Isanette: A Common and Common Sense Knowledge Base for Opinion Mining

Erik Cambria*, Yangqiu Song†, Haixun Wang† and Amir Hussain ‡
*Temasek Laboratories, National University of Singapore, cambria@nus.edu.sg
†Microsoft Research Asia, {yangqiu.song, haixun.wang}@microsoft.com
‡COSIPRA Lab, University of Stirling, ahu@cs.stir.ac.uk

***Abstract*—The ability to understand natural language text is far from being emulated in machines. One of the main hurdles to overcome is that computers lack both the common and the common sense knowledge humans normally acquire during the formative years of their lives. If we want machines to really understand natural language, we need to provide them with this kind of knowledge rather than relying on the valence of keywords and word co-occurrence frequencies. In this work, we blend the largest existing taxonomy of common knowledge with a natural-language-based semantic network of common sense knowledge, and use multi-dimensionality reduction techniques on the resulting knowledge base for opinion mining and sentiment analysis.**

***Keywords*-Knowledge-Based Systems; Semantic Networks; Natural Language Processing; Opinion Mining.**

I. INTRODUCTION

The ever-growing amount of available information in the Social Web fostered the proliferation of many business and research activities around the relatively new fields of opinion mining and sentiment analysis. The automatic analysis of user generated contents 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 multi-faceted and multi-disciplinary problem as opinions and sentiments can be expressed in a multitude of forms and combinations in which it is extremely difficult to find any kind 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 human mind.

Most of current approaches to opinion mining and sentiment analysis rely on rather unambiguous affective keywords extracted from an existing knowledge base (e.g., WordNet [1]) or from a purpose-built lexicon based on a domain-dependent corpus [2], [3], [4], [5]. Such approaches are still far from being able to perfectly extract the cognitive and affective information associated with natural language and, hence, often fail to meet the golden standard of human annotators.

Especially when dealing with social media, in fact, contents are often very diverse and noisy and the use of a limited number of affect words or a domain-dependent training corpus is simply not enough (see Table I). In order to enable computers to intelligently process open-domain textual resources, we need to provide them with both the common and common sense knowledge humans normally acquire during the formative years of their lives, as relying just on valence of keywords and word co-occurrence frequencies does not allow a deep understanding of natural language. Common knowledge represents human general knowledge acquired from the world, e.g., "canine distemper is a domestic animal disease". Common sense knowledge is some obvious thing that people normally know but usually leave unstated, e.g., "cat can hunt mice" and "cat is cute".

It is through the combined use of common and common sense knowledge that we can have a grip on both low and high level concepts in natural language sentences and, hence, effectively communicate with other people without having to continuously ask for definitions and explanations. Common sense knowledge, moreover, enables the propagation of sentiment from affect words, e.g., 'happy' and 'sad', to general concepts, e.g., 'birthday gift', 'school graduation', 'cancer' and 'canine distemper', which is useful for tasks such as sentiment elicitation and polarity detection. In this work, we blend ProBase [6], the largest existing taxonomy of common knowledge, with ConceptNet [7], a natural-language-based semantic network of common sense knowledge, and use multi-dimensionality reduction techniques on the resulting knowledge base for opinion mining and sentiment analysis.

The structure of the paper is as follows: Section II presents related works in the field of opinion mining, Section III discusses how and why blending common and common sense knowledge is important for the development of domain independent sentiment analysis system, Section IV explains in detail the strategies adopted to build the common and common sense knowledge base, Section V illustrates the dimensionality reduction techniques employed to perform reasoning on the newly built knowledge base, Section VI presents the development of an opinion mining engine and its evaluation, Section VII, eventually, comprises concluding remarks and future directions.

978-0-7695-4409-0/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDMW.2011.106

Table I
LIVE JOURNAL POSTS WHERE AFFECTIVE INFORMATION IS NOT CONVEYED THROUGH AFFECT WORDS.

Sentiment	Live Journal Posts	ProBase Instances
Happy	Finally I got my student cap ! I am officially high school graduate now ! Our dog Tanja, me, Timo (our art teacher) and EmmaMe, Tanja, Emma and Tiia Only two weeks to Japan !!	student; school graduate; Japan
Happy	I got a kitten as an early birthday gift on Monday. Abby was smelly, dirty, and knawing on the metal bars of the kitten carrier though somewhat calm when I picked her up. We took her. She threw up on me on the ride home and repeatly keeps sneezing in my face.	kitten; birthday gift; metal bar; face
Sad	Hi. Can I ask a favor from you? This will only take a minute. Please pray for Marie, my friends' dog a labrador, for she has canine distemper. Her lower half is paralyzed and she's having locked jaw. My friends' family is feeding her through syringe.	friends; dog; labrador; canine distemper; jaw; syringe
Sad	my uncle paul passed away on febuary 16, 2008. he lost his battle with cancer. i remember spending time with him and my aunt nina when they babysat me. we would go to taco bell and i would get nachos.	uncle; battle; cancer; aunt; taco bell; nachos

II. RELATED WORK

Early works in the field of opinion mining and sentiment analysis aimed to classify entire documents as containing rating scores (e.g., 1-5 stars) of reviews [8] or overall positive or negative polarity [9]. These were mainly supervised approaches relying on manually-labeled samples, such as movie or product reviews where the opinionist's overall positive or negative attitude was explicitly indicated. However, opinions and sentiments do not occur only at document level, nor they are limited to a single valence or target. Contrary or complementary attitudes toward the same topic 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 non-sentimental sections, e.g., by using graph-based techniques for segmenting sections of a document on the basis of their subjectivity [10] or by performing a classification based on some fixed syntactic phrases likely to be used to express opinions [2] or by bootstrapping using a small set of seed opinion words and a knowledge base such as WordNet [3].

In recent works, text analysis granularity has been taken down to sentence level, e.g., by using presence of opinion-bearing lexical items (single words or n-grams) to detect subjective sentences [4], [5] or by using semantic frames defined in FrameNet [11] for identifying the topics (or targets) of sentiment [12] or by exploiting an affective common sense knowledge base for a feature-based analysis of product reviews [13]. These approaches, however, are still far from being able to infer the cognitive and affective information associated with natural language as they mainly rely on semantic knowledge bases which are still too limited to efficiently process text at sentence level. Moreover, text analysis granularity might still not be enough as a single sentence may express more than one opinion [14].

The main aim of this work is to build possibly the most comprehensive resource of common and common sense knowledge in order to perform a domain-independent clause-level analysis of opinion and sentiments on the Web.

III. COMMON AND COMMON SENSE KNOWLEDGE

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 comprehends 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 [15] (vocabulary knowledge).

However, when people communicate with each other, in fact, they also 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 we call common sense – obvious things people normally know and usually leave unstated.

A. Common Knowledge Base

Attempts to build a common knowledge base are countless and comprehend both resources crafted by human experts or community efforts, such as WordNet, a lexical knowledge base of about 25,000 words grouped into an ontology of synsets, or Freebase [16], a social database of 1,450 concepts, and automatically-built knowledge bases, such as WikiTaxonomy [17], a taxonomy of about 127,000 concepts extracted from Wikipedia's category links, YAGO [18], a semantic knowledge base of 149,162 instances derived from Wikipedia, WordNet and GeoNames [19], and ProBase [20].

ProBase contains about 12 million concepts learned iteratively from 1.68 billion web pages in Bing [20] web 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 evidences found in web data, search log data and other available data. The core taxonomy consists of the "IsA" relationships extracted by using syntactic patterns such as the Hearst patterns [21]. For example, a segment like "artists such as Pablo Picasso" can be considered as a piece of evidence for the claim that 'pablo picasso' is an instance of the concept 'artist'.

B. Common Sense Knowledge Base

One of the biggest projects aiming to build a comprehensive common sense knowledge base is Cyc [22]. Cyc requires knowledge engineers working on some specific languages, and contains just 120,000 concepts as this is labor intensive and time consuming. A more recent project is Open Mind Common Sense (OMCS), which has been collecting pieces of knowledge from volunteers on the Internet since 2000 by enabling the general public 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, a semantic network of 173,398 nodes. Table II shows the comparison of this collaboratively constructed common sense knowledge base with WordNet and Probase. As we can see from the table, WordNet contains very detailed descriptions of every word's various senses but it does not include enough general Web information. ProBase, which provides more concepts, includes pieces of knowledge that match general distribution of human knowledge. ConceptNet, in turn, contains implicit knowledge that people rarely mention on the Web, which is a good complementary material to Probase.

To this end, in this work we blend ProBase and ConceptNet and, hence, build a comprehensive knowledge base that can be seen as one of the first attempts to emulate how tacit and explicit knowledge are organized in human mind and how this can be exploited to perform reasoning within natural language tasks. Providing a machine with a database of millions of common and common sense concepts, in fact, would still be not enough for it to be intelligent: it needs to be taught how to handle this knowledge, retrieve it when necessary, make analogies and learn from experience.

To test our newly built knowledge base, we apply multi-dimensionality reduction techniques on it and exploit the resulting system to perform opinion mining tasks, for which a good trade-off between common and common sense knowledge is particularly needed in order to infer both the topical clues and the polarity conveyed by natural language.

IV. Building the Knowledge Base

In this work, we focus on IsA relationships to build a semantic network, which we call Isanette (IsA net). It represents hyponym-hypernym common knowledge as a $2{,}715{,}218 \times 1{,}331{,}231$ matrix having instances (e.g., 'pablo picasso') as rows and concepts (e.g., 'artist') as columns. Performing reasoning on Isanette as it is, however, is not very convenient as it is a very large and fat matrix that contains noise and multiple forms, since all of the evidences are automatically extracted from the Web. To this end, we firstly clean it by applying different natural language processing (NLP) techniques (Section IV-A) and, secondly, enhance its consistency and further reduce its sparseness by adding complementary common sense knowledge (Section IV-B).

A. Cleaning Isanette

We build Isanette out of 23,066,575 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, namely multiple word forms and low connectivity.

We address the former issue by processing both subjects and objects of triples with OMCS lemmatizer, which groups together the different inflected forms of words (different cases, plurals, verb tenses, etc.) so that they can be stored in Isanette as a single item. In case of duplicates, we simply consider the triple with higher confidence score.

As for Isanette's connectivity, if we want to apply dimensionality reduction techniques on it in order to find similar patterns, we would like the matrix to be as less sparse as possible. To this end, we firstly want to get rid of hapax legomena, that is instances/concepts with singular out-/in-degree. These nodes can be useful for specific tasks such as finding the meaning of uncommon instances or give an example of a rare concept. For more general reasoning tasks, however, hapax legomena are very bad as they enlarge dimensionality without providing overlapping information that can be useful for finding similar patterns and perform analogies. In this work, we choose to discard not only hapax legomena but also the other nodes with low connectivity, in order to heavily reduce Isanette's sparseness.

In particular, we used a trial and error approach and found that the best trade-off between size and sparseness is achieved by setting the minimum node connectivity equal to 10. This cut-off operation leaves out almost 40% of nodes and makes Isanette a strongly connected core. Moreover, we exploit dimensionality reduction techniques to infer negative evidence such as 'carbonara' is not a kind of 'fuel' or 'alitalia' is not a 'country', which is very useful to further reduce Isanette's sparseness and improve reasoning algorithms (more details in Section V).

B. 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 is not usually stated in web pages (or at least not that often to be extracted by Hearst patterns with a high enough confidence score). To overcome this problem, we enrich Isanette with complementary hyponym-hypernym common sense knowledge from ConceptNet.

In particular, we extract from the Open Mind corpus 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". We exploit these assertions to generate a directed graph of about 15,000 nodes (interconnected by IsA edges), representing subsumption common sense knowledge.

Table II
COMPARISON BETWEEN DIFFERENT KNOWLEDGE BASES.

Term	WordNet Hypernyms	ConceptNet Assertions	Probase Concepts
Cat	Feline; Felid; Adult male; Man; Gossip; Gossiper; Gossipmonger; Rumormonger; Rumourmonger; Newsmonger; Woman; Adult female; Stimulant; Stimulant drug; Excitant; Tracked vehicle; ...	Cats can hunt mice; Cats have whiskers; Cats can eat mice; Cats have fur; cats have claws; Cats can eat meat; cats are cute; ...	Animal; Pet; Species; Mammal; Small animal; Thing; Mammalian species; Small pet; Animal species; Carnivore; Domesticated animal; Companion animal; Exotic pet; Vertebrate; ...
Dog	Canine; Canid; Unpleasant woman; Disagreeable woman; Chap; Fellow; Feller; Lad; Gent; Fella; Scoundrel; Sausage; Follow, ...	Dogs are mammals; A dog can be a pet; A dog can guard a house; You are likely to find a dog in kennel; An activity a dog can do is run; A dog is a loyal friend; A dog has fur; ...	Animal; Pet; Domestic animal; Mammal; Species; Companion animal; Domesticated animal; Household pet; Small animal; Carnivore; Family pet; Follower; ...
iPhone	N/A;	An iPhone is a kind of a telephone; An iPhone is a kind of computer; An IPhone can display your position on a map; An IPhone can send and receive emails; An IPhone can display the time; ...	Device; Smartphones; Mobile device; Apple product; Mobile platform; Fancy phone; Popular phone brand; Latest popular brand; ...
Birthday gift	Present;	Card is birthday gift; Present is birthday gift; Buying something for a loved one is for a birthday gift; ...	Gift; Occasion; Expense; Flower delivery service; Birthday candle related product; ...
School Graduation	N/A	N/A	Occasion; Milestone family event; Large event; School function; Outcome; Accomplishment; Milestone
Canine distemper	Distemper	N/A	Disease; Viral disease; Infectious disease; Domestic animal disease; Epidemic infection; ...

To merge this subsumption common sense knowledge base with Isanette, we use blending [23], a technique that performs inference over multiple sources of data simultaneously, taking advantage of the overlap between them.

Blending combines two sparse matrices linearly into a single matrix in which the information between the two initial sources is shared. This alignment operation yields a new strongly connected core, $C \in \mathcal{R}^{m\times n}$, in which common and common sense knowledge coexist, i.e., a matrix 340,000 $\times$ 200,000 whose rows are instances such as 'birthday party' and 'china', whose columns are concepts like 'special occasion' and 'country', and whose values indicate truth values of assertions.

V. REASONING ON THE KNOWLEDGE BASE

In this section, we apply dimensionality reduction techniques to build a vector space representation of the instance-concept relationship matrix (Section V-A) and employ a partitioning clustering technique to segment the reduced space into conceptual classes (Section V-B).

A. Vector Space Representation

In order to more compactly represent the information contained in C and encode the latent semantics between its instances, we build a multi-dimensional vector space representation by applying truncated singular value decomposition (SVD) [24]. For the Eckart–Young theorem [25], the resulting lower-dimensional space represents the best approximation of C, in fact:

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

where C has the form $C = U\ \Sigma\ V^T$, $\tilde{C}$ has the form $\tilde{C} = U_d\, S_d\, V_d^T$ ($U_d \in \mathcal{R}^{m\times d}$, $V_d \in \mathcal{R}^{n\times d}$, $S_d \in \mathcal{R}^{d\times d}$), and d is the lower dimension of the latent semantic space. From the rank constraint, i.e., S_d has d non-zero diagonal entries, the minimum of the above 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 $\sigma_i = s_i$ $(i = 1, ..., d)$ and the corresponding singular vectors are the same as those of C. If we choose to discard all but the first d principal components and consider $\tilde{C}_U = U_d\, S_d$, 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, i.e., its basis $e = (e^{(1)},...,e^{(d)})^T$. We used a trial and error approach and found that the best compromise is achieved when d assumes values around 500.

We can use such 500-dimensional vector space 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).

B. Semantic Clustering

In order to perform concept-level topic-spotting in natural language opinions, we want to assign to each instance different degrees of membership to different classes. To this end, we cluster $\tilde{C}_U$ 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 k-medoids approach. Differently from the k-means algorithm, which does not pose constraints on centroids, k-medoids do assume that centroids must coincide with k observed points. The most commonly used algorithm for finding the k-medoids is the partitioning around medoids (PAM) algorithm. The PAM algorithm determines a medoid for each cluster selecting the most centrally located centroid within the cluster. After selection of medoids, clusters are rearranged so that each point is grouped with the closest medoid. For the purpose of this work, we use a modified version of the algorithm recently proposed by Park and Jun [26], which runs similarly to the k-means clustering algorithm. This has shown to have similar performance when compared to PAM algorithm while taking a significantly reduced computational time. Generally, the initialization of clusters for clustering algorithms is a problematic task as the process often risks to get stuck into local optimum points, depending on the initial choice of centroids [27].

However, for this study, we decide to use as initial centroids the most representative (highest confidence score) instances of Isanette's hub concepts. For this reason, what is usually seen as a limitation of the algorithm can be seen as advantage for this study, since we are not looking for the 5,000 centroids leading to the best 5,000 clusters but indeed for the 5,000 centroids identifying the top 5,000 hub concepts (i.e., the centroids should not 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 $D(e_i, e_j) = \sqrt{\sum_{s=1}^{d'} \left(e_i^{(s)} - e_j^{(s)}\right)^2}$, the used algorithm can be summarized as follows:

1) Each centroid $\bar{e}_i \in \mathbb{R}^{d'}$ $(i = 1, 2, ..., k)$ 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 $i(i') = 1, 2, ..., k$
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 step 2 and 3 until no changes on centroids are observed

VI. Exploiting the Knowledge Base

The main aim of this research work is to build a comprehensive common and common sense knowledge base and to exploit it for efficiently analyzing natural language text at semantic level. In particular, we choose to exploit such resource for opinion mining tasks, for which a good trade-off between common and common sense knowledge is specifically needed, in order to infer both the topical clues and the polarity conveyed by natural language opinions.

In this section, we show in detail how we embedded the newly built knowledge base into a software engine for the analysis of on-line opinions (Section VI-A) and how we evaluated the performances of such engine (Section VI-B).

A. Opinion Mining Engine

In order to effectively mine opinions and sentiments, we developed a software engine able to infer both the cognitive and affective information associated with natural language text. This software engine consists of four main components: a pre-processing 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 pre-processing module firstly interprets all the affective valence indicators usually contained in opinionated text such as special punctuation, complete upper-case words, onomatopoeic repetitions, 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 concepts using a lexicon based on sequences of lexemes that represent multiple-word concepts extracted from ConceptNet, WordNet and other linguistic resources. These n-grams are not used blindly as fixed word patterns but exploited as reference for the module, in order to extract multiple-word concepts from information-rich sentences. So, differently from other shallow parsers, the module can recognize complex concepts also when irregular verbs are used or when these are interspersed with adjective and adverbs, e.g., the concept 'buy christmas present' in the sentence "I bought a lot of very nice Christmas presents".

The semantic parser, additionally, provides, for each retrieved concept, the relative frequency, valence and status, that is the concept's occurrence in the text, its positive or negative connotation and the degree of intensity with which the concept is expressed. For each clause, the module outputs a small bag of concepts (SBoC), which is later on analyzed separately by the target spotting module and the affect interpreter to infer the cognitive and affective information associated with the input text, respectively.

The target spotting module aims to individuate one or more opinion targets, such as people, places, events and ideas, from the input concepts. This is done by projecting the concepts of each SBoC into the vector space representation of Isanette, in order to assign these to a specific conceptual class. The categorization does not consist in simply labeling each concept but also in assigning a confidence score to each category label, which is directly proportional to the value of belonging to a specific conceptual cluster (dot product).

The affect interpreter, similarly, projects the concepts of each SBoC into the vector space clustered using sentic medoids [28] (an approach similar to k-medoids that uses the Hourglass of Emotions [29] as emotion categorization model for clustering). The module, in particular, assigns concepts to a specific affective class and, hence, calculates polarity in terms of Pleasantness, Attention, Sensitivity and Aptitude, according to the formula proposed in [30]:

$$p = \sum_{i=1}^{N} \frac{Plsnt(c_i) + |Attnt(c_i)| - |Snst(c_i)| + Aptit(c_i)}{9N}$$

where c_i is an input concept, N the size of the SBoC and 9 the normalization factor (as the Hourglass dimensions are defined as float $\in$ [-3,+3]). In the formula, Attention and Sensitivity are taken in absolute value since both their positive and negative intensity values correspond to merely positive or negative polarity values respectively (e.g., ‘surprise’ is negative in the sense of lack of Attention but positive from a polarity point of view, and ‘anger’ is positive in the sense of level of activation of Sensitivity but negative in terms of polarity).

As an example of how the software engine works, we can examine intermediate and final outputs obtained when a natural language opinion is given as input to the system. We choose the tweet “I think iPhone4 is the top of the heap! OK, the speaker is not the best i hv ever seen bt touchscreen really puts me on cloud 9... camera looks pretty good too!”. After the pre-processing and semantic parsing operations, we obtain the following SBoCs:

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

Table III
STRUCTURED OUTPUT EXAMPLE OF OPINION MINING ENGINE

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

These are then concurrently processed by the target spotting module and the affect interpreter, which detect the opinion targets and output, for each of them, the relative affective information both in a discrete way, with one or more emotional labels, and in a dimensional way, with a polarity value $\in$ [-1,+1] (as shown in Table III).

B. Evaluation

In order to evaluate the different facets of the opinion mining engine from different perspectives, we used three different resources, namely a Twitter [31] hashtag repository, a LiveJournal [32] database and a Patient Opinion [33] dataset, and compared results obtained using WordNet, ConceptNet and Probase. The first resource is a collection of 3,000 tweets crawled from Bing web repository by exploiting Twitter hashtags as category labels, which is useful to test the engine’s target spotting performances. In particular, we selected hashtags about electronics (e.g., IPhone, XBox, Android and Wii), companies (e.g., Apple, Microsoft and Google), countries, cities, operative systems and cars. In order to test the resource’s consistency and reliability, we performed a manual evaluation of 100 tweets, which showed that hashtags are accurate to 89%.

The second resource is a 5,000 blogpost database extracted from LiveJournal, a virtual community of more than 23 millions users who keep a blog, journal or diary. An interesting feature of this website is that bloggers are allowed to label their posts with both a category and a mood tag, by choosing from predefined categories and mood themes. Since the indication of mood tags is optional, posts are likely to reflect the true mood of the authors, which is not always true for category tags. After a manual evaluation of 200 posts, in fact, the category tags turned out to be very noisy (53% accuracy). The mood tags, however, showed a good enough reliability (78% accuracy) so we used them to test the engine’s affect recognition performances.

The third resource, eventually, is a dataset obtained from Patient Opinion, a social enterprise pioneering an on-line feedback service for users of the UK national health service to enable people to share their recent experience of local health services on-line. It is a manually tagged dataset of 2,000 patient opinions [34] that associates to each post a category (namely, clinical service, communication, food,

Table IV
PRECISION VALUES RELATIVE TO TWITTER EVALUATION

	WordNet	ConceptNet	Probase	Isanette
electronics	34%	45%	76%	79%
companies	26%	51%	82%	82%
countries	38%	65%	89%	85%
cities	25%	59%	81%	80%
operative systems	37%	51%	79%	77%
cars	13%	22%	74%	76%

parking, staff and timeliness) and a positive or negative polarity. We used it to test the detection of opinion targets and the polarity associated with these.

Evaluations were performed by firstly lemmatizing text and searching for matches between words appearing in it and concepts contained in the adopted knowledge bases respectively. Secondly, retrieved concepts were located in the graph structure of the respective semantic networks in order to semantically/affectively categorize them according to predefined category/affect nodes.

As for the Twitter evaluation, results show that Probase and Isanette perform significantly better than WordNet and ConceptNet, as these lack factual knowledge concepts such as Wii or Ford Focus (see Table IV). Probase and Isanette topic spotting precision, on the other hand, are comparable as Probase hyponym-hypernym common knowledge is enough for this kind of task. It actually even outperforms Isanette sometimes as this contains just a subset of Probase instances (hub instances) and common sense knowledge does not play a key role in this type of classification.

As for the LiveJournal evaluation, we chose to evaluate the capability of the software engine 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, Probase is consistently outperformed by WordNet, ConceptNet and Isanette as it is based on semantic rather than affective relatedness of concepts (F-measure values are reported in Table V). In Probase graph representation, in fact, instances like 'joy', 'surprise' and 'anger' are all close to each other, although they convey different affective valence, for being associated with the same hyponym-hypernym relationships.

As for the Patient Opinion evaluation, eventually, Isanette turns out to be the best choice as it represents the best trade-off between common and common sense knowledge, which is particularly needed when aiming to infer both the cognitive and affective information associated with text (F-measure values are reported in Table VI). As also 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 meaning underpinning words.

Table V
F-MEASURE VALUES RELATIVE TO LIVEJOURNAL EVALUATION

	WordNet	ConceptNet	Probase	Isanette
joy-sadness	47%	55%	33%	75%
anticipation-surprise	30%	41%	19%	62%
anger-fear	43%	49%	25%	60%
trust-disgust	27%	39%	12%	58%

Table VI
F-MEASURE VALUES RELATIVE TO PATIENT OPINION EVALUATION

	WordNet	ConceptNet	Probase	Isanette
clinical service	35%	49%	56%	78%
communication	41%	50%	43%	71%
food	39%	45%	40%	65%
parking	47%	51%	49%	73%
staff	32%	37%	51%	69%
timeliness	44%	50%	41%	62%

VII. CONCLUSION

In this work, we blended together common and common sense knowledge in order to build a comprehensive resource that can be seen as one of the first attempts to emulate how tacit and explicit knowledge is organized in human mind and how this can be exploited to perform reasoning within natural language tasks.

It is usually hard to take advantage of a knowledge base in systems different from the one the resource was conceived for. Indeed, its underlying symbolic framework and content, whilst being very efficient for its original purpose, are not flexible enough to be fruitfully exported and embedded in any application. Isanette is different as it is an open-domain resource and it exploits reasoning techniques able to infer cognitive and affective information, which can be used for many different tasks such as opinion mining, affect recognition, text auto-categorization, etc.

Whilst this study has shown encouraging results, further research studies are now planned to investigate if a better trade-off between size and sparseness of Isanette can be found. At the same time, we plan to explore new multi-dimensionality reduction techniques to perform reasoning on the knowledge base.

Even if we manage to teach a machine 15 million and such things, in fact, it will still be not enough for it to be intelligent: it needs to be taught how to handle this knowledge, retrieve it when necessary, make analogies and learn from experience.

ACKNOWLEDGMENT

This work has been part-funded by Microsoft Research Asia in Beijing (P.R. China), the UK Engineering and Physical Sciences Research Council (EPSRC Grant Reference: EP/G501750/1) and Sitekit Solutions Ltd. (UK).

REFERENCES

[1] C. Fellbaum, *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, 1998.

[2] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Proceedings of ACL*, Philadelphia, 2002, pp. 417–424.

[3] J. Kamps, M. Marx, R. Mokken, and M. de Rijke, "Using WordNet to Measure Semantic Orientation of Adjectives," in *Proceedings of LREC*, Lisbon, 2004, pp. 1115–1118.

[4] S. Kim and E. Hovy, "Automatic Detection of Opinion Bearing Words and Sentences," in *Proceedings of IJCNLP*, Jeju Island, 2005, pp. 61–66.

[5] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," in *Proceedings of EMNLP*, Sapporo, 2003, pp. 105–112.

[6] W. Wu, H. Li, H. Wang, and K. Zhu, "Towards a Probabilistic Taxonomy of Many Concepts," MSRA, Tech. Rep. MSR-TR-2011-25, 2011.

[7] C. Havasi, R. Speer, and J. Alonso, "ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge," in *Proceedings of RANLP*, Borovets, 2007.

[8] B. Pang and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales," in *Proceedings of ACL*, Ann Arbor, 2005, pp. 115–124.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of EMNLP*, Philadelphia, 2002, pp. 79–86.

[10] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," in *Proceedings of ACL*, Barcelona, 2004, pp. 271–278.

[11] C. Baker, C. Fillmore, and J. Lowe, "The Berkeley FrameNet Project," in *Proceedings of COLING/ACL*, Montreal, 1998, pp. 86–90.

[12] S. Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," in *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, Sydney, 2006.

[13] E. Cambria, M. Grassi, A. Hussain, and C. Havasi, "Sentic Computing for Social Media Marketing," *Multimedia Tools and Application. DOI:10.1007/s11042-011-0815-0*, 2011.

[14] T. Wilson, J. Wiebe, and R. Hwa, "Just How Mad Are You? Finding Strong and Weak Opinion Clauses," in *Proceedings of AAAI*, San Jose, 2004, pp. 761–769.

[15] "Wikipedia," 2011. [Online]. Available: http://wikipedia.org

[16] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge," in *Proceedings of SIGMOD*, Vancouver, 2008, pp. 1247–1250.

[17] S. Ponzetto and M. Strube, "Deriving a Large-Scale Taxonomy from Wikipedia," in *AAAI*, Vancouver, 2007, pp. 1440–1445.

[18] F. Suchanek, G. Kasneci, and G. Weikum, "Yago: a Core of Semantic Knowledge," in *WWW*, Banff, 2007, pp. 697–706.

[19] "Geonames," 2011. [Online]. Available: http://geonames.org

[20] "Bing," 2011. [Online]. Available: http://bing.com

[21] M. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," in *COLING*, Nantes, 1992, pp. 539–545.

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

[23] C. Havasi, R. Speer, J. Pustejovsky, and H. Lieberman, "Digital Intuition: Applying Common Sense Using Dimensionality Reduction," *IEEE Intelligent Systems*, vol. 24, no. 4, pp. 24–35, 2009.

[24] M. Wall, A. Rechtsteiner, and L. Rocha, "Singular Value Decomposition and Principal Component Analysis," in *A Practical Approach to Microarray Data Analysis*, D. Berrar, W. Dubitzky, and M. Granzow, Eds. Springer US, 2003, pp. 91–109.

[25] C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[26] H. Park and C. Jun, "A Simple and Fast Algorithm for K-Medoids Clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336–3341, 2009.

[27] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons Inc, 1973.

[28] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, "Sentic Medoids: Organizing Affective Common Sense Knowledge in a Multi-Dimensional Vector Space," ser. Lecture Notes in Computer Science, vol. 6677. Berlin Heidelberg: Springer-Verlag, 2011, pp. 601–610.

[29] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "SenticSpace: Visualizing Opinions and Sentiments in a Multi-Dimensional Vector Space," ser. Lecture Notes in Artificial Intelligence. Berlin Heidelberg: Springer, 2010, vol. 6279, pp. 385–393.

[30] E. Cambria, R. Speer, C. Havasi, and A. Hussain, "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," in *Proceedings of AAAI CSK*, Arlington, 2010, pp. 14–18.

[31] "Twitter," 2011. [Online]. Available: http://twitter.com

[32] "Livejournal," 2011. [Online]. Available: http://livejournal.com

[33] "Patient Opinion," 2011. [Online]. Available: http://patientopinion.org.uk

[34] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, "Sentic Computing for Patient Centered Applications," in *Proceedings of IEEE ICSP*, Beijing, 2010, pp. 1279 – 1282.