et al., 2014a) used a convolutional neural network (CNN) to obtain word embeddings for words frequently used in tweets. Dos Santos et al. (dos Santos and Gatti, 2014) also focused on deep CNN for sentiment detection in short texts. Recent approaches also focus on developing word embeddings based on sentiment corpora (Tang et al., 2014b). Such word vectors include more affective clues than regular word vectors and produce better results for tasks such as emotion recognition (Poria et al., 2016b) and aspect extraction (Poria et al., 2016a).

Statistical methods, however, are generally semantically weak (Cambria and White, 2014). This means that, with the exception of obvious affect keywords, other lexical or co-occurrence elements in a statistical model have little predictive value individually. As a result, statistical text classifiers only work with acceptable accuracy when given a sufficiently large text input. Hence, while these methods may be able to affectively classify user's text on the page or paragraph level, they do not work well on smaller text units such as sentences. Concept-level sentiment analysis, instead, focuses on a semantic analysis of text through the use of web ontologies or semantic networks, which allows for the aggregation of the conceptual and affective information associated with natural language opinions (Cambria and Hussain, 2015; Gezici et al., 2013; Araújo et al., 2014; Bravo-Marquez et al., 2014; Recupero et al., 2014).

By relying on large semantic knowledge bases, such approaches step away from the blind use of keywords and word co-occurrence counts, relying instead on the implicit features associated with natural language concepts. Unlike purely syntactic techniques, concept-based approaches are also able to detect sentiments 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 bag-of-words. In the latter, in fact, concepts like `pretty_ugly` or `sad_smile` would be split into two separate words, disrupting both semantics and sentics of the input sentence.

## 3 Conceptual Primitives

It is inherent to human nature to try to categorize things, events and people, finding patterns and forms they have in common. One of the most intuitive ways to relate two entities is through their similarity. According to Gestalt theory (Smith, 1988), similarity is one of six principles that guide human perception of the world.

Similarity is a quality that makes one thing or person like another and ‘similar’ means having characteristics in common. There are many ways in which objects can be perceived as similar, based on things like color, shape, size and texture. If we move away from mere visual stimuli, we can apply the same principles to define similarity between concepts based on shared semantic features. Previous versions of SenticNet exploited this principle to cluster natural language concepts sharing similar affective properties. Finding groups of similar concepts, however, does not ensure full coverage of all possible semantic inflections of multiword expressions.

In this work, we leverage on such similarities to deduce conceptual primitives that can better generalize SenticNet’s commonsense knowledge. This generalization is inspired by different theories on conceptual primitives, including Roger Schank’s conceptual dependency theory (Schank, 1972), Ray Jackendoff’s work on explanatory semantic representation (Jackendoff, 1976), and Anna Wierzbicka’s book on primes and universals (Wierzbicka, 1996), but also theoretical studies on knowledge representation (Minsky, 1975; Rumelhart and Ortony, 1977). All such theories claim that a decompositional method is necessary to explore conceptualization. In the same manner a physical scientist understands matter by breaking it down into progressively smaller parts, a scientific study of conceptualization proceeds by decomposing meaning into smaller parts. Clearly, this decomposition cannot go on forever: at some point we must find semantic atoms that cannot be further decomposed. This is the level of conceptual structure; mental representation that encodes basic understanding and commonsense by means of primitive conceptual elements out of which meanings are built.

In SenticNet, this ‘decomposition’ translates into the generalization of multiword expressions that convey a specific set of emotions and, hence, carry a particular polarity. The motivation behind this process of generalization is that there are countless ways to express the same concept in natural language and having a comprehensive list of all the possible concept inflections is almost impossible. While lexical inflections such as conjugation and declension can be solved with lemmatization, semantic inflections such as the use of synonyms or semantically-related concepts need to be tackled by analogical reasoning.

If multiword expressions like `attain_knowledge` and `acquire_know-how` are encountered in text, SenticNet 3 is unable to process them because there is no entry for such concepts in the knowledge base. SenticNet 3, however, does contain a multiword expression that is highly semantically related to those two concepts, that is, `acquire_knowledge`. By working at the primitive level, SenticNet 4 is able to bridge this semantic gap, as `attain_knowledge`, `acquire_know-how`, and `acquire_knowledge` are all represented by the same conceptual primitive: `GET_INFORMATION`.

By automatically inferring conceptual primitives for SenticNet concepts, we aim to broadly extend the coverage of the commonsense knowledge base and better perform sentiment analysis tasks such as polarity detection and emotion recognition from text. As shown in the next two sections, this is done via generalizing noun concepts by means of hierarchical clustering as well as discovering conceptual primitives for verb concepts by means of dimensionality reduction.

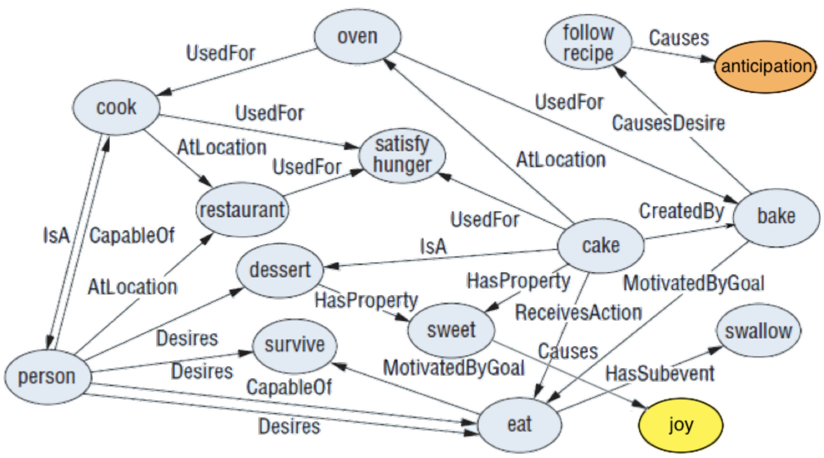



Figure 1: A sketch of the AffectNet graph showing part of the semantic network for the concept `cake`.

## 4 Noun Concept Generalization

The first step towards generalizing multiword expressions in SenticNet is to build a hierarchical classification of its noun concepts (or object concepts) so that nouns such as `cat`, `dog` or `pet` can be identified as `ANIMAL`. Such classification is implemented by applying hierarchical clustering on a semantic network of commonsense knowledge. It is important to note that each generalization inherits the emotional information and the polarity of its instance concepts. In the case of `cat` and `dog`, for example, the primitive is actually `ANIMAL+` since `cat` and `dog` are associated with positive emotions. Conversely, for animals that are associated with negative emotions such as fear (e.g., `white_shark`) or disgust (e.g., `cockroach`), the corresponding primitive is `ANIMAL-`.

### 4.1 AffectNet

AffectNet (Cambria and Hussain, 2015) is an affective commonsense knowledge base built upon ConceptNet (Speer and Havasi, 2012), the graph representation of the Open Mind corpus, and WordNet-Affect (Strapparava and Valitutti, 2004), a linguistic resource for the lexical representation of affect (Fig. 1). The resource is represented as a semantic network where nodes are multiword expressions of commonsense knowledge and the links between these are relations that interconnect them. The knowledge encoded by AffectNet is constantly expanding as new versions of ConceptNet are continuously released and new affective commonsense knowledge is crowdsourced through games. AffectNet is first converted into a matrix by dividing each assertion into two parts: a concept and a feature, where a feature is simply the assertion with the first or the second concept left unspecified such as 'a wheel is part of' or 'is a kind of liquid'.

The entries in the resulting matrix are positive or negative numbers, assigned according to the reliability of the assertions, with their magnitude increasing logarithmically with the confidence score. Because the AffectNet graph is made of triples based on the format <concept-relationship-concept>, the entire knowledge repository can be visualized as a large matrix, with every known concept of some statement being a row and every known semantic feature (relationship+concept) being a column. Such a representation has several advantages including the possibility to perform cumulative analogy (Tversky, 1977), executed by first selecting a set of nearest neighbors (in terms of similarity) of the input concept and then by projecting known properties of this set onto unknown properties of the concept (Table 1).

### 4.2 Group Average Agglomerative Clustering

Direct objects in verb+noun concepts, such as `buy_cake` or `eat_burger`, exhibit semantic coherence in that they tend to generate lexical items and phrases with related semantics. Most words related to the same verb tend to share some semantic characteristics. Our commonsense-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. In AffectNet, concepts inter-define one another, with directed edges indicating semantic dependencies between concepts.

| Concepts | Semantic Features (*relationship+concept*) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | .. | *Causes* joy | *IsA* event | *UsedFor* housekeeping | *LocatedAt* party_venue | *PartOf* celebration | *MotivatedByGoal* clean_room | .. |
| ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| `wedding` | .. | 0.94 | 0.86 | 0 | 0.79 | 0.88 | 0 | .. |
| `broom` | .. | 0 | 0 | 0.83 | 0 | 0 | 0.87 | .. |
| **`buy_cake`** | .. | **?** | **0.78** | **0** | **0.80** | **0.91** | **0** | .. |
| `birthday` | .. | 0.97 | 0.85 | 0 | 0.99 | 0.98 | 0 | .. |
| `sweep_floor` | .. | 0 | 0 | 0.79 | 0 | 0 | 0.91 | .. |
| ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

Table 1: Cumulative analogy allows for the inference of new pieces of knowledge by comparing similar concepts, e.g., `buy_cake` causes joy because `wedding` and `birthday` (which are similar) do so.

The traditional way to define features for any particular concept $c$ in a semantic network is to consider the set of concepts reachable via outbound edges from $c$. The proposed algorithm exploits hierarchical clustering to generate from such features conceptual primitives, which represent the core semantics of each concept. Based on experiments with various clustering algorithms, e.g., $k$-means and expectation-maximization clustering, we determined that group average agglomerative clustering (GAAC) provides the highest accuracy. 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, concepts are represented as vectors of commonsense features extracted from AffectNet. 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 and it is 0 in other situations. Cosine distance is used as the distance metric. Agglomerative algorithms are bottom-up in nature. GAAC, in particular, 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.

The resulting dendrogram 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:

$$X_{sum} = \sum_{c_m \in \omega_i \cup \omega_i} \quad \sum_{c_n \in \omega_i \cup \omega_j, c_n \neq c_m} \vec{c_n}.\vec{c_m} \tag{1}$$

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

where $\vec{c}$ is the vector of the concept of length $c$, vector entries are boolean (1 if the feature is present, 0 otherwise), and $N_i, N_j$ is the number of features in $\omega_i$ and $\omega_j$, respectively (which denote clusters). The main drawback of the hierarchical clustering algorithm is its running complexity (Berkhin, 2006), which averages $\theta(N^2 log\, N)$. We chose to utilize average link clustering as our clustering is connectivity-based. The concept proximity matrix consists of features from AffectNet and 'good' connections are deemed to occur when two concepts share multiple features. After clustering, the number of clusters is determined and the dendrogram is pruned accordingly.

Each cluster is later split into a positive and a negative sub-cluster. Cluster instances are assigned to either the positive or the negative sub-cluster depending on their polarity in SenticNet 3. For example, `cobra` and `cat` end up being in the same cluster (`ANIMAL`) after applying GAAC but, since they have opposite polarity in SenticNet 3, they are later assigned to different sub-clusters (`ANIMAL-` and `ANIMAL+`, respectively). Noun concepts for which no specific categorization is discovered are grouped under one of three most general noun primitives, namely: `SOMETHING`, `SOMEONE`, and `SOMEWHERE` (also divided into positive and negative sub-clusters). Table 2 provides an example of the results of polarity-driven feature-based clustering for 24 noun concepts.

| SOMETHING- ANIMAL- | SOMETHING+ ANIMAL+ | SOMEONE- PROFESSIONAL- | SOMEONE+ PROFESSIONAL+ | SOMEWHERE- NATURE- | SOMEWHERE+ NATURE+ |
|---|---|---|---|---|---|
| cockroach | horse | gravedigger | doctor | dry_steppe | oasis |
| rat | cat | coroner | scientist | desert | sandy_beach |
| cobra | puppy | executioner | teacher | wild_forest | natural_park |
| termite | pet | mortician | sea_captain | polar_desert | seaside |

Table 2: Example of feature-based clustering for polarity-driven conceptual primitive inference.

## 5 Verb Concept Generalization

The second step in generalizing SenticNet concepts is to define conceptual primitives for verb concepts (or action concepts) so that, for example, verbs like `acquire`, `attain` or `collect` can be identified as `GET`. Such classification is implemented by applying dimensionality reduction techniques on the vector space representation of AffectNet. As with noun concepts, verb concepts are also associated with a polarity but, in this case, polarity is more relevant to the opposite meanings (or outcomes) these action concepts represent, as in `INCREASE` versus `DECREASE`. This allows for reasoning about verb+noun combinations to be as per algebraic multiplication, where negative multiplied by positive (or vice versa) results in a negative, e.g., `DECREASE_GAIN` (or `INCREASE_LOSS`), multiplying two positives produces a positive, e.g., `INCREASE_PLEASURE`, and negative multiplied by negative results in a positive, e.g., `DECREASE_PAIN`.

### 5.1 AffectiveSpace

The human mind constructs intelligible meanings by continuously compressing over vital relations (Fauconnier and Turner, 2003). The compression principles aim to transform diffuse and distended conceptual structures into more focused versions so they can become more congenial for human understanding. In order to emulate such a process, we use simple but powerful meta-algorithms which underlie neuronal learning (Lee et al., 2011). These meta-algorithms should be fast, scalable, effective, with few-to-no specific assumptions and biologically plausible. Optimizing all the $\approx 10^{15}$connections formed through the last few million years of evolution is very unlikely. Objectively speaking, however, nature probably only optimizes the global connectivity (mainly the white matter) but leaves the other details to randomness (Balduzzi, 2013).

In this work, we use random projections (Bingham and Mannila, 2001) on the matrix representation of AffectNet in order to compress the semantic features associated with commonsense concepts and, hence, better perform analogical reasoning on these. Random projections are a data-oblivious method to map an original high-dimensional dataset into a much lower-dimensional subspace by using a Gaussian $\mathrm{N}(0,1)$ matrix, while at the same time, preserving pair-wise distances with high probability. This theoretically-solid and empirically-verified statement follows Johnson and Lindenstrauss's Lemma (Balduzzi, 2013), which states that, with high probability, for all pairs of points $\mathrm{x}, \mathrm{y} \in \mathrm{X}$ simultaneously:

$$\sqrt{\frac{m}{d}} \parallel x - y \parallel_2 (1-\varepsilon) \leq \parallel \Phi x - \Phi y \parallel_2 \leq \sqrt{\frac{m}{d}} \parallel x - y \parallel_2 (1+\varepsilon) \tag{3}$$

where $X$ is a set of vectors in Euclidean space, $d$ is the original dimension of this Euclidean space, $m$ is the dimension of the space we wish to reduce the data points to, $\varepsilon$ is a tolerance parameter measuring the maximum allowed distortion extent rate of the metric space, and $\Phi$ is a random matrix. Structured random projections for making matrix multiplication much faster was introduced in (Sarlos, 2006). When the number of features is much larger than the number of training samples ($d \gg n$), subsampled randomized Hadamard transform (SRHT) is preferred, as it behaves very much like Gaussian random matrices but accelerates the process from $\mathcal{O}(n\,d)$ to $\mathcal{O}(n\,log\,d)$ time (Lu et al., 2013). Following (Tropp, 2011; Lu et al., 2013), for $d = 2^p$ (where $p$ is any positive integer), a SRHT can be defined as:

$$\Phi = \sqrt{\frac{d}{m}}\mathrm{RHD} \tag{4}$$

where • $m$ is the number we want to subsample from $d$ features randomly;

• R is a random $m \times d$ matrix. The rows of R are $m$ uniform samples (without replacement) from the standard basis of $\mathbb{R}^d$;

• H$\in \mathbb{R}^{d\times d}$ is a normalized Walsh-Hadamard matrix, which is defined recursively:

$H_d = \begin{bmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & -H_{d/2} \end{bmatrix}$ with $H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$;

• D is a $d \times d$ diagonal matrix and the diagonal elements are i.i.d. Rademacher random variables.

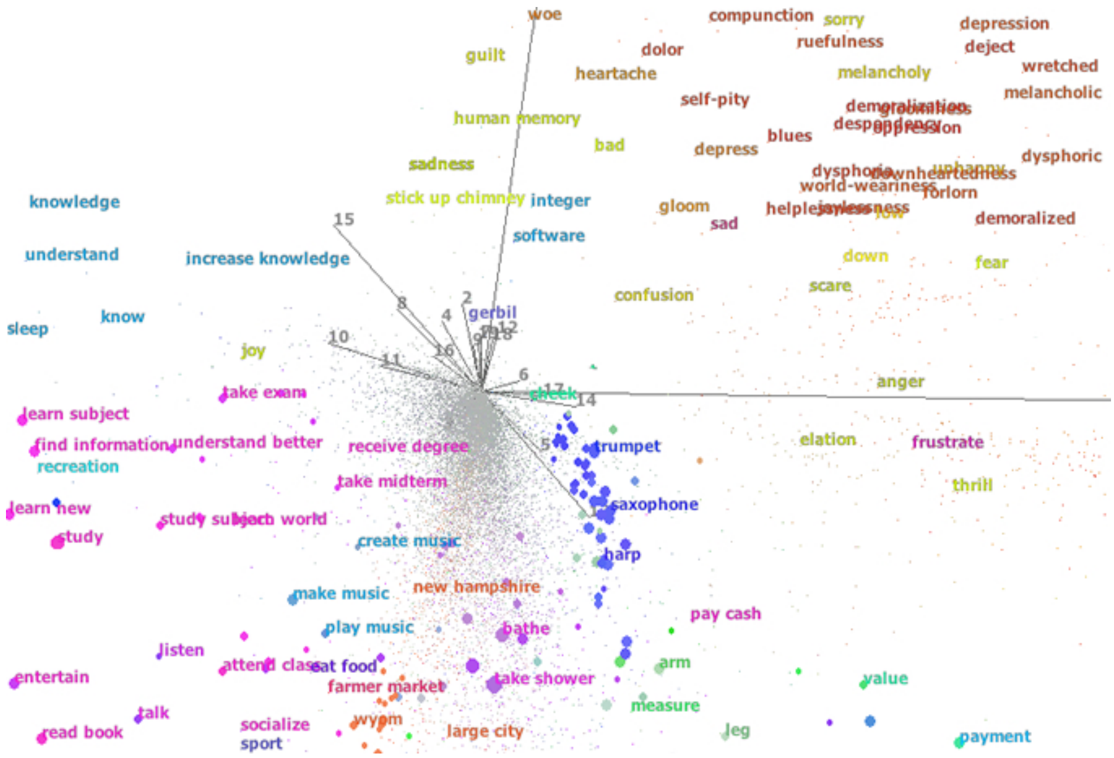



Figure 2: In AffectiveSpace, commonsense concepts gravitate around positive and negative emotions.

Our subsequent analysis only relies on the distances and angles between pairs of vectors (i.e., the Euclidean geometry information) and it is sufficient to set the projected space to be logarithmic in the size of the data (Ailon and Chazelle, 2010) and, hence, apply SRHT. The result is AffectiveSpace (Cambria et al., 2015), a vector space model where commonsense concepts and emotions are represented by vectors of $m$ coordinates (Fig. 2).

By exploiting the information sharing property of random projections, concepts with the same semantic and affective valence are likely to have similar features – that is, concepts conveying the same meaning and emotion tend to fall near each other in AffectiveSpace. Similarity does not depend on concepts' absolute position in the vector space, but rather on the angle these make with the origin. For example, concepts such as `birthday_party`, `celebrate`, and `buy_cake` are found very closely positioned in the vector space, while concepts like `lose_faith`, `depressed`, and `shed_tear` are found in a completely different direction (nearly opposite with respect to the centre of the space).

### 5.2 Semi-Supervised Verb Propagation

AffectiveSpace allows for analogical reasoning about multiword expressions so that concepts such as `buy_groceries` and `go_shopping` will be detected as being semantically similar. In order to generalize verb concepts, however, we need to discriminate such reasoning according to actions, so that a concept like `buy_groceries` would be associated with concepts related to the verb `buy`, e.g., `buy_milk` or `purchase_vegetable`.

To this end, we leverage on VerbNet (Schuler, 2005), the largest English verb lexicon currently available, and Sentic LDA (Poria et al., 2016c), a classification framework that integrates commonsense in the calculation of word distributions in the linear discriminant analysis (LDA) algorithm. In particular, we use a semi-supervised version of Sentic LDA in order to incorporate both supervised (VerbNet-labeled) and unsupervised information in such a way that a proper semantic space which reflects the desired information (verb concepts) is obtained. Given a set of verbs and a large amount of unlabeled instances in AffectiveSpace, the between-class scatter is to be maximized and the within-class scatter of VerbNet instances is to be minimized, keeping the semantic relatedness of all the other instances simultaneously.

Each instance is denoted as $v_i \in \mathcal{A}^m$, which is the $m$-dimensional vector after being processed by random projections. For each verb instance, there is a label $y_i \in \{1, \dots, q\}$, where $q$ is the number of verb classes. Then, the between-class scatter and the within-class scatter matrices are defined as follows:

$$S_w = \sum_{j=1}^{q} \sum_{i=1}^{l_j} (v_i - \mu_j)(v_i - \mu_j)^T \tag{5}$$

$$S_b = \sum_{j=1}^{q} l_j(\mu_j - \mu)(\mu_j - \mu)^T \tag{6}$$

where $\mu_j = \frac{1}{l_j} \sum_{i=1}^{l_j} v_i \ (j = 1, 2, ..., q)$ is the mean of the samples in class $j$, $l_j$ is the number of verb instances in class $j$ and $\mu = \frac{1}{l} \sum_{i=1}^{l} v_i$ is the mean of all the labeled samples. A total scatter matrix on all the instances in AffectiveSpace is also defined:

$$S_t = \sum_{i=1}^{k} (v_i - \mu_k)(v_i - \mu_k)^T \tag{7}$$

where $k$ is the total number of instances in AffectiveSpace and $\mu_k$ is the mean of all the instances. Our objective is then to find a projection matrix $W$ to project the semantic space to a lower-dimensional space, which is more discriminative towards verb concepts:

$$W^* = \arg \max_{W \in \mathcal{A}^{m \times m'}} \frac{|W^T S_b W|}{|W^T (S_w + \lambda_1 S_t + \lambda_2 I) W|} \tag{8}$$

where $I$ is identity matrix, and $\lambda_1$ and $\lambda_2$ are control parameters, obtained through a grid search, for balancing the trade-off between verb discriminant and semantic regularizations. The optimal solution is given by:

$$(S_w + \lambda_1 S_t + \lambda_2 I) w_j^* = \eta_j S_b w_j^* \quad j = 1, ..., m' \tag{9}$$

where $w_j^*$ $(j = 1, ..., m')$ are the eigenvectors corresponding to the $m'$ largest eigenvalues of $(S_w + \lambda_1 S_t + \lambda_2 I)^{-1} S_b$. Here, $m' = q - 1$ is selected, where $q$ is the total verb primitive number. After the projection, the new space preserves both semantic relatedness and action concept grouping based on the information coming from AffectNet and VerbNet, respectively.

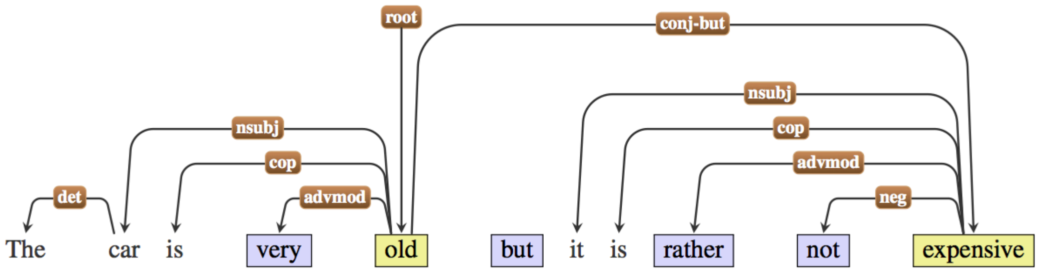
## 6 Evaluation

In order to perform a qualitative evaluation of SenticNet 4 (available both as a standalone XML repository[1] and as an API[2]), we asked five annotators to judge the plausibility of inferred conceptual primitives. We obtained an overall accuracy of 91% with Cohen's kappa score of 0.84. As for the quantitative evaluation, we tested SenticNet 4 against two well-known sentiment resources, namely: the Blitzer Dataset (Blitzer et al., 2007) and the Movie Review Dataset (Pang and Lee, 2005).

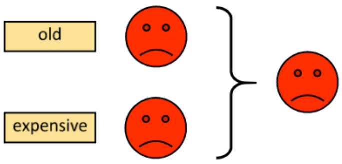
### 6.1 Performing Polarity Detection with SenticNet

While SenticNet can be used as any other sentiment lexicon, e.g., concept matching or bag-of-concepts model, the right way to use the knowledge base for the task of polarity detection is in conjunction with sentic patterns (Poria et al., 2015). Sentic patterns are sentiment-specific linguistic patterns that infer polarity by allowing affective information to flow from concept to concept based on the dependency relation between clauses. The main idea behind such patterns can be best illustrated by analogy with an electronic circuit, in which few 'elements' are 'sources' of the charge or signal, while many elements operate on the signal by transforming it or combining different signals. This implements a rudimentary type of semantic processing, where the 'meaning' of a sentence is reduced to only one value: its polarity.

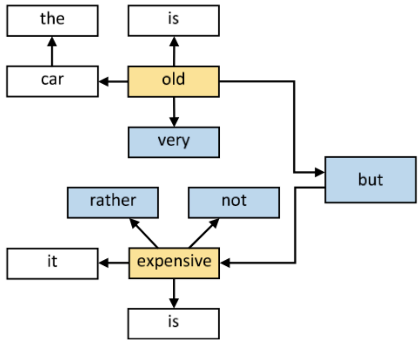
[1] `http://sentic.net/senticnet-4.0.zip`
[2] `http://sentic.net/api`



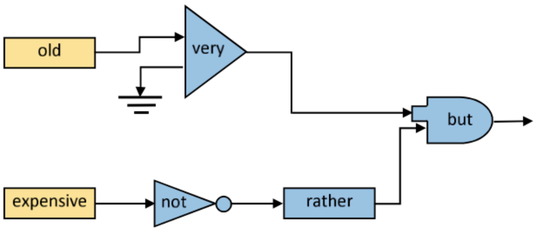

(a) Dependency tree of a sentence.



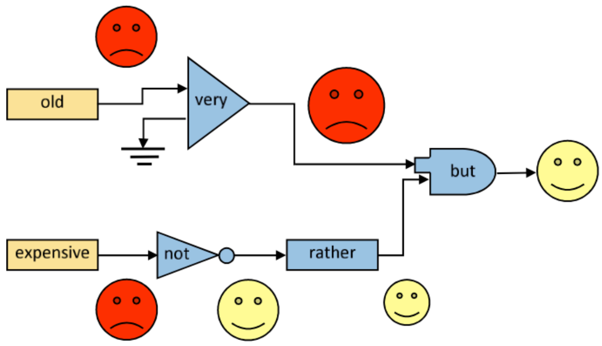

(b) The old way: averaging over a bag of sentiment words. The overall polarity of a sentence is given by the algebraic sum of the polarity values associated with each affect word divided by the total number of words.




(c) The dependency tree of a sentence resembles an electronic circuit: words shown in blue can be thought as a sort of "boolean operations" acting on other words.




(d) The electronic circuit metaphor: sentiment words are "sources" while other words are "elements", e.g., *very* is an amplifier, *not* is a logical complement, *rather* is a resistor, *but* is an OR-like element that gives preference to one of its inputs.




(e) The final sentiment data flow of the "signal" in the "circuit".

Figure 3: In sentic patterns, the structure of a sentence is like an electronic circuit where logical operators channel sentiment data-flows to output an overall polarity.

Sentic patterns are applied to the dependency syntactic tree of a sentence, as shown in Figure 3(a). The only two words that have intrinsic polarity are shown in yellow color; the words that modify the meaning of other words in the manner similar to contextual valence shifters (Polanyi and Zaenen, 2006) are shown in blue. A baseline that completely ignores sentence structure, as well as words that have no intrinsic polarity, is shown in Figure 3(b): the only two words left are negative and, hence, the total polarity is negative. However, the syntactic tree can be re-interpreted in the form of a 'circuit' where the 'signal' flows from one element (or subtree) to another, as shown in Figure 3(c). After removing the words not used for polarity calculation (in white), a circuit with elements resembling electronic amplifiers, logical complements, and resistors is obtained, as shown in Figure 3(d),

Figure 3(e) illustrates the idea at work: the sentiment flows from polarity words through shifters and combining words. The two polarity-bearing words in this example are negative. The negative effect of the word 'old' is amplified by the intensifier 'very'. However, the negative effect of the word 'expensive' is inverted by the negation, and the resulting positive value is decreased by the 'resistor'. Finally, the values of the two phrases are combined by the conjunction 'but', so that the overall polarity has the same sign as that of the second component (positive).

### 6.2 SenticNet 4 vs. SenticNet 3

The Blitzer Dataset consists of product reviews in seven different domains. For each domain there are 1,000 positive and 1,000 negative reviews. In evaluating SenticNet 4, we only used reviews under the *electronics* category. From these, we randomly extracted 7,210 non-neutral sentences: 3,800 of which were marked as positive and 3,410 as negative. We then compared the performance of SenticNet 4 with its predecessor SenticNet 3 for the task of sentence-level polarity detection, using sentic patterns. The results are shown in Table 3.

Table 3: Comparison on the Blitzer Dataset

| Framework | Accuracy |
|---|---|
| Sentic Patterns and SenticNet 3 | 87.0% |
| Sentic Patterns and SenticNet 4 | 91.3% |

### 6.3 SenticNet 4 vs. Statistical Methods

The Movie Review Dataset includes 1,000 positive and 1,000 negative movie reviews collected from Rotten Tomatoes[3]. Originally, Pang and Lee manually labeled each review as positive or negative. Later, Socher et al. (Socher et al., 2012; Socher et al., 2013) annotated this dataset at sentence level. They extracted 11,855 sentences from the reviews and manually labeled them using a fine-grained inventory of five sentiment labels: *strong positive*, *positive*, *neutral*, *negative*, and *strong negative*. Since in this work we consider only binary classification, we removed neutral sentences from the dataset and merged germane labels. Thus, the final dataset contained 4,800 positive sentences and 4,813 negative ones. The results of the classification with SenticNet 3 and SenticNet 4 are shown in Table 4.

Table 4: Comparison on the Movie Review Dataset

| Framework | Accuracy |
|---|---|
| Socher et al., 2012 | 80.0% |
| Socher et al., 2013 | 85.4% |
| Sentic Patterns and SenticNet 3 | 86.2% |
| Sentic Patterns and SenticNet 4 | 90.1% |

## 7 Conclusion

The distillation of knowledge from the huge amount of unstructured information on the Web is a key factor for tasks such as social media marketing, brand positioning, and financial prediction. Commonsense reasoning is a good solution for sentiment analysis but the scalability of commonsense knowledge bases is a major factor that jeopardizises the efficiency of concept extraction and polarity detection. A first possible step in solving this problem is to generalize pieces of commonsense knowledge in terms of conceptual primitives that could catch most semantic inflections of natural language concepts.

In this work, we used an ensemble of hierarchical clustering and dimensionality reduction for automatically discovering the primitives for both noun and verb concepts in SenticNet. This generalization process allowed us to largely extend the coverage of the commonsense knowledge base and, hence, to boost the accuracy of SenticNet for sentence-level polarity detection in comparison with both the previous version of the resource and with state-of-the-art statistical sentiment analysis research.

In the future, we plan to discover new conceptual primitives in a more automatic and scalable way by means of dependency-based word embeddings. In particular, we will exploit the internally-learned context embeddings of the skip-gram model in conjunction with the standard target word embeddings, to weigh context compatibility together with word similarity.

[3] `http://rottentomatoes.com`

## References


Nir Ailon and Bernard Chazelle. 2010. Faster dimension reduction. *Communications of the ACM*, 53(2):97–104.

Matheus Araújo, Pollyanna Gonçalves, Meeyoung Cha, and Fabrício Benevenuto. 2014. iFeel: A system that compares and combines sentiment analysis methods. In *WWW*, pages 75–78.

David Balduzzi. 2013. Randomized co-training: from cortical neurons to machine learning and back again. *arXiv preprint arXiv:1310.6536*.

Pavel Berkhin. 2006. A survey of clustering data mining techniques. *Grouping multidimensional data*, pages 25–71.

Ella Bingham and Heikki Mannila. 2001. Random projection in dimensionality reduction: applications to image and text data. In *ACM SIGKDD*, pages 245–250.

John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447.

Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. 2014. Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, 69:86–99.

Erik Cambria and Amir Hussain. 2015. *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Springer, Cham, Switzerland.

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

Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. 2014. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In *AAAI*, pages 1515–1521, Quebec City.

Erik Cambria, Jie Fu, Federica Bisio, and Soujanya Poria. 2015. AffectiveSpace 2: Enabling affective intuition for concept-level sentiment analysis. In *AAAI*, pages 508–514, Austin.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Cıcero Nogueira dos Santos and Maıra Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69–78.

Gilles Fauconnier and Mark Turner. 2003. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Gizem Gezici, Rahim Dehkharghani, Berrin Yanikoglu, Dilek Tapucu, and Yucel Saygin. 2013. Su-sentilab: A classification system for sentiment analysis in twitter. In *International Workshop on Semantic Evaluation*, pages 471–477.

Ray Jackendoff. 1976. Toward an explanatory semantic representation. *Linguistic Inquiry*, 7(1):89–150.

Anil Jain and Richard Dubes. 1988. *Algorithms for clustering data*. Prentice-Hall, Inc.

Honglak Lee, Roger Grosse, Rajesh Ranganath, and A. Y. Ng. 2011. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10):95–103.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan and Claypool.

Yichao Lu, Paramveer Dhillon, Dean P Foster, and Lyle Ungar. 2013. Faster ridge regression via the subsampled randomized hadamard transform. In *Advances in Neural Information Processing Systems*, pages 369–377.

Prem Melville, Wojciech Gryc, and Richard D Lawrence. 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *ACM SIGKDD*, pages 1275–1284.

Marvin Minsky. 1975. A framework for representing knowledge. In Patrick Winston, editor, *The psychology of computer vision*. McGraw-Hill, New York.

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL*, pages 115–124, Ann Arbor.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–135.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *EMNLP*, volume 10, pages 79–86.

Livia Polanyi and Annie Zaenen. 2006. Contextual valence shifters,. In *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series*, pages 1–10. Springer, Berlin, Germany.

Soujanya Poria, Erik Cambria, Alexander Gelbukh, Federica Bisio, and Amir Hussain. 2015. Sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine*, 10(4):26–36.

Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016a. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.

Soujanya Poria, Erik Cambria, Newton Howard, Guang-Bin Huang, and Amir Hussain. 2016b. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, 174:50–59.

Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Federica Bisio. 2016c. Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In *IJCNN*.

Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Nuzzolese. 2014. Sentilo: Frame-based sentiment analysis. *Cognitive Computation*, 7(2):211–225.

David Rumelhart and Andrew Ortony. 1977. The representation of knowledge in memory. In C Anderson, R Spiro, and W Montague, editors, *Schooling and the acquisition of knowledge*. Erlbaum, Hillsdale, NJ.

Tamas Sarlos. 2006. Improved approximation algorithms for large matrices via random projections. In *FOCS*, pages 143–152.

Roger Schank. 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3:552–631.

Karin Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania, Computer and Information Science.

Barry Smith, editor. 1988. *Foundations of Gestalt Theory*. Munich and Vienna: Philosophia Verlag.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *EMNLP*, pages 1201–1211.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, pages 1642–1654.

Robert Speer and Catherine Havasi. 2012. ConceptNet 5: A large semantic network for relational knowledge. In Eduard Hovy, M Johnson, and G Hirst, editors, *Theory and Applications of Natural Language Processing*, chapter 6. Springer.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: An affective extension of WordNet. In *LREC*, pages 1083–1086, Lisbon.

Duyu Tang, Furu Wei, Bing Qin, Ting Liu, and Ming Zhou. 2014a. Coooolll: A deep learning system for twitter sentiment classification. In *SemEval*, pages 208–212.

Duyu Tang, Furu Wei, Nan Yang, Ting Liu, and Bing Qin. 2014b. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL*, pages 1555–1565.

Joel A Tropp. 2011. Improved analysis of the subsampled randomized hadamard transform. *Advances in Adaptive Data Analysis*, 3(01n02):115–126.

Amos Tversky. 1977. Features of similarity. *Psychological Review*, 84(4):327–352.

Anna Wierzbicka. 1996. *Semantics: Primes and Universals*. Oxford University Press.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP*, pages 129–136.