Semantic Multidimensional Scaling for Open-Domain Sentiment Analysis

Erik Cambria, *Nanyang Technological University*

Yangqiu Song, *University of Illinois at Urbana-Champaign*

Haixun Wang, *Microsoft Research Asia*

Newton Howard, *Massachusetts Institute of Technology*

The largest existing taxonomy of common knowledge is blended with a natural-language-based semantic network of commonsense knowledge. Multidimensional scaling is applied on the resulting knowledge base for open-domain opinion mining and sentiment analysis.

The ever-growing amount of available information in the Social Web fosters the proliferation of business and research activities around the relatively new fields of opinion mining and sentiment analysis. The automatic analysis of user-generated content such as online news, reviews, blogs, and tweets, in fact, can be extremely valuable for tasks such as mass opinion estimation, corporate reputation measurement, political orientation categorization, stock market prediction, customer preference, and public opinion study. Distilling useful information from such unstructured data, however, is a multifaceted and multidisciplinary problem, as opinions and sentiments can be expressed in a variety of forms and combinations in which it's extremely difficult to find any type of regular behavior. A lot of conceptual rules, in fact, govern the expression of opinions and sentiments and there exist even more clues that can convey these concepts from realization to verbalization in the human mind.

Most current approaches to opinion mining and sentiment analysis rely on rather unambiguous affective keywords extracted from an existing knowledge base, for example, WordNet,[1] or from a purpose-built lexicon based on a domain-dependent corpus.[2–4] (See the related sidebar for further details.) Such approaches are far from being able to perfectly extract the conceptual and affective information associated with natural language and, hence, often fail to meet the gold standard of human annotators. Especially when dealing with social media, in fact, content is often diverse and noisy, and the use of a limited number of affect words or a domain-dependent

1541-1672/14/$31.00 © 2014 IEEE
Published by the IEEE Computer Society

Related Work in Opinion Mining

Early works in the field of opinion mining and sentiment analysis aimed to classify entire documents as containing overall positive or negative polarity[1] or rating scores (for example, 1 to 5 stars) of reviews.[2] These were mainly supervised approaches relying on manually labeled samples such as movie or product reviews, where the commentator's overall positive or negative attitude was explicitly indicated. However, opinions and sentiments don't occur only at the document level, nor are they limited to a single valence or target. Contrary or complementary attitudes toward the same or multiple topics can be present across the span of a document.

Later works adopted a segment- or paragraph-level opinion analysis aiming to distinguish sentimental from nonsentimental sections, for example, by performing a classification based on some fixed syntactic phrases likely to be used to express opinions[3] or by bootstrapping using a small set of seed opinion words and a knowledge base such as WordNet.[4] Other works have taken down text analysis granularity to the sentence level, for example, by using the presence of opinion-bearing lexical items (single words or *n*-grams) to detect subjective sentences[5] or by using semantic frames for identifying the sentiment topics (or targets).[6]

The aim of our work (described in the article's main text) is to build the most comprehensive resource of common and common-sense knowledge and apply multidimensional scaling (MDS) to perform a domain-independent, concept-level analysis of opinion and sentiments on the Web. Charles Osgood and his colleagues conducted pioneering work on understanding and visualizing the affective information associated with natural language text through MDS.[7] Osgood used MDS to create visualizations of affective words based on the words' similarity ratings provided to subjects from different cultures. In Osgood's work, words can be thought of as points in a multidimensional space and the similarity ratings represent the distances between these words. MDS projects such distances to points in a smaller dimensional space.

In our work, similarly, we apply MDS to a common and common-sense knowledge base to grasp the semantic and affective similarity between different concepts after plotting them into a multidimensional vector space. Differently from Osgood's space, however, the building blocks of our vector space aren't simply a limited set of similarity ratings between affect words, but rather millions of confidence scores related to pieces of common-sense knowledge linked to a hierarchy of affective domain labels. Rather than merely determined by a few human annotators and represented as a word–word matrix, in fact, our vector space is built upon a common-sense knowledge base represented as a concept-feature matrix.

References

1. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," *Proc. Conf. Empirical Methods on Natural Language Processing*, 2002, pp. 79–86.
2. B. Pang and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales," *Proc. 43rd Ann. Mtg. Assoc. Computational Linguistics*, 2005, pp. 115–124.
3. P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Ann. Mtg. Assoc. Computational Linguistics*, 2002, pp. 417–424.
4. J. Kamps et al., "Using WordNet to Measure Semantic Orientation of Adjectives," *Proc. Int'l Conf. Language Resources and Evaluation*, 2004, pp. 1115–1118.
5. E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," *Proc. Conf. Empirical Methods on Natural Language Processing*, 2003, pp. 105–112.
6. S. Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," *Proc. Workshop Sentiment and Subjectivity in Text*, 2006.
7. C. Osgood, W. May, and M. Miron, *Cross-Cultural Universals of Affective Meaning*, Univ. of Illinois Press, 1975.

training corpus is simply not enough. To intelligently process open-domain textual resources, computers need to be provided with both the common and common-sense knowledge humans normally acquire during the formative years of their lives, because relying just on the valence of keywords and word co-occurrence frequencies doesn't allow a deep understanding of natural language.

Here, we blend ProBase,[5] the largest existing taxonomy of common knowledge, with ConceptNet,[6] a natural-language-based semantic network of common-sense knowledge. We apply multidimensional scaling (MDS) on the resulting knowledge base for sentiment analysis.

Common and Common Sense

In standard human-to-human communication, people usually refer to existing facts and circumstances and use this knowledge to build new useful, funny, or interesting information. This common knowledge encompasses information usually found in news, articles, debates, lectures, and so on (that is, factual knowledge), but also includes principles and definitions found in collective intelligence projects such as Wikipedia (that is, vocabulary knowledge). Moreover, when people communicate with each other, they rely on similar background knowledge, for example, 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.

Common Knowledge Base

Attempts to build a common knowledge base are countless and include human expert or community effort crafted resources such as WordNet, with its 25,000 synsets, or Freebase,[7] a social database of 1,450 concepts; automatically built knowledge bases such as YAGO,[8] a semantic database with 149,162 instances derived from Wikipedia, WordNet, and GeoNames (see http://geonames.org); and ProBase.

ProBase contains approximately 12 million concepts learned iteratively from 1.68 billion webpages in the Bing repository. The taxonomy is probabilistic, which means every claim in ProBase is associated with some probabilities that model the claim's correctness, ambiguity, and other characteristics. The probabilities are derived from evidence found in Web, search log, and other available data. The core taxonomy consists of the IsA relationships extracted by using syntactic patterns. For example, the segment "artists such as Pablo Picasso" can be considered a piece of evidence for the claim that "pablo picasso" is an instance of the concept "artist."

Common-Sense Knowledge Base

One of the biggest projects aiming to build a comprehensive common-sense knowledge base is Cyc.[9] Cyc, however, requires the involvement of experts working on specific languages and contains just 120,000 concepts, as the knowledge engineering is labor-intensive and time-consuming. A more recent and scalable project is Open Mind Common Sense (OMCS), which has been collecting pieces of knowledge from volunteers on the Internet since 1999 by enabling the general public to enter common sense into the system. OMCS exploits these pieces of common-sense knowledge to automatically build ConceptNet, a semantic network of 173,398 nodes upon which many other common-sense resources, for example, SenticNet (see http://sentic.net), are built.

WordNet contains detailed descriptions of every word's various senses, but it doesn't include enough general Web information. ProBase, which provides more concepts, includes pieces of knowledge that match the general distribution of human knowledge. ConceptNet, in turn, contains implicit knowledge that people rarely mention on the Web, which acts as complementary material to Probase.

Building the Knowledge Base

In other work, Probase IsA relationships were exploited to build a semantic network, termed Isanette (IsA net), representing hyponym–hypernym common knowledge as a matrix having instances (for example, "pablo picasso") as rows and concepts (for example, "artist") as columns.[10] In this work, we use an extended version of Probase and the new Isanette matrix is 4,622,119 × 2,524,453. Because Isanette is a large and fat matrix that contains noise and multiple forms, we first clean it by applying different natural language processing and MDS techniques. Second, we enhance Isanette's consistency (and further reduce its sparseness) by adding complementary common-sense knowledge.

Cleaning Isanette

Isanette is built out of approximately 40 million IsA triples extracted with the form <instance, concept, confidence score>. Before generating the matrix from these statements, however, we need to solve two main issues: multiple concept forms and low connectivity.

The first issue is addressed by exploiting both word similarity and MDS. The concept "barack obama," for example, appears in the triples in many different forms such as "president obama," "mr barack obama," "president barack obama," and so on. Trying to disambiguate these types of instances a priori by simply using word similarity could be dangerous because, for example, concepts like "buy christmas present" and "present christmas event" have different meanings although they have high word similarity. Hence, we perform an a posteriori concept deduplication by exploiting concept semantic relatedness after Isanette is built. That is, we merge concepts with high word similarity if they're close enough to each other in the vector space generated from Isanette.

The second issue is addressed by discarding hapax legomena, that is, instances and concepts with a singular out- and in-degree. If we are to apply MDS to find similar patterns, in fact, Isanette needs to be as populated as possible. In this work, we not only discard hapax legomena but other long-tail concepts to heavily enhance Isanette's graph connectivity. In particular, we used a trial-and-error approach to find that the best tradeoff between size and sparseness is achieved by setting the minimum node connectivity equal to 10. This cut-off operation leaves out almost 40 percent of nodes and makes Isanette a strongly connected core. Moreover, MDS is exploited to infer negative evidence, such as "carbonara" is not a kind of "fuel" or "alitalia" is not a "country," which is useful to further reduce Isanette's sparseness and improve reasoning algorithms.

Blending Isanette

As a subsumption common knowledge base, Isanette lacks information like a "dog" is a "best friend" (rather than simply an "animal") or a "rose" is a kind of "meaningful gift" (rather than simply a kind of "flower"), that is, common sense that isn't usually stated in webpages (or at least not that often to be extracted by Hearst patterns with a high-enough confidence score). To overcome this problem, we enriched Isanette with complementary hyponym–hypernym common-sense knowledge from ConceptNet. In particular, all the assertions involving IsA relationships with a non-null confidence score, such as "dog is man's best friend" or "a birthday party is a special occasion," are extracted from the OMCS corpus. Such assertions are exploited to generate a directed graph of about 15,000 nodes (interconnected by IsA edges), representing subsumption common-sense knowledge.

To merge this subsumption commonsense knowledge base with Isanette, we employ the blending technique.[11] Blending is a technique that performs inference over multiple sources of data simultaneously, taking advantage of the overlap between them. The approach combines two sparse matrices linearly into a single matrix in which the information between the two initial sources is shared. This alignment operation yields IsaCore (see http://sentic.net/isacore.zip), a strongly-connected core (hereafter referred as *C*, for the sake of simplicity) in which common and common-sense knowledge coexist, that is, a matrix of 500,000 × 300,000 whose rows are instances (for example, "birthday party" and "china"), whose columns are concepts (for example, "special occasion" and "country"), and whose values indicate truth values of assertions.

Reasoning on the Knowledge Base

In this section, we apply MDS to build a vector-space representation of the instance–concept relationship matrix. We then employ a semisupervised learning algorithm to further discriminate affective information, and use a partitioning clustering technique to segment the reduced space into conceptual classes.

Semantic Multidimensional Scaling

To more compactly represent the information contained in $C \in \mathcal{R}^{m \times n}$ and encode the latent semantics between its instances, we build a multidimensional vector space representation by applying truncated singular value decomposition (SVD). The resulting lower-dimensional space represents the best approximation of *C*, in fact

$$\min_{\tilde{C}|rank(\tilde{C})=d} |C-\tilde{C}| = \min_{\tilde{C}|rank(\tilde{C})=d} |\Sigma - U_d^T \tilde{C} V_d| = \min_{\tilde{C}|rank(\tilde{C})=d} |\Sigma - S_d|,$$

where *C* has the form $C = U\Sigma V^T$, $\tilde{C}$ has the form $\tilde{C} = U_d S_d V_d^T$ ($U_d \in \mathcal{R}^{m \times d}$, $V_d \in \mathcal{R}^{n \times d}$, and $S_d \in \mathcal{R}^{d \times d}$ is the diagonal matrix), and *d* is the lower dimension of the latent semantic space. From the rank constraint, that is, S_d has *d* nonzero diagonal entries, the minimum of the previous statement is obtained as follows:

$$\min_{\tilde{C}|rank(\tilde{C})=d} |\Sigma - S_d| = \min_{s_i} \sqrt{\sum_{i=1}^{n} (\sigma_i - s_i)^2} = \min_{s_i} \sqrt{\sum_{i=1}^{d} (\sigma_i - s_i)^2 + \sum_{i=d+1}^{n} \sigma_i^2} = \sqrt{\sum_{i=d+1}^{n} \sigma_i^2}.$$

Therefore, $\tilde{C}$ of rank *d* is the best approximation of *C* in the Frobenius-norm sense when $\sigma_i = s_i$ ($i = 1, \ldots, d$) and the corresponding singular vectors are the same as those of *C*. If all but the first *d* principal components are discarded and $\tilde{C}_U = U_d S_d$ is considered, we obtain a space in which common and common-sense instances are represented by vectors of *d* coordinates.

These coordinates can be seen as describing instances in terms of eigenconcepts that form the axes of the vector space, that is, its basis $e = (e^{(1)}, \ldots, e^{(d)})^T$. We use a trial-and-error approach and find that the best compromise is achieved when *d* assumes values around 500. Such a 500-dimensional vector space can be used for making analogies (given a specific instance, find the instances most semantically related to it), for making comparisons (given two instances, infer their degree of semantic relatedness), and for classification purposes (given a specific instance, assign it to a predefined cluster).

Semisupervised Affective Propagation

After applying SVD, the obtained $\tilde{C}_U$ doesn't lead to meaningful affective relatedness results, as the vector space represents semantic relatedness of instances according to IsA relationships. Affectively opposite instances such as "smile" and "cry," in fact, are likely to be found close to each other in $\tilde{C}_U$ because they both relate to emotions. Hence, to build an appropriate space that is both semantic and sentiment preserving, we adopt a semisupervised linear discriminant analysis (LDA) algorithm.

Differently from other classifiers, semisupervised LDA can incorporate both supervised (affective keywords) and unsupervised (nonaffective keywords) information in such a way that a proper semantic space that reflects the sentiment information is obtained. In this work, we prefer LDA to other classifiers for its analytical simplicity and computational efficiency. More in-depth motivations for the choice of LDA can be found in other work.[12]

To infer affective information from natural language and use it for tasks such as emotional labelling and opinion polarity detection, existing approaches rely on a relatively small set of affective words extracted from manually-labelled lexicons, for example, WordNet, and a few emotional labels, for example, Ekman's six universal emotions. Because such categorical approaches classify emotions using a list of labels, they usually fail to describe the complex range of emotions that can occur in daily communication.

To overcome this problem, we employ the Hourglass of Emotions[13] categorization model (see Figure 1). Because such a biologically-inspired and psychologically-motivated model goes beyond mere categorical and dimensional approaches, it can potentially describe any human emotion in terms of four independent but concomitant dimensions.

Given a set of affective labels and a large amount of unlabeled instances in *C*, the between-class scatter is to be

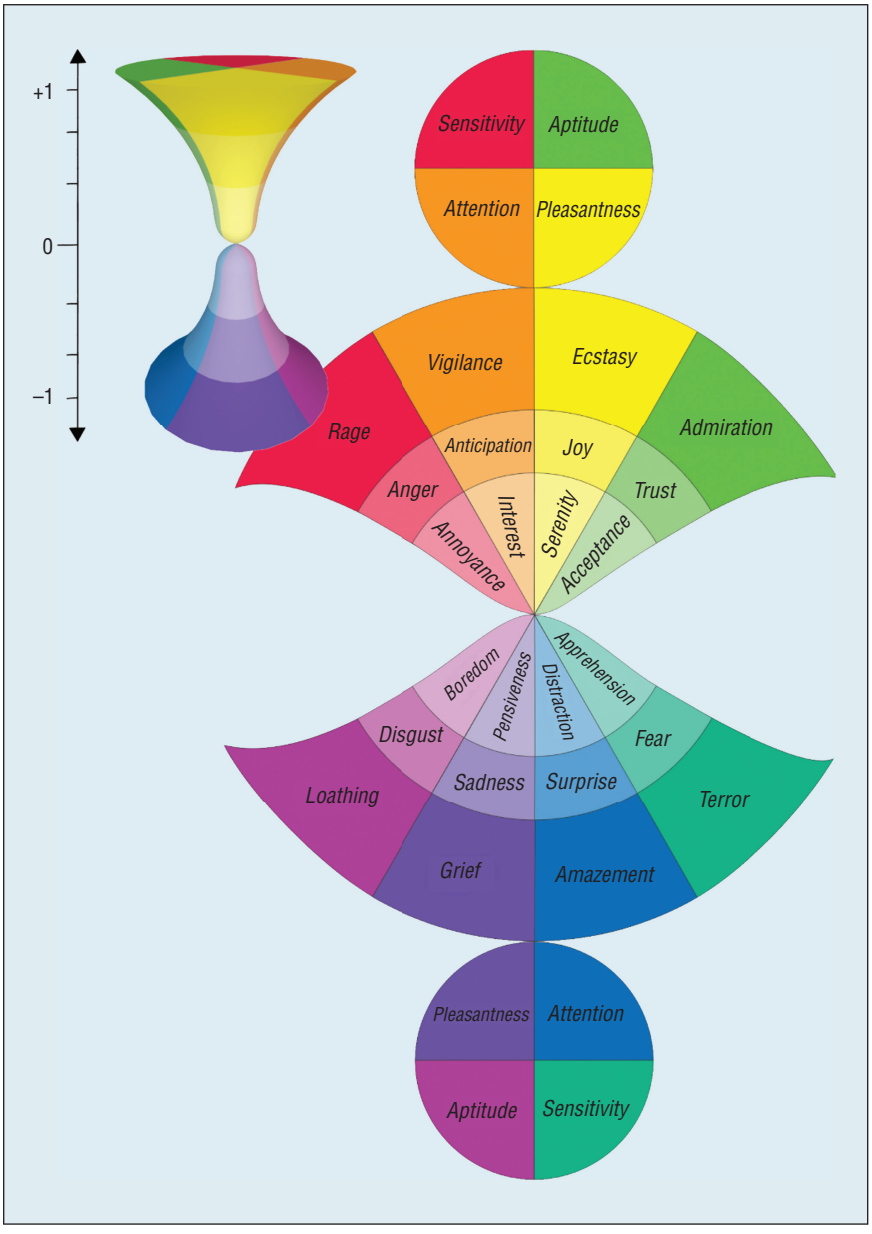

Figure 1. Hourglass of emotions.

maximized and the within-class scatter of expressly affective instances (from the Hourglass model) is to be minimized, as well as the semantic relatedness of all the other instances simultaneously is to be kept.

We denote each instance as $e_i \in \mathcal{R}^d$, which is a d-dimensional vector after processing with SVD. For each expressly affective instance, there is a label $y_i \in \{1, \ldots, q\}$, where q is the number of sentiment classes. Then, the between-class scatter and the within-class scatter matrices are defined as follows:

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

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

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

$$S_t = \sum_{i=1}^{m}(e_i - \mu_m)(e_i - \mu_m)^T,$$

where m is the total number of instances in C and μ_μ 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 affectively discriminative:

$$W^* = \arg \max_{W \in \mathcal{R}^{d \times d'}} \times \frac{|W^T S_b W|}{|W^T(S_w + \lambda_1 S_t + \lambda_2 I)W|},$$

where I is identity matrix, and λ_1 and λ_2 are control parameters obtained through a grid search, which balance the tradeoff between sentiment discriminant and semantic regularizations. The optimal solution is given by

$$(S_w + \lambda_1 S_t + \lambda_2 I)w_j^* = \eta_j S_b w_j^*$$
$$j = 1, \ldots, d',$$

where $w_j^*(j = 1, \ldots, d')$ are the eigenvectors corresponding to the d' largest eigenvalues of $(S_w + \lambda_1 S_t + \lambda_2 I)^{-1} S_b$. Here $d' = q - 1$ is selected, where q is the total emotion number. After the projection, the new space preserves both semantic and sentiment property based on the instance–concept relationships and affective labels.

Semantic Clustering

To perform concept-level topic-spotting in natural language opinions, our method assigns different membership degrees to different classes for each instance. To this end, $\tilde{C}_U$ is clustered

into k distinct categories represented by Isanette's hub concepts, that is, the top 5,000 concepts with highest in-degree in Isanette.

We employ a sentic medoids[14] approach. Differently from the k-means algorithm (which doesn't pose constraints on centroids), sentic medoids assume that centroids must coincide with k observed points, which allows for better clustering of a commonsense-knowledge vector space.[14] The sentic medoids approach is similar to the partitioning around medoids (PAM) algorithm, which determines a medoid for each cluster by selecting the most centrally located centroid within that cluster. Unlike other PAM techniques, however, the sentic medoids algorithm runs similarly to k-means and, hence, requires significantly reduced computational time.

Generally, the initialization of clusters for clustering algorithms is a problematic task as the process often risks getting stuck in local optimum points, depending on the initial centroid choice. For this study, however, we use the most representative (highest confidence score) instances of Isanette's hub concepts as initial centroids. For this reason, what is usually a limitation of the algorithm is an advantage for this study, because what we are seeking isn't the 5,000 centroids leading to the best 5,000 clusters, but the 5,000 centroids identifying the top 5,000 hub concepts (that is, the centroids shouldn't be too far from the most representative instances of these concepts). Therefore, given that the distance between two points in the space is defined as

$$D(e_i, e_j) = \sqrt{\sum_{s=1}^{d'} \left(e_i^{(s)} - e_j^{(s)}\right)^2},$$

we can summarize the adopted algorithm as follows:

1. Each centroid $\bar{e}_i \in \mathbb{R}^{d'} (i = 1, 2, \ldots, k)$ is set as one of the k most representative instances of the top hub concepts.
2. Assign each instance e_j to a cluster $\bar{e}_i$ if $D(e_j, \bar{e}_i) \leq D(e_j, \bar{e}_{i'})$ where $i(i') = 1, 2, \ldots, k$.
3. Find a new centroid $\bar{e}_i$ for each cluster c so that $\sum_{j \in Cluster\ c} D(e_j, \bar{e}_i) \leq \sum_{j \in Cluster\ c} D(e_j, \bar{e}_{i'})$.
4. Repeat steps 2 and 3 until no changes on centroids are observed.

Exploiting the Knowledge Base

To assess the accuracy of IsaCore, we developed an opinion-mining engine able to infer both the conceptual and affective information associated with natural language text. This engine consists of four main components: a preprocessing module, which performs a first skim of the opinion; a semantic parser, whose aim is to extract concepts from the opinionated text; a target spotting module, which identifies opinion targets; and an affect interpreter, for emotion recognition and polarity detection.

The preprocessing module firstly interprets special punctuation, complete upper-case words, cross-linguistic onomatopoeias, exclamation words, negations, degree adverbs, and emoticons. Secondly, it converts text to lower-case and, after lemmatizing it, splits the opinion into single clauses according to grammatical conjunctions and punctuation.

Then, the semantic parser deconstructs text into small bags of concepts (SBoCs) using a lexicon based on sequences of lexemes that represent multiple-word concepts extracted from ConceptNet and Isanette. These n-grams are not used blindly as fixed word patterns but exploited as reference for the module to extract multiple-word concepts from information-rich sentences. So, differently from other shallow parsers, the module can recognize complex concepts when they are interspersed with adjectives and adverbs, for example, the concept "buy christmas present" in the sentence "I bought a lot of very nice Christmas presents."

The target-spotting module aims to individuate one or more opinion targets, such as people, places, and events, from the input concepts. This is done by projecting the concepts of each SBoC into $\tilde{C}_U$, clustered according to Isanette's hub concepts. The categorization doesn't consist of simply labeling each concept, but also assigns a confidence score to each category label, which is directly proportional to the value of belonging (dot product) to a specific conceptual cluster.

The affect interpreter, similarly, projects the concepts of each SBoC into $\tilde{C}_U$, clustered according to the Hourglass labels, and, hence, calculates polarity in terms of the Hourglass dimensions (pleasantness, attention, sensitivity, and aptitude) according to the formula[13]

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

where c_i is an input concept, N the SBoC size, and 3 the normalization factor.

To evaluate the different facets of the opinion-mining engine from different perspectives, we used three resources, namely a Twitter hashtag repository, a LiveJournal database, and a PatientOpinion (see http://patientopinion.org.uk) dataset, and compared the obtained results using WordNet, ConceptNet, and Isanette.

The first resource is a collection of 3,000 tweets crawled from the Bing Web repository by exploiting Twitter hashtags as category labels, which we find useful in testing the engine's target-spotting performances. In particular, hashtags about electronics

Table 1. Precision values relative to Twitter evaluation and F-measure values relative to LiveJournal and PatientOpinion evaluations.

Database	Category	WordNet (%)	ConceptNet (%)	Isanette (%)	IsaCore (%)
Twitter	Electronics	34.5	45.3	79.1	79.2
Twitter	Companies	26.4	51.0	82.3	82.3
Twitter	Countries	38.2	65.4	85.2	84.9
Twitter	Cities	25.3	59.3	80.4	81.8
Twitter	Operative systems	37.3	51.4	77.8	75.6
Twitter	Cars	13.1	22.2	76.5	76.7
LiveJournal	Joy–sadness	47.5	55.1	75.5	81.8
LiveJournal	Anticipation–surprise	30.2	41.4	62.3	73.0
LiveJournal	Anger–fear	43.3	49.0	60.6	71.6
LiveJournal	Trust–disgust	27.3	39.5	58.8	69.9
PatientOpinion	Clinical service	35.1	49.5	78.3	82.9
PatientOpinion	Communication	41.0	50.4	71.6	79.7
PatientOpinion	Food	39.3	45.4	65.9	81.6
PatientOpinion	Parking	47.3	51.6	73.4	77.8
PatientOpinion	Staff	32.9	37.2	69.8	73.9
PatientOpinion	Timeliness	44.0	50.4	62.8	80.8

(for example, iPhone, Xbox, and Wii), companies (for example, Apple, Microsoft, and Google), countries, cities, operative systems, and cars are selected.

The second resource is a 5,000-blogpost database extracted from LiveJournal, a virtual community of more than 23 million users who keep a blog, journal, or diary. An interesting feature of this website is that bloggers are allowed to label their posts with a mood tag by choosing from predefined mood themes or by creating new ones. Because the indication of mood tags is optional, posts are likely to reflect the authors' true mood.

The third resource, finally, is a dataset obtained from PatientOpinion, a social enterprise pioneering an online feedback service for users of the UK national health service. This is a manually-tagged dataset of 2,000 patient opinions that associates to each post a category (namely, clinical service, communication, food, parking, staff, and timeliness) and a positive or negative polarity.

To assess the accuracy of IsaCore, we carried out a comparison study by replacing it with state-of-the-art knowledge bases in the opinion mining engine. In particular, we first swapped WordNet, ConceptNet, and Isanette with IsaCore to compare topic-spotting performance and emotion-recognition capabilities on the Twitter hashtag repository and the LiveJournal database. Secondly, we repeated the same evaluation process to assess the engine's topic-spotting and polarity-detection capabilities on the PatientOpinion dataset.

For the Twitter evaluation, results show that Isanette and IsaCore perform significantly better than WordNet and ConceptNet, as these lack factual knowledge concepts such as Nintendo Wii or Ford Focus (see Table 1). Isanette's and IsaCore's topic-spotting precision, on the other hand, is comparable as Isanette hyponym–hypernym common knowledge is enough for this task type. Isanette actually outperforms IsaCore sometimes, because IsaCore contains just a subset of Isanette instances (hub instances) and common-sense knowledge doesn't play a key role in this type of classification.

As for the LiveJournal evaluation, we evaluated the software engine's ability to properly categorize antithetical affective pairs from the Hourglass model (namely joy–sadness, anticipation–surprise, anger–fear, and trust–disgust). Results show that, in this case, IsaCore consistently outperformed Isanette, because Isanette is based on semantic, rather than affective, relatedness of concepts (F-measure values are reported in Table 1). In Isanette's vector-space representation, in fact, instances like "joy," "surprise," and "anger" are all close to each other, although they convey different affective valence, even though they're associated with the same hyponym–hypernym relationships.

As for the PatientOpinion evaluation, finally, IsaCore turns out to be the best choice as it represents the best tradeoff between common and common-sense knowledge, which is particularly needed when aiming to infer both the conceptual and affective information associated with text. As shown by previous experiments, in fact, common knowledge is particularly functional for tasks such as open-domain text auto-categorization, while common-sense knowledge is notably useful for natural language understanding and inference of implicit meanings underpinning words.

In this work, common and common-sense knowledge were blended together to build a comprehensive resource that can be seen as an attempt to emulate how tacit and explicit knowledge is organized in the human mind, and how this can be exploited to perform reasoning within natural language tasks such as opinion mining and sentiment analysis. It's usually difficult to take advantage of a knowledge base in systems different from the one for which the resource was conceived. Indeed, the knowledge base's underlying symbolic framework and content, while being efficient for its original purpose, aren't flexible enough to be fruitfully exported and embedded in any application.

IsaCore is different because it's an open-domain resource and exploits reasoning techniques to infer general conceptual and affective information, which can be used for many tasks such as opinion mining, affect recognition, text auto-categorization, and so on. While this study has shown encouraging results, we're planning further research studies to investigate the possibility of a better tradeoff between size and sparseness in IsaCore. At the same time, we'll explore new semantic MDS techniques to perform reasoning on the knowledge base. ■

References

1. C. Fellbaum, *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*, MIT Press, 1998.
2. P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Ann. Mtg. Assoc. Computational Linguistics*, 2002, pp. 417–424.
3. J. Kamps et al., "Using WordNet to Measure Semantic Orientation of Adjectives," *Proc. Int'l Conf. Language Resources and Evaluation*, 2004, pp. 1115–1118.
4. E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," *Proc. Conf. Empirical Methods on Natural Language Processing*, 2003, pp. 105–112.
5. W. Wu et al., "Probase: A Probabilistic Taxonomy for Text Understanding," *Proc. Sigmod*, 2012, pp. 481–492.
6. R. Speer and C. Havasi, "ConceptNet 5: A Large Semantic Network for Relational Knowledge," *Theory and Applications of Natural Language Processing*, E. Hovy, M. Johnson, and G. Hirst, eds. Springer, 2012, ch. 6.
7. K. Bollacker et al., "Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge," *Proc. Sigmod*, 2008, pp. 1247–1250.
8. F. Suchanek, G. Kasneci, and G. Weikum, "YAGO: A Core of Semantic Knowledge," *Proc. Int'l World Wide Web Conf.*, 2007, pp. 697–706.
9. D. Lenat and R. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Addison-Wesley, 1989.
10. E. Cambria et al., "Isanette: A Common and Common Sense Knowledge Base for Opinion Mining," *Proc. Int'l Conf. Data Mining*, 2011, pp. 315–322.
11. C. Havasi et al., "Digital Intuition: Applying Common Sense Using Dimensionality Reduction," *IEEE Intelligent Systems*, vol. 24, no. 4, 2009, pp. 24–35.
12. Y. Song et al., "A Unified Framework for Semi-Supervised Dimensionality Reduction," *Pattern Recognition*, vol. 41, no. 9, 2008, pp. 2789–2799.
13. E. Cambria, A. Livingstone, and A. Hussain, "The Hourglass of Emotions," *Cognitive Behavioral Systems*, A. Esposito et al., eds., LNCS 7403, Springer, 2012, pp. 144–157.
14. E. Cambria et al., "Sentic Medoids: Organizing Affective Common Sense Knowledge in a Multi-Dimensional Vector Space," *Advances in Neural Networks*, D. Liu et al., eds., LNCS 6677, Springer, 2011, pp. 601–610.

THE AUTHORS

Erik Cambria is an assistant professor in the School of Computer Engineering of Nanyang Technological University. His research interests include concept-level sentiment analysis, affective common-sense reasoning, and intention awareness. Cambria has a PhD in computing science and mathematics from the University of Stirling. He's on the editorial board of Springer's *Cognitive Computation* and is the chair of many international conferences such as Extreme Learning Machines (ELM) and workshop series such as ICDM SENTIRE. Contact him at cambria@ntu.edu.sg.

Yangqiu Song is a postdoctoral researcher at the University of Illinois at Urbana-Champaign. His current research focuses on using machine learning and data mining to extract and infer insightful knowledge from big data, including the techniques of large-scale learning algorithms, natural language understanding, text mining and visual analytics, and knowledge engineering. Song has a PhD from Tsinghua University, China. Contact him at yqsong@illinois.edu.

Haixun Wang is a senior researcher at Microsoft Research Asia, where he manages the data management, analytics, and services group. Also, he is currently with Google Research. His research interests include text analytics and natural language processing. Wang has a PhD in computer science from the University of California, Los Angeles. He's on the editorial board of *Distributed and Parallel Databases, IEEE Transactions on Knowledge and Data Engineering, Knowledge and Information System,* and *Journal of Computer Science and Technology.* Contact him at haixun@google.com.

Newton Howard is one of the directors of the Synthetic Intelligence Project and a resident scientist at the Massachusetts Institute of Technology. He is the board director of the Center for Advanced Defense Studies and is a national security advisor to several US Government organizations. His current research focuses on the molecular basis for human intelligence. Howard has a doctoral degree in cognitive informatics and mathematics from La Sorbonne in France. Contact him at nhmit@mit.edu.