Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining

Soujanya Poria,[1] Alexander Gelbukh,[2] Amir Hussain,[3] Dipankar Das,[4] Sivaji Bandyopadhyay[1]

[1] *Computer Science and Engineering Department, Jadavpur University, Kolkata 700 032, India*
soujanya.poria@ieee.org, sbandyopadhyay@cse.jdvu.ac.in

[2] *Centro de Investigación en Computación, Instituto Politécnico Nacional, 07738, Mexico City, Mexico*
gelbukh@cic.ipn.mx

[3] *Division of Computing Science and Mathematics, School of Natural Sciences, University of Stirling, FK9 4LA, Scotland, UK*
ahu@cs.stir.ac.uk

[4] *Computer Science and Engineering Department, National Institute of Technology (NIT), Shillong 793003, Meghalaya, India*
dipankar.dipnil2005@gmail.com

Abstract—**Recent studies show that concept-based approaches to opinion mining perform better than more canonical methods based on keyword spotting or word co-occurrence frequencies. SenticNet 1.0 is one of the most widely used publicly available resources for concept-based opinion mining. It gives polarity scores for a large number of single- and multi-word common sense concepts. However, developing high-quality opinion mining and sentiment analysis systems also requires affective information associated with the concepts. In this work, we present a methodology for enriching SenticNet concepts with affective information by assigning to them an emotion label. The created resource is freely available for academic use.**

Keywords——**SenticNet, Sentic computing, Sentiment analysis, opinion mining, emotion lexicon, WordNet-Affect**

I. INTRODUCTION

The Internet contains important information on its user's opinions and sentiments. The extraction of such unstructured web data is known as opinion mining and sentiment analysis, a recent and explosively growing research field widely employed by the industry for purposes such as marketing, customer service, and financial market prediction.

A number of lexical resources have been developed in the past few years for different opinion mining tasks, e.g., WordNet-Affect (WNA) [1] and SentiWordNet [2], though most of them are rather incomplete and noisy. In particular, SenticNet [3] is a concept-based resource containing 5,732 single- or multi-word concepts along with a quantitative polarity score in the range from –1 to +1; e.g., *aggravation*: –0.925; *accomplish goal*: +0.967; *December*: +0.111.

However, it is often desirable to have a more complete resource containing affective labels for the concepts and not only polarity scores. Currently the main lexical resource for emotion detection in natural language texts is WNA, which assigns emotion labels to words, e.g.: *wrath*: ANGER, *nausea*: DISGUST, *triumph*: JOY. It is, however, rather small, and gives only qualitative affective information—an emotion label—but not quantitative information on the intensity of the emotion: e.g., *sulk* and *offend* have the same emotion label in WNA.

Our goal was to create a resource resulting from automatically merging SenticNet and WNA, i.e., containing both SenticNet polarity scores and WNA emotion categories for the concepts. Since the WNA's vocabulary is almost a subset of that of SenticNet, the task was to automatically extend the emotion labels from WNA to the rest of SenticNet's vocabulary.

On the one hand, the obtained resource can be viewed as an extended WNA, and the proposed methodology can be applied to extend WNA or similar resources that provide emotion labels for concepts. There are a large number of applications of WNA ranging from social, economical, and commercial to health care, political, and governmental uses.

For example, a company needs to know what emotions the customers express in the reviews of its products; the improvements thus made to the products translate in both better income for the businesses and better quality of life for the consumers. A political party or government needs to know the emotions prevailing in the press materials related to its actions; the improvements in decision-making that result from such analysis represent more efficient democracy: democracy in real time, in contrast to elections as a means of punishing bad governors when it is too late or giving credit to a party's promises when it is too early to judge.

In all such applications, a larger resource will give more precise and reliable results because more words in the analyzed texts will contribute emotion labels to the statistics.

It is more interesting, on the other hand, to view the obtained resource as SenticNet augmented with affective labels. Such a resource will give rise to a range of novel applications combining the polarity information from SenticNet with the affective information that we have added to it. In this way, a company could obtain information on which products or features the customers like and which they do not (polarity) and information on the specific emotions they feel with respect to them (affect): are they SURPRISE-d, ANGER-ed, or JOY-ful with the product.

What is more, polarity information can be used as a measure of degree of the corresponding emotion: while both *sulk* and *offend* have the same emotion label ANGER in WNA, *sulk* has polarity –0.303 in SenticNet and *offend* –0.990; thus *offend* is stronger ANGER than *sulk* is. This information is important for weighted emotion detection. For example, consider the following review of a phone: "*the keyboard is*

comfortable (+0.120, JOY) *and the interface is amicable* (+0.214, JOY), *but the color is queasy* (–0.464, DISGUST)." While it has two JOY words and only one DISGUST word, weighting the labels by the polarity score indicates that the main emotion felt by the customer was rather DISGUST.

In this paper we describe a methodology to automatically assign emotion labels to SenticNet concepts. We trained a classifier on the subset of SenticNet concepts present in WNA. As classification features, we used several concept similarity measures, as well as various psychological features available in an emotion-related corpus. Of various classifiers we tried, the highest accuracy was obtained with Support Vector Machine (SVM).

This work lays the foundation for a new concept-level opinion mining resource and presents new features and measures for generating such resources. The created resource, along with associated programs and other relevant data, is freely available for academic use from http://www.gelbukh.com/resources/emo-senticnet.

The paper is organized as follows. Section II gives a brief overview of related work; Section III describes the lexical resources we used; Section IV addresses text processing issues; Section V presents the features used for classification; Section VI describes the proposed classification procedure; Section VII shows the evaluation results; Section VIII gives some statistics of the developed resource; Section IX, finally, proposes concluding remarks and future work.

II. RELATED WORK

All known approaches to opinion mining crucially depend on the availability of adequate lexical resources that provide emotion-related information. A number of methods for their semi-automatic building have been suggested, both for English [4] and other languages [5][6].

SentiWordNet or WNA are the most widely used such resources. They are rather small and are mostly limited to affective information for single words.

Recent research [7] shows that concept-based sentiment analysis and opinion mining outperform word-based methods. This approach relies on polarity and affective information for common-sense knowledge concepts such as *accomplish goal*, *bad feeling*, *celebrate special occasion*, which are often used to express viewpoints and affect.

SenticNet is a publicly available affective common sense resource for sentic computing [7], a new paradigm that exploits Artificial Intelligence, Semantic Web, and Affective Computing [8] techniques to better recognize, interpret, and process natural language opinions over the Web. SenticNet was developed through the ensemble application of graph-mining and dimensionality-reduction techniques over multiple common sense knowledge bases [9]. It has been exploited for the development of applications in fields such as social media marketing, human-computer interaction, and e-health.

While SenticNet is much larger than WNA, it does not provide the specific emotion labels for the concepts. In this work we fill this gap by extending the WNA labels to other SenticNet concepts.

III. LEXICAL RESOURCES USED

The main aim of this work is to assign WNA emotion labels to SenticNet's concepts. For this, we use a supervised machine learning approach. The intuition behind this approach is that words that have similar features are similar in their use and meaning and, in particular, are related with similar emotions. This allows us to extend the emotion labels from the seed words for which they are known from WNA to words absent from WNA that share features with the seed words.

It is important in this respect to select linguistically relevant features, which we extracted from various relevant text corpora and dictionaries. In this section, we describe the corresponding lexical resources. In addition, we used standard resources such as WordNet 3.0.

A. *SenticNet*

As target lexicon and source of polarity information for our polarity-based concept similarity measure, we used SenticNet. This freely available dictionary assigns polarity values, in the interval from –1 to 1, to single- or multi-word concepts.

The version 1.0 [1] of SenticNet we used contains 5,732 concepts, of which 2,690 are multi-word concepts, e.g., *animate flesh dead person*, *ban harry potter*, *why happen*. Of the 5,732 SenticNet concepts, 3,303 are found in WordNet 3.0 and 2,429 are not. Of the latter set, most are multi-word concepts such as *access internet* or *make mistake*, except for 68 single-word concepts, such as *against* or *telemarketer*.

B. *The WordNet-Affect Emotion Lists*

As training corpus, i.e., the source of examples, we used the WNA lists [2] provided as part of the SensEval 2007 data. This dataset consists of six wordlists corresponding to Ekman's six basic emotions [10]: ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE; see [11] for an overview of different sets of basic emotions proposed in the literature.

The dataset contains 606 synsets, of which all but two are assigned exactly one label each. If the synsets are broken down into individual concepts, the dataset contains 1,536 concepts. Only 63 concepts are multi-word expressions, e.g., *with hostility* or *jump for joy*. All but 72 concepts (93%) are present in SenticNet vocabulary.

C. *The ISEAR Dataset*

As a source of various features and corpus-based similarity measures between concepts, we used the International Survey of Emotion Antecedents and Reactions (ISEAR) [3] dataset [12].

The survey, conducted in 1990s across 37 countries, consists of short texts called statements, obtained from approximately 3,000 respondents who were instructed to describe a situation in which they felt a particular emotion. A statement contains 2.37 sentences on average. The dataset

---

[1] http://sentic.net/senticnet-1.0.zip, downloaded on July 12, 2012.
[2] http://www.cse.unt.edu/~rada/affectivetext/data/ WordNetAffectEmotionLists.tar.gz, downloaded on July 12, 2012.
[3] http://www.affective-sciences.org/system/files/page/2636/ ISEAR.zip, downloaded on July 14, 2012. See http://www.affective-sciences.org/researchmaterial

contains 7,666 statements, 18,146 sentences, and 449,060 running words.

For each statement, the dataset contains 40 numeric or categorical parameters that give various kinds of information, such as age of the respondent, the emotion felt in the situation described in the statement, its intensity, etc. The majority of these parameters are numerical scores with a small (around 3–4) number of discrete values expressing different degrees of the parameter, such as degree of intensity of the emotion.

IV. Pre-processing and Tokenizing

A number of features we used were based on occurrences of the concepts in the ISEAR dataset. For locating concepts in the text we used pre-processing tools from Rapidminer's text plug-in;[4] for lemmatizing, WordNet lemmatizer was used. Of the 5,732 concepts contained in SenticNet, 2,729 were found at least once in ISEAR, either directly or after lemmatizing.

V. Features used for Classification

We extracted from ISEAR a number of statistical features related with occurrences and co-occurrences of SenticNet concepts (we treated multi-word concepts as a single token). We used two kinds of features: those based on the parameters provided in the ISEAR dataset and those based on various similarity measures between concepts.

A. Features Based on Parameters Provided in ISEAR

Some of the 40 parameters provided for each statement in ISEAR are not informative for our goals, such as the respondent ID. We used the following 16 parameters:

– Background data related to the respondent: age; gender; religion; father's occupation; mother's occupation; country;
– General data related to the emotion felt in the situation described in the statement: intensity; timing; longevity;
– Physiological data: ergotropic arousals; trophotropic arousals, felt change in temperature;
– Expressive behavior data: movement, non-verbal activity; paralinguistic activity;
– Emotion felt in the situation described in the statement.

Even for values expressing the degree or intensity of a parameter, we used each value of the degree (there are only 3–4 discrete values) as independent feature and the frequency of occurrences of a given concept under given value of the parameter as the value of the feature. Namely, for each occurrence of each concept in a statement of the ISEAR dataset, the corresponding parameters were extracted from the ISEAR dataset, and the data for multiple occurrences of the same concept in the whole corpus were aggregated in a feature vector for that concept as the corresponding frequencies.

B. Features Based on Similarity Measures

Another kind of feature we used for classification was various similarity values between concepts. We used two types of similarity measures: those based on lexical resources such as SenticNet and WordNet and those based on co-occurrence of concepts. The intuition behind similarity-based features is that if the distances from two data points in Euclidian space to a number of other points are similar, then it is probable that these two points are close to each other.

Given a concept, we treated as independent dimensions of its feature vector the value of each similarity measure described below, obtained by comparing this given concept with each other concept in the vocabulary.

1) SenticNet Score-Based Similarity

We define the distance between two SenticNet concepts a and b as $D_{SN}(a,b) = |p(a) - p(b)|$, where p(·) is the polarity specified for these concepts in SenticNet. The similarity is the inverse of the distance: $Sim_{SN}(a,b) = 1 / D_{SN}(a,b)$.

2) Nine WordNet Distance-Based Similarity Measures

We used English WordNet 3.0 to measure the semantic distance between two words. WordNet::Similarity[5] is an open-source package developed at the University of Minnesota for calculating nine different lexical similarity measures between pairs of word senses. This gave us nine different similarity measures.

Since WordNet similarity is defined for specific senses of words, for each pair of concepts found in WordNet, we defined the corresponding similarity as the maximum similarity between all senses of the first one and all senses of the second one.

For the concepts that were not found in WordNet, the similarity between such a concept and any other concept was set to a random value. An alternative would be using 0 or some other "not found" category, but this deteriorated the results by making all such "not found" concepts appear similar to each other because a huge number of coinciding values in their feature vectors.

3) Point-Wise Mutual Information

Point-wise mutual information (PMI) is a similarity measure based on co-occurrences of two concepts within the same sentence. For concepts a and b, it is defined as

$$Sim_{PMI} = \log \frac{p(a,b)}{p(a)\,p(b)}, \qquad (1)$$

where $p(x)$ is the probability for a sentence in the corpus to contain the concept x, i.e., the number of sentences where x occurs normalized by the total number of sentences in the corpus, and $p(x,y)$ is the probability for a sentence to contain both x and y, i.e., the normalized number of sentences that contain both x and y.

4) Emotional Affinity

We define the emotional affinity between two concepts a and b in the same way as PMI but at the level of complete statements of ISEAR, i.e., $p(x)$ in (1) above is defined as the

---

[4] http://rapid-i.com/content/view/181/190, visited November 3, 2012.

[5] http://www.d.umn.edu/tpederse/similarity.html

corresponding number of statements instead of sentences, normalized by the total number of statements.

While PMI often reflects syntactic relatedness of the words—for example, it is high for a verb and its typical object, or for parts of a frequent multi-word expression—emotional affinity incorporates a wider notion of relatedness within the same real-world situation, as well as synonymy and rephrasing. Since ISEAR statements have strong emotional contents and are related with one emotion each, it is probable that the words that co-occur in the same ISEAR statement are related with the same emotion.

5) ISEAR Text Distance Similarity

Positional information of the concept tokens in the ISEAR statements was also used to measure the similarity between concepts. For this, we calculated the average minimum distance between the pairs of tokens of SenticNet concepts that co-occurred in the statements of the ISEAR dataset. The similarity was defined as the inverse of the distance. If the concepts did not co-occur in any statement, then we considered similarity between them to be zero.

VI. CLASSIFICATION PROCEDURE

We cast the task as a 6-way categorization task, i.e., as assigning exactly one of the six WNA emotion labels to each considered concept. We conducted two sets of experiments. In one of them we took into account the features that relied on ISEAR and in the other, we did not use those features.

In the experiments that relied on features based on occurrences or co-occurrences of concepts in the ISEAR statements, only those 2,729 SenticNet concepts that were found in the ISEAR data (and thus had valid ISEAR-based features) participated in further processing and were finally assigned the emotion labels. In contrast, in the experiments without ISEAR-based features, all 5,732 SenticNet concepts were assigned the labels.

Hence, we constructed feature vectors for 2,729 and 5,732 SenticNet concepts, correspondingly. As training and test data, we used the intersection between the corresponding set of concepts and WNA vocabulary (for which we had the gold standard emotion labels); this intersection consisted of 1,202 and 1,436 concepts, correspondingly. For evaluation, the corresponding set of available labeled data was randomly divided into training and test data, using 66.7% of the set for training and 33.3% for testing. For construction of the final resource, we used 100% of available labeled data for training.

Finally, we used various machine learning algorithms trained on the training data to derive the labels for the test or unlabeled data; see Table 1.

VII. EVALUATION AND DISCUSSION

As a gold standard data for evaluation, we used concepts found in WNA. A label was considered assigned correctly to a concept if the WNA data assigned the same label to the same concept (in very few cases when the WNA assigned two labels to a concept, we considered our label assignment correct if our assigned label was one of those two labels).

Table 1: Accuracy obtained with different feature sets, data sets, and classifiers used

Feature set	S+W+I		S+W
Total data set	$S \cap I$		S
Total data size	2,729		5,732
Labeled data set	$S \cap I \cap A$		$S \cap A$
Labeled data size	1,202		1,436
Naïve Bayes	71.20%	51.23%	53.21%
Multi-layer perceptron	74.12%	55.54%	52.12%
SVM	**88.64%**	57.77%	**59.23%**

Table 1 summarizes our experiments with different feature sets, data sets, and classifiers. In this table, S stands for SenticNet, W for WordNet, I for ISEAR, and A for WNA. The S+W+I feature set stands for using the features derived from SenticNet, WordNet, and ISEAR. Only 2,729 concepts in the intersection of SenticNet and ISEAR ($S \cap I$) had valid ISEAR-based features and thus participated in those experiments. Correspondingly, S+W stands for feature vectors not containing features that rely on the ISEAR data, i.e., consisting only of the SenticNet similarity and the nine WordNet::Similarity measures between the given concept and all other concepts in the dataset. All 5,732 SenticNet concepts (S) were processed.

Accordingly, as labeled data set we used the intersection of the total data set with WNA; for ISEAR-based experiments this set ($S \cap I \cap A$) contained 1,202 labeled concepts and for experiments not relying on ISEAR, 1,436 concepts ($S \cap A$). SVM showed the best performance, obtaining an accuracy of 88.64% with the ISEAR-based features. Table 2 shows the corresponding confusion matrix.

Table 2: Confusion matrix on the test set of 396 concepts out of the 1,202 labeled ones (left: true, top: predicted).

Classified as:	S	J	S	A	F	D	precision	recall
SURPRISE	39	4	–	–	–	1	85%	89%
JOY	3	91	2	3	1	1	91%	90%
SADNESS	–	3	52	1	–	–	85%	93%
ANGER	2	2	3	72	4	2	90%	85%
FEAR	–	–	3	1	63	2	91%	91%
DISGUST	2	–	1	3	1	34	85%	83%

Comparing the first and second data columns in Table 1, which reports the experiments on the same data but with different features, shows that ISEAR-based features are very important for accurate classification in our task. The third column shows that the task would benefit from more data: increasing of the training data set improves accuracy.

In the rest of this paper, we will report the results for the corpus of 2,729 concepts obtained with the ISEAR features.

While experimenting with different subsets of features, we observed that similarity-based features performed better than the ISEAR parameter-based features [13]. Indeed, similarity measures identify whether two concepts have similar properties and thus should be placed in the same category.

SenticNet-based similarity had a positive impact. The use of all available features gave best results both when only similarity-based features were used and when ISEAR parameter-based features were taken into account [13].

Probably this is because of balancing the two sources of information: while similarity measures forced the classifier to place two concepts in a same category or, on the contrary, to separate them, the ISEAR parameter-based features provided affective information on the concept, which helped to assign a correct emotion label to it.

A. *Agreement between Polarity and Labels*

Even when the emotion label assigned by the algorithm does not coincide exactly with the one present in the gold standard data, it can share important properties with the correct label and thus can be considered correct in some relaxed sense.

Considering JOY and SURPRISE as positive emotions and the rest as negative, and ignoring the confusion within the areas marked with dotted line in Table 2, we observe a 95.71% agreement with the gold standard—the WNA data—in whether the emotion is positive or negative. Here are some examples of agreement and disagreement in the polarity of labels; two labels agree if both are positive or both are negative:

Concept	Gold standard		Our label		Agree?
frustration	ANGER	neg.	FEAR	neg.	yes
offensive	DISGUST	neg.	ANGER	neg.	yes
triumph	JOY	pos.	SURPRISE	pos.	yes
favor	JOY	pos.	ANGER	neg.	no
bored	SADNESS	neg.	FEAR	neg.	yes
wonder	SURPRISE	pos.	JOY	pos.	yes

When we used the whole set of 1,202 labeled examples used as both training and test data, agreement was 96.8%. Agreement between our results and the gold standard in affective polarity sign (negative or positive emotion) per emotion label was as follows:

Polarity	Emotion label	Agree	Disagree	Total
Positive	JOY	422	8	8
	SURPRISE	97	8	8
Negative	ANGER	198	3	3
	DISGUST	148	10	10
	FEAR	123	8	8
	SADNESS	175	2	2
	total	1,163	39	39

Here, again, by agreement we mean that both labels are positive or both negative.

B. *Agreement with the Hourglass Model*

The Hourglass of Emotions [14] reinterprets Plutchik's model by organizing primary emotions around four dimensions, whose different levels of activation make up the total emotional state of the mind:

Sentic level	Pleasantness	Attention	Sensitivity	Aptitude
1	ecstasy	vigilance	rage	admiration
2	**joy**	anticipation	**anger**	trust
3	serenity	interest	annoyance	acceptance
4	pensiveness	distraction	apprehension	boredom
5	**sadness**	**surprise**	**fear**	**disgust**
6	grief	amazement	terror	loathing

Based upon the Hourglass model, SenticNet 2 [15] aims to assign one primary emotion and one secondary emotion to each SenticNet concept.

If we map the six emotion labels used in our research (shown above in boldface) to the affective dimensions of the Hourglass model, then JOY and SADNESS are mapped to the same dimension, and so do ANGER and FEAR. Ignoring these differences—that is, ignoring the confusion within the areas marked with dashed line in Table 2—we have a 91.16% agreement with the gold standard in identifying the affective dimensions, or 93.5% when all labeled data were used as both training and test set.

C. *Agreement with the Sign of the SenticNet Polarity Score*

A possible way of an indirect evaluation of our algorithm on non-WNA data is comparing the polarity of our assigned labels (considering JOY and SURPRISE as positive and other labels as negative) with the sign of the polarity score given in SenticNet. Here are some examples:

Concept	SenticNet Score	Our Label	Agree?
better grade	–0.185	JOY (positive)	no
collect information	–0.309	JOY (positive)	no
common sense	+0.588	SURPRISE (positive)	yes
dislike	–0.917	DISGUST (negative)	yes
efficiency	–0.136	SURPRISE (positive)	no
rescue	+0.963	JOY (positive)	yes

We obtained a 95.30% agreement. Note that sometimes the disagreement can be attributed to a possible problem with SenticNet score rather than with our algorithm, as in the case of *better grade* or *efficiency*.

VIII. STATISTICS OF THE DEVELOPED RESOURCE

We obtained two resources: one with the use of ISEAR data and the other without it. The former one is smaller (2,729 vs. all 5,732 SenticNet concepts) but more accurate. Table 1 shows the lower bounds on the accuracy of each resource: 88.64% and 59.23%, accordingly. In the sequel, we discuss only the former, more accurate but smaller, resource.

The resource contains 2,729 SenticNet concepts along with their polarity scores and emotion labels, e.g.:

Concept	Polarity	Label
annoyed	–0.755	ANGER
birthday	+0.309	JOY
December	+0.111	JOY
feel guilty	–0.540	SADNESS
make mistake	–0.439	SADNESS
weekend	+0.234	JOY

Distribution of the concepts per emotion label is as follows:

Label	Concepts	Label	Concepts
ANGER	421	JOY	718
DISGUST	334	SADNESS	426
FEAR	303	SURPRISE	527
		total	2,729

IX. CONCLUSIONS AND FUTURE WORK

SenticNet is an important lexical resource for opinion mining and sentiment analysis. We have automatically supplied SenticNet with affective labels compatible with WordNet-Affect (WNA), using a machine learning algorithm. On a large subset of SenticNet concepts, the accuracy of our algorithm was 88.64%. We used various features extracted from ISEAR, an emotion-related dataset, as well as similarity measures that rely on the polarity data provided in SenticNet, those based on WordNet, and ISEAR distance-based measures, including point-wise mutual information, and emotional affinity. The developed resource is currently the largest freely available dictionary for opinion mining and sentiment analysis containing both quantitative polarity scores and qualitative affective labels.

The main contributions of this paper include the development of a new resource for opinion mining and sentiment analysis: SenticNet with emotion labels; information on how ISEAR could be leveraged in the opinion mining domain, suggestion of various features and measures useful for the task, and a general methodology for extending WNA and similar resources.

The work opens a number of directions for future research, such as using other types of monolingual or multilingual [16] corpora as a source of features to improve the accuracy or to label a greater number of concepts, as well as the use of more elaborated classification techniques in combination with syntactic and psychological clues to improve accuracy. It is also interesting to see how ISEAR information on gender, country, etc. affects the results.

REFERENCES

[1] Strapparava, C., and Valitutti, A. 2004. WordNet-Affect: An affective extension of WordNet. In: Proc. of LREC.

[2] Esuli, A., Sebastiani, F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In: Proc. of LREC.

[3] Cambria, E., Speer, R., Havasi, C., Hussain, A. 2010. SenticNet: A publicly available semantic resource for opinion mining. In: Proc. of AAAI CSK, 14–18.

[4] Mohammad, S., Turney, P.D. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In: Proceedings of the NAACL-HLT, Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, 26–34.

[5] Arora, P., Bakliwal, A., Varma, V. 2012. Hindi subjective lexicon generation using WordNet graph traversal. International Journal of Computational Linguistics and Applications, 3(1):25–39.

[6] Wawer, A. 2012. Extracting emotive patterns for languages with rich morphology. International Journal of Computational Linguistics and Applications, 3(1):11–24.

[7] Cambria, E., Hussain, A. 2012. Sentic computing: Techniques, tools, and applications. Dordrecht, Netherlands: Springer, 153 pp.

[8] Picard, R. 1997. Affective computing. The MIT Press.

[9] Cambria, E., Olsher, D., Kwok, K. 2012. Sentic activation: A two-level affective common sense reasoning framework. In: Proc. of AAAI, 186–192.

[10] Ekman, P. 1993. Facial expression and emotion. American Psychologist, 48(4):384–392.

[11] Ortony, A., Turner, T. J. 1990. What's basic about basic emotions? In: Psychological Review, 97, 315–331.

[12] Scherer K.R. 2005. What are emotions? And how can they be measured? In: Social Science Information, 44(4):693–727.

[13] Poria, S., Gelbukh, A., Cambria, E., Yang, P., Hussain, A., Durrani, T. 2012. Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis. In: Proc. of ICSP.

[14] Cambria, E., Livingstone, A., Hussain, A. 2012. The Hourglass of Emotions. LNCS, vol. 7403, 144–157.

[15] Cambria, E., Havasi, C., Hussain, A. 2012. SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. Proceedings of FLAIRS, 202–207.

[16] Sidorov, G., Posadas-Durán, J.-P., Jiménez Salazar, H., Chanona-Hernandez, L. 2011. A new combined lexical and statistical based sentence level alignment algorithm for parallel texts. International Journal of Computational Linguistics and Applications, 2(1–2):257–263.

Soujanya Poria is a student at the Department of Computer Science and Engineering of the Jadavpur University, India, and author of various publications in Natural Language Processing and Machine Learning; see www.soujanyaporia.in.

Alexander Gelbukh, PhD, is a Research Professor at the Computing Research Center of the National Polytechnic Institute, Mexico, and author of ca. 500 publications in Natural Language Processing; see www.gelbukh.com.

Amir Hussain, PhD, is a Professor of Computing Science at the University of Stirling, UK, and author of more than 200 publications in computational intelligence and cognitive computing technology and applications; see cs.stir.ac.uk/~ahu.

Dipankar Das, PhD, is an Assistant Professor at the Department of Computer Science and Engineering, National Institute of Technology (NIT), Meghalaya, India, with research interests in Natural Language Processing; see www.dasdipankar.com.

Sivaji Bandyopadhyay, PhD, is a Professor at the Department of Computer Science and Engineering of the Jadavpur University, India, and author of more than 100 publications in Natural Language Processing and Machine Learning; see www.sivajibandyopadhyay.com.