

Enriching SenticNet Polarity Scores Through Semi-Supervised Fuzzy Clustering

Soujanya Poria¹, Alexander Gelbukh², Erik Cambria³, Dipankar Das¹, Sivaji Bandyopadhyay¹

¹Computer Science and Engineering Department, Jadavpur University, India

²Center for Computing Research, National Polytechnic Institute, Mexico

³Temasek Laboratories, National University of Singapore, Singapore

soujanya.poria@gmail.com, gelbukh@gelbukh.com, cambria@nus.edu.sg

dipankar.dipnil2005@gmail.com, sbandyopadhyay@cse.jdvu.ac.in

Abstract—SenticNet 1.0 is one of the most widely used freely-available resources for concept-level opinion mining, containing about 5700 common sense concepts and their corresponding polarity scores. Specific affective information associated to such concepts, however, is often desirable for tasks such as emotion recognition. In this work, we propose a method for assigning emotion labels to SenticNet concepts based on a semi-supervised classifier trained on WordNet-Affect emotion lists with features extracted from various lexical resources.

Keywords—SenticNet, Sentic computing, Sentiment analysis, WordNet, WordNet-Affect, ISEAR dataset, Fuzzy clustering

I. INTRODUCTION

Opinion mining is a rapidly developing area of natural language processing that has received growing attention from both research community and industry in recent years. It helps a company to know what customers feel about its products, it helps a political party or government to know what the voters feel about their actions and proposals. On the other hand, it helps customers or voters to choose wisely and in an informed way by knowing what their peers feel about a product or a political candidate. With this, opinion mining is of great importance for aiding economy and democracy.

The basic “feeling” about something can be described on the scale of approval vs. disapproval, good vs. bad, positive vs. negative. Such a scale is termed *polarity*. The basic question of whether customers or voters are satisfied with a product or actions can be answered by detecting the average polarity of their comments. For this purpose, one of the most widely used resources for concept-level opinion mining is SenticNet [1]. For about 5,700 multiword expressions, SenticNet provides a quantitative polarity score in the range between -1 (bad) to $+1$ (good), with neutral scores being around zero; e.g., *aggravation*: -0.925 (bad), *accomplish goal*: $+0.967$ (good), *December*: $+0.111$ (neutral).

This allows detecting from natural language concepts if a customer review is, say, 0.567 positive or 0.876 negative. However, more detailed information is often desirable. Citizens dislike the governing party—do they specifically feel anger, fear, or disgust? Customers like a product—is this predominantly because of joy or surprise?

One of the main lexical resources currently available for detecting emotional labels in text is WordNet-Affect (WNA) [2], which indicates whether a given word is related to one or more of six basic emotions: ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE. WNA, however, contains a relatively small number of words.

In this work, we propose a method for automatically assigning quantitative WNA emotion scores to a significant subset of the SenticNet entries. The paper is organized as follows: Section II discusses related work in the field of publicly available resource for opinion mining; Section III provides an overview of the proposed method; Section IV introduces the lexical resources used in this work; Section V illustrates the corresponding features used for classification; Sections VI, VII, and VIII describe the three main steps of the algorithm: fuzzy clustering, mapping of the obtained clusters to labels, and final hard categorization; Section IX explains the data preparation for the experimental evaluation, which is presented in Section X; Section XII, finally, presents conclusions and future work.

II. RELATED WORK

Attempts to build affective corpora are countless and comprehend both resources crafted by human experts or community efforts and automatically-built lexicons.

A. Emotion Lexicons

Alm et al. [3] built a supervised emotion lexicon upon a corpus of ca. 180 children stories, such as those authored by Grimms, Andersen, and Potter. For classification, they extracted several lexical and stylometric features such as upper case words, first sentence of the story, part-of-speech (POS), etc., achieving 69% accuracy.

Yang et al. [4] built an emotion lexicon from a weblog corpus by using a collocation model to learn an emotion lexicon from weblog articles. For determining emotions, a classification based on a support vector machine (SVM) was exploited.

Banea et al. [5] used a rule-based approach for subjectivity lexicon generation, achieving 69% accuracy as well, which represented an 18% improvement over their previous semi-supervised approach.

Their method allows acquiring a large subjectivity lexicon by bootstrapping from a few manually selected seeds. In the present work, we achieve slightly better accuracy than the best one of the mentioned works, Yang et al., by using much less data.

B. Sentic Computing and SenticNet

Currently available lexical resources for affect recognition such as SentiWordNet [6] or WNA are known to be rather noisy and limited. These resources mainly provide emotion labels at syntactical-level, leaving out polarity and affective information for common sense knowledge concepts such as *accomplish goal*, *bad feeling*, *celebrate special occasion*, *lose temper* or *be on cloud nine*, which are usually found in natural language text to express viewpoints and affect. In contrast, recent approaches to sentiment analysis aim for a semantic-level analysis of natural language opinions.

Sentic computing [7], for example, is a multidisciplinary approach to sentiment analysis that exploits both computer and social sciences to better recognize, interpret, and process opinions at concept-level. Developed under the umbrella of sentic computing, SenticNet is a lexical resource that associates polarity values with 5,700 common sense knowledge concepts. The efficiency of such a resource over other word-based knowledge bases has been extensively demonstrated through the development of applications in fields such as Social Web, human-computer interaction (HCI), and e-health [7].

III. OVERVIEW OF THE PROPOSED METHOD

The proposed method consists in assigning emotion labels to a significant subset of SenticNet concepts, namely, to those concepts for which we have all the necessary features from other lexical resources. Labels can be assigned to a greater number of concepts with lower accuracy, if fewer features are used. Specifically, the proposed method classifies concepts into ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE, by extending the labels from the seed concepts, for which they are known, to other concepts, using a variety of features extracted from a number of lexical resources.

We use a semi-supervised approach. First, concepts are clustered into six clusters in an unsupervised manner. In particular, a fuzzy clustering methodology is adopted: each concept belongs to each of the six categories to a different degree, termed *membership value*. Then, such clusters are mapped to specific emotion labels (fuzzy categorization).

It would be tempting to just pick for each concept the category with the maximum membership value as the final classification output. We show, however, that sometimes picking the second-best one significantly improves the accuracy. To choose between the top-2 categories, we use a classifier trained for each pair of emotions: for example, we train a separate classifier to distinguish between ANGER and DISGUST, another one for ANGER vs. FEAR, etc.

In addition to the same set of features used for the fuzzy clustering, such classifiers use the membership values from

the fuzzy clustering as six additional features, which further improves accuracy. We tried using top- K categories instead of top-2, with $K = 1, \dots, 6$, but $K = 2$ gave the best result by a large margin (Table IV). Note that while $K = 6$ does not restrict the confusion set for the hard classifier at all, the classifier would still use the membership values as six additional features.

IV. LEXICAL RESOURCES USED

Our goal is to assign WNA emotion labels to the concepts that appear in the SenticNet dictionary using, among others, corpus-based similarity measures between concepts. In this section, we describe in detail the lexical resources exploited.

A. SenticNet

As the target lexicon and the source of polarity information for our polarity-based concept similarity measure, we used SenticNet, which has been developed through an ensemble application of graph-mining and dimensionality-reduction techniques [8].

SenticNet 1.0¹ contains 5,726 concepts, of which 2,690 are multi-word concepts, e.g., *buy Christmas present*, *animate flesh dead person*, *ban harry potter*. Of the concepts in SenticNet 1.0, 3,303 are also present in WordNet 3.0. The remaining concepts are mostly multiword concepts such as *access internet* or *make mistake*, except for 68 single-word concepts, such as *against* or *telemarketer*. However, some of those 68 single-word entries that are found in SenticNet 1.0 but not in WordNet seem to contain typos, such as *alcohol*, *exercize*, *baloon*, *happine*, or *gossipe*.

B. WordNet-Affect Emotion Lists

As an inventory of the target labels and a source of training examples for the supervised classification, we used the emotion lists² extracted from WNA provided for the SemEval 2007 task 14: Affective text. These are six lists corresponding to the six emotion labels mentioned above.

This dataset assigns emotion labels to *synsets*—groups of words or concepts that are synonymous in the corresponding senses: e.g., a synset *{puppy love, calf love, crush, infatuation}* is assigned the label JOY. We, however, ignored the synonymy information contained in the data and used the labels for individual words or concepts, i.e., we used *puppy love* → JOY, *calf love* → JOY, *crush* → JOY, *infatuation* → JOY. Statistics of the synsets or concepts by label is given in Table I.

The dataset contains 532 different synsets, of which two are assigned to two distinct labels each: *{cliff-hanging, suspenseful, suspensive}* → {fear, joy} and *{persecute, oppress, harass}* → {anger, sadness} (thus the numbers of synsets in Table I sum up to 534).

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

TABLE I.
The number of synsets and concepts by emotion label

	Synsets	Concepts
JOY	204	400
ANGER	115	255
SADNESS	95	202
FEAR	76	147
SURPRISE	27	71
DISGUST	17	53
total:	534	1,128

If the synsets are broken down into individual concepts (single- or multi-word expressions), the dataset contains 1,113 concepts, of which 15 are assigned two distinct labels each (thus the total of 1,128 in Table I). Apart from the words from the two ambiguous synsets, these are words that enter in different unambiguous synsets (in different senses), e.g., *awful* \rightarrow {FEAR, SURPRISE}; when breaking synsets into individual concepts, we lose the sense disambiguation information.

Of the concepts, 63 are multi-word expressions, e.g., *with hostility* or *jump for joy*; all of them are unambiguous, i.e., are assigned only one label. Only 42 synsets contain multi-word concepts. Of the concepts included in the lists, all but 72 (93%) are contained in the SenticNet dictionary. Stemming reduces the number of exceptions to 46; however, it is not always clear to which SenticNet concept with the same stem a stemmed word should be mapped.

By considering the emotions JOY and SURPRISE as positive and ANGER, DISGUST, FEAR, and SADNESS as negative, we can assign binary polarity to the listed concepts. The number of concepts is then as follows:

positive	464
negative	643
ambiguous	6

We count a concept as ambiguous if it has two labels with contradicting polarity, e.g., *suspensive* \rightarrow {JOY, FEAR}. Of the 1,041 concepts present both in the SenticNet and in the WNA lists, 5 have ambiguous WNA polarity, and the rest is distributed according to the following confusion matrix:

In WordNet-	In SenticNet	
	positive	negative
positive	417	8
negative	11	601

This shows a very good (98.1%) agreement between WNA lists and SenticNet as to the sign of the polarity. In some cases of disagreement, the polarity assigned by the WNA lists seems to be inappropriate, e.g., *worry* appears under JOY (actually, as a synonym of ‘interest’), or *stupid* under SURPRISE (as a synonym of ‘dazed’). In other cases of disagreement, the SenticNet’s scores seem to be somewhat counterintuitive, such as -0.503 for *satisfy* or $+0.875$ for *greedy*.

C. The ISEAR Dataset

As a source of various features and similarity measures between concepts, we used the International Survey of Emotion Antecedents and Reactions (ISEAR)³ dataset [9]. The survey was conducted in 1990s across 37 countries and had approximately 3,000 respondents.

The respondents were instructed to describe a situation or event in which they felt a particular emotion, in the form of a *statement*—a short text of a couple of sentences (2.37 on average). Here is an example of a complete statement:

I had the window open and the door was shut so that the kitten would not go out. My partner came in and started talking about something and I forgot about the window and suddenly I saw the kitten hanging from the window frame. I was rigid with fright till I got hold of her.

The dataset contains 7,666 such statements, which include 18,146 sentences, 449,060 running words. Of the 5,732 concepts contained in SenticNet, 2,729 were found in ISEAR. Since only for those concepts we have important features extracted from ISEAR, only those 2,729 concepts were assigned emotion labels in this work.

The statements in the ISEAR dataset are arranged in a table, with 40 additional numeric or categorical columns that give various kinds of information for each statement. Some of these columns are not informative for our goals, such as the statement ID, the respondent ID, etc. Of the 40 data columns, we carried out experiments with just 16 columns, presented in Table II for convenience of discussion in 5 groups.

TABLE II.
The data columns of ISEAR dataset used in this work

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

The majority of these parameters (except age, gender, religion, country, and emotion) are numerical scores with a small (around 3–4) number of discrete values expressing different degrees of the parameter. For example, the values for ergotropic arousal are: 1 for change in breathing, 2 for heart beating faster, 3 for muscles tensing / trembling, 4 for perspiring / moist hands; the values for trophotropic arousal are: 1 for lump in throat, 2 for stomach troubles, 3 for crying / sobbing. One of the columns gives the name of the emotion felt.

³ <http://www.affective-sciences.org/system/files/page/2636/ISEAR.zip>, downloaded on July 14, 2012. Linked from <http://www.affective-sciences.org/researchmaterial>

Seven values are used: ANGER, DISGUST, FEAR, GUILT, JOY, SADNESS, and SHAME; for example, the example statement given above is labeled as FEAR. This set of seven emotions is different from our target set of six basic emotions used in WNA lists (ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE).

We do not directly use the ISEAR emotion labels to assign out target emotion label, but do use them as one of feature types for our classification, alongside with many other features. The choice of ISEAR as the source of corpus-based information is motivated by the following considerations:

- ISEAR corpus is particularly rich in emotion-related words as compared with more standard corpora used in natural language processing;
- ISEAR statements are consistent as to emotion contents. Thus, for its statements what we call (rephrasing Gale et al. [10]) “one emotion per discourse” principle is applied: if two expressions co-occur in the same statement, they are related with the same emotion.

In our sample statement the concepts *window open*, *forget*, *suddenly*, *hang*, *rigid with fright* are all associated with the same emotion, FEAR. This property makes the ISEAR database particularly suitable for co-occurrence-based emotion similarity measures between concepts.

V. FEATURES FOR SEMI-SUPERVISED CLASSIFICATION

The data items that we categorized into emotion classes are concepts, i.e., single- or multi-word expressions present in the SenticNet vocabulary. For our classification, we use two kinds of features of such concepts: those based on the data columns provided in the ISEAR dataset (averaged by the occurrences of a concept as a token in the ISEAR statements), and those based on various similarity measures between concepts.

A. ISEAR data-based features

We used the 16 ISEAR data columns listed in Table II as independent features. We treated all the features as categorical features stated in Table III with the following specification. For example, the *country* column has 16 different numerical codes (different countries are represented by different numeric codes), so we used 16 different features corresponding to each specific country.

Hence, the feature set dimension of country is 16. As for the value, we used term frequency: if the concept occurs in the ISEAR dataset 3 times under country code 1 and 5 times under country code 2, then the corresponding part of the feature vector was (... , 3, 5, ...).

The values expressing the degree or intensity of various parameters were, for simplicity, treated in the same way; there are around 3–4 discrete values per such data column in the ISEAR dataset.

We did not use numeric data types for the values to avoid problems in aggregating (e.g., averaging) values for different occurrences of the same concept. This gave us about 100 categorical features, which were used as different dimensions of the feature vector.

B. Features based on similarity measures

Another kind of features is given by similarity measures between concepts. If we can measure somehow the similarity between pair of concepts, this gives us N distinct dimensions of the feature vector: the similarity between the given concept and each concept in the vocabulary; here N is the total number of the concepts we considered (2,729 in our case, as explained in Section IV). The intuition behind the approach 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.

To define such similarity-based features for classification, we used the following 13 similarity measures: 10 lexical resources-based measures—SenticNet score-based similarity and 9 WordNet-based similarity measures—and 3 co-occurrence-based measures—text distance-based similarity, point-wise mutual information, and emotional affinity, as described below.

This gave us other $13N$ dimensions in the feature vectors. The 3 co-occurrence-based measures (and in fact some of the WordNet similarity measures that incorporate corpus-based co-occurrence information) are highly correlated but still reflect different granularity levels of the text, so we decided to include all of them in the feature vectors.

1) SenticNet score-based similarity

The distance between two concepts a, b found in SenticNet is defined as $D_{SN}(a,b) = |p(a) - p(b)|$, where $p(\cdot)$ is the polarity specified for the concepts in SenticNet; the similarity is the inverse of the distance: $Sim_{SN}(a,b) = 1 / D_{SN}(a,b)$.

All concepts that we used were taken from SenticNet and thus have valid SenticNet scores.

2) WordNet distance-based similarity

The open-source package WordNet::Similarity⁴ [11] with English WordNet 3.0 was used to measure the distance between tokens. This package provides 9 similarity measures based on the analysis of the WordNet hierarchy, glosses, and other data present in WordNet:

In this work, we used all the 9 measures as independent sources of information, corresponding to its own N dimensions each in the feature vectors. As mentioned in Section IV, not all concepts from SenticNet 1.0 are present in WordNet 3.0. Of those 2,729 concepts that were found in the ISEAR database and with which we worked, 1,436 were directly found in WordNet.

Those that were not found were examined manually and rephrased. For example, if a multi-word concept contained a main word, such as in *make mistake*, it was reduced to this word, *mistake* this case. After this, we could map 169 more concepts to WordNet, so in total 1,605 concepts received meaningful WordNet similarity scores. For the remaining 1,124 concepts that we could not map to WordNet, the similarity between such a concept and any other concept was set to a random value in the interval $\{0, 1\}$.

⁴ <http://www.d.umn.edu/tpederse/similarity.html>

We did not set those values to 0 or some other fixed value because this would make all concepts not found in WordNet far from other concepts and very similar to each other, about 70% of coordinates ($9N$) in their feature vectors being exactly equal (we did check that doing so did deteriorate the final results). In contrast, setting them to random values better expresses the idea of unknown similarity: such vectors lie in the feature space rather far from all others and, most importantly, far from each other.

Since all 9 similarity scores are defined for specific senses and not for character strings, for a pair of concepts found in WordNet, the similarity was defined as the maximum similarity between all senses of the first one and all senses of the second one.

3) ISEAR text distance-based similarity

The positional information of the concept tokens in the ISEAR statements was used to measure the similarity between them. Suppose, if the tokens a and b occur in a statement S as words number a_1, \dots, a_n and b_1, \dots, b_n , correspondingly, then the distance between them in this statement is defined as $D_{ISEAR}(a, b, S) = \min(|a_i - b_j|)$, and the distance over the entire ISEAR dataset is defined by averaging over individual statements S_k : $D_{ISEAR}(a, b) = \text{avg } D_{ISEAR}(a, b, S_k)$. Note that if the two words appear as a bigram, then $D_{ISEAR}(a, b, S) = 1$.

Now, the similarity is defined as the inverse of the distance: $Sim_{ISEAR}(a, b) = 1 / D_{ISEAR}(a, b)$. If the concepts do not co-occur in any statement, then we considered $Sim_{ISEAR}(a, b) = 0$.

4) Point-wise Mutual Information

The point-wise mutual information (PMI) between concepts measures the degree of co-occurrence between them within a sentence. For concepts a and b , it is defined as

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

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

5) 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, i.e., $p(\cdot)$ in (1) is defined as the 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 multiword expression—emotional affinity incorporates a wider notion of relatedness within the same real-world situation, as well as synonymy and rephrasing. Due to our “one emotion per discourse” principle for the ISEAR dataset, the concepts with high emotional affinity tend to be related with the same emotion.

Similarly to previously adopted clustering techniques [12], the first step in our process is unsupervised: we do not use the emotion labels known for some concepts from the WordNet Affinity list. We consider fuzzy clusters, so that a concept can belong to several clusters with different degree of membership. Consequently, we used fuzzy clustering to determine the preliminary affinity for a concept to be related with each group.

The features we used for the fuzzy clustering were those described in Section V, and the result of the clustering, for each concept and each of the six groups, was the membership function of a given concept in the given class defined by the given emotion label—a number in the range between 0 and 1 representing our estimation of the probability for the given token to express the given emotion.

A. Fuzzy C-means Clustering Algorithm

For fuzzy clustering, we used the fuzzy c-means clustering algorithm [13] with a modified objective function as described in Section B below.

The well-known fuzzy c-means clustering algorithm takes as input a set of N of *points* x_1, x_1, \dots, x_N described via their coordinates in a P -dimensional *feature space*: $x_k = (x_{k1}, x_{k2}, \dots, x_{kP})$. As output, it constructs two sets: a set of c centroids v_1, v_2, \dots, v_c , which are points in the same feature space that represent the c clusters its found (c is a given parameter), and a set of cN *membership functions* $\mu_{ik}, i = 1, \dots, c; k = 1, \dots, N$, which represent the *degree of membership* of a point x_k in a class c_i .

The membership function can be interpreted as the share of a point that a cluster has, or as a probability for the point to belong to a cluster, so it is considered that $0 \leq \mu_{ik} \leq 1$ and they sum up to a unity for each point:

$$\sum_{i=1}^c \mu_{ik} = 1, \quad k = 1, \dots, N. \quad (2)$$

To find the optimal distribution of points by clusters and optimal placement of the centroids, it uses an given *objective function* J , which is minimized when the distribution is optimal: $(\mu_0, v_0) = \arg \min J(\mu, v)$, where $\mu = \{\mu_{ik}\}$ and $v = \{v_i\}$ represent the sets of the variables to be found and μ_0, v_0 are the optimal solutions. An expression often used for J is

$$J_p(\mu, v) = \sum_{i=1}^c \sum_{k=1}^N \mu_{ik}^p \|x_k - v_i\|^2, \quad (3)$$

where the power $p > 1$ is a given parameter that controls the degree of fuzziness of the obtained clusters (we used $p = 2$), and

$$\|x - y\| = \sum_{i=1}^P (x_i - y_i)^2 \quad (4)$$

is the Euclidean distance in the feature space.

The optimal solution of the constraint optimization problem defined by (2) and (3) is given by the following formula [14]:

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^p x_k}{\sum_{k=1}^N \mu_{ik}^p}, \quad \mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{p-1}}} \quad (5)$$

Now, a stationary point (μ_0, v_0) of the system (5) is found by iterations:

- Start from assigning random values to all μ_{ik} ; the values are normalized to satisfy the constraints (2).
- Iteratively re-calculate the values for all v_i and then all μ_{ik} according to (5);
- Stop when the objective function J changes from the previous iteration less than by a small number ε , a given parameter (we used $\varepsilon = 0.01$).

B. Modified Objective Function

To achieve more compact clusters in which the most similar elements are clustered together, we incorporated an additional term in the original objective function (3):

$$J_p(\mu, v) = \sum_{i=1}^c \sum_{k=1}^N \mu_{ik}^p \left(\|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right), \quad (6)$$

where the parameter ρ is intended to control the effect of the new term (we used $\rho = 1$) and N_k is the set constructed in the following way:

- For each data point x , we identified the nearest centroid
$$v(x) = \arg \min_i \|v_i - x\| \quad (7)$$
(in case of a tie an arbitrary one was chosen);
- Now, $N_k = \{x \mid v(x) = v(x_k)\}$ is the set of all data points with the same nearest centroid as x_k .

This additional term hints the algorithm to increase the membership function of a data point in the cluster with the nearest centroid thus better grouping similar point together.

In our implementation we constructed these sets on the fly while re-calculating the positions of the centroids according to (8) below, which is a modification of (5). That is, when re-calculating v_2 , we considered in (7) already re-calculated value for v_1 . The change of the expression for the objective function required modification of the formulas (5):

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^p \left(x_k + \rho \sum_{x \in N_k} x \right)}{\sum_{k=1}^N \mu_{ik}^p (1 + \rho |N_k|)}, \quad (8)$$

and

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2}{\|x_k - v_j\|^2 + \rho \sum_{x \in N_k} \|x - v_j\|^2} \right)^{\frac{1}{p-1}}}; \quad (9)$$

VII. MAPPING FUZZY CLASSES TO EMOTION LABELS

After the fuzzy clustering process had been completed, we identified which one of the $c = 6$ classes corresponded to which one of the six emotion labels. For this, first we converted the fuzzy clustering into hard clustering; in our implementation we chose for each data point x_k a cluster $C(x_k) = \arg \max_i \mu_{ik}$ (in case of a tie, arbitrary class was chosen).

The emotion label for each hard cluster was chosen by majority voting. Namely, we considered all concepts in the cluster that were in the emotion lists and thus had their emotion labels known; calculated the number of times each label occurred (for those concepts that had more than one emotion label in the emotion lists, all their labels were counted), and chose a label that occurred the greater number of times.

This procedure does not guarantee for a hard cluster to be not empty or for the majority voting not to result in a tie, or that two clusters do not share the same emotion label, in which case some labels would not be assigned at all. However, this is low probable and did not happen in our experiments.

Moreover, correctness of the obtained mapping of the classes to emotion labels is confirmed by the fact that we obtained over 90% accuracy of our final results; this is not possible with incorrectly mapped labels.

VIII. CLASSIFICATION

While several concepts appear in more than one WNA list (for example, *harass* is listed under SADNESS and ANGER, *suspensive* under JOY and FEAR), the great majority most of them have only one emotion label. Predicting whether a word is ambiguous is outside the scope of the present paper, and we are not even sure if such ambiguity was not an error in the WNA lists. Therefore, to simplifying things, we chose to assign only one emotion label to each concept.

In our evaluation, we consider a label to be assigned correctly if WNA lists assign this label to the concept—even if it also assigns another label to it. There are too few cases of double labels present in the lists for this decision to significantly affect our results.

To choose only one class for a token under classification, we use a two-step process.

C. Reducing the confusion set

For each data point, we chose K classes for which the fuzzy clustering gave the highest value of the membership function. The hard clustering technique used afterwards was only allowed to choose between those K labels pre-selected for a given concept.

In case of $K = 1$ no further processing is needed and the final result are determined by the greatest membership function of the fuzzy clustering. In case of $K = 6$, no reducing of the confusion set happens and this step is in effect omitted. In case of $K = 2$ or $K = 3$ —the values with which we experimented—the confusion set is reduced to 2 or 3 options, correspondingly. We show in Section X that reducing the confusion set to 2 candidates increased the accuracy.

D. Final hard categorization by supervised classification

Given the K options left after reducing the confusion set, we trained different classifiers for each of the $\binom{6}{K}$ possible combinations of K emotion labels: for example, with $K = 2$, a separate classifier was trained for choosing between FEAR and DISGUST, another one to choose between FEAR and SADNESS, etc. For $K = 2$, therefore, 15 different classifiers were trained. To assign a label to data point, the K emotion labels for the point are selected as explained above, e.g., FEAR and SADNESS, and then the corresponding classifier is used.

For training, only those concepts from the WNA lists we used for a given classifier that have any of the corresponding labels in the lists. For example, to train a classifier for the confusion set {FEAR, SADNESS}, all concepts extracted from the ISEAR corpus that are in the WNA lists and have their either the label FEAR or the label SADNESS were used (the few concepts with double labels, such as *harass*, were excluded from the training data for the confusion sets containing both labels).

As features, we used the same feature vectors as for fuzzy clustering, extended by 6 extra dimensions: the membership values generated by the fuzzy classifier for the 6 emotion labels, except the experiments where the fuzzy clustering was not used. As classes, the K selected labels were used for each classifier; in case of $K = 2$ the classification was binary.

As a supervised classification algorithm, we used the SVM framework. Specifically, we used the `libsvm` library of the WEKA toolset, which, for the case of $K > 2$, provides an implementation of a multiclass SVM. As a result, we obtained one emotion label for each concept in the dataset.

IX. DATA PREPARATION

A set of standard pre-processing techniques such as tokenizing and lemmatizing were used in the process as described below. For this, we used the tools provided by Rapidminer’s text plug-in, except for lemmatizing, for which the WordNet lemmatizer was used (a lemmatizer differs from a stemmer in that it provides a complete form: for *feet*, it provides *foot*).

For each SenticNet concept, we identified all its occurrences in the text of ISEAR statements. For multi-word SenticNet concepts, such as *after summer*, *to person*, etc., we allowed any number of stop-words to appear the position of the space, so that in the SenticNet vocabulary matched, e.g., *to a person* or *to the person* in the text.

Lemmatizing both the SenticNet concepts and the ISEAR text before matching would generate some number of false matches, while not using any lemmatization would result in many words in the text to fail to match with the concepts because they appear in the text in a different form, e.g., *made mistake* in the text vs. *make mistake* in the vocabulary.

To minimize both undesired effects, for each concept in the SenticNet vocabulary and for each ISEAR statement, we first tried to find the concept in the statement; if zero occurrences of the concept were found in this particular statement, then we lemmatized both the concept and all words in the statement and repeated the search.

A total of 2,729 SenticNet concepts were found at least once in the ISEAR dataset. Only these concepts participated in further processing and were finally assigned the emotion labels.

X. EVALUATION

As training and test data, we took the intersection between the sets of concepts found in the WNA lists (for which we had the gold standard emotion labels) and those 2,729 SenticNet concepts found in the ISEAR texts (for which we had ISEAR-specific features); this intersection consists of 1,202 concepts.

A. Impact of different features

Table III presents the result of ten-fold cross validation with different subsets of features. The method achieves 92.15% accuracy when all features are used.

The results are given for classification based on the maximum membership function of the fuzzy classifier without the need for the hard classifier ($K = 1$); SVM hard categorization without the previous fuzzy clustering ($K = 6$ and not using the fuzzy membership functions as additional features); and the complete algorithm as presented in this paper.

Left-hand side columns of Table III correspond to the following feature sets described in Section V: S stands for SenticNet similarity; W for WordNet similarity; T for ISEAR text similarity; A to D for the groups of ISEAR features described in Table II; F for the membership values of the fuzzy clustering used as additional six features for the hard clustering (then SVM-only classification is not possible).

TABLE III
Accuracy with different feature combinations
and different classifier combinations

Feature Combination								Fuzzy	SVM	Fuzzy +SVM
S	W	T	A	B	C	D	F			
×								25.27	29.07	32.23
	×							47.66	50.59	52.02
×	×							56.64	57.77	61.21
		×						24.12	27.87	31.09
×		×						25.23	28.21	33.01
	×	×						41.32	43.09	45.15
×	×	×						58.11	59.27	63.50
×	×	×	×					69.78	73.23	78.26
×	×	×		×				63.18	65.12	66.04
×	×	×			×			59.17	60.22	65.77
×	×	×				×		55.34	58.00	61.10
×	×	×	×	×				61.02	61.12	62.02
×	×	×	×		×			60.15	60.24	61.00
×	×	×	×			×		58.12	58.45	58.75
×	×	×		×	×			55.77	56.34	58.14
×	×	×		×		×		65.35	66.12	69.04
×	×	×	×	×	×	×		80.11	85.12	88.85
×	×	×	×	×	×	×	×	83.41	—	92.15

B. Impact of the fuzzy clustering and hard categorization

In addition to the data presented in Table III for fuzzy-only and hard-only classifiers, we experimented with different values of K : the size of the confusion set after reduction based on the result of fuzzy clustering; see Table IV:

- $K = 1$ means that the final classification is made basing on the results of the fuzzy clustering and no further hard clustering is necessary;
- $K = 2$ means that the hard classification has to do only binary choices;
- $K = 3$ means reduction of the confusion set for the hard classification to three choices;
- $K = 6$ means no reduction of the confusion set. It is not the same as not to use the fuzzy clustering phase at all, because the fuzzy clustering results are still used as additional features for final categorization.

TABLE IV
Impact of the selection of most likely fuzzy cluster

K	Accuracy
1	83.41%
2	92.15%
3	67.45%
6	65.43%

We can see that SVM performed better on choosing between the category with the highest membership value and that of the second highest one. Here, we used all features, which corresponds to the last row of Table III.

C. Error Analysis

Two main types of errors were observed. The first one occurs when two emotion words are present in a statement in two separate sentences. Then the lexical affinity between such remote occurrences does not always carry the similarity between their lexical patterns.

The other type is related with the concepts for which the fuzzy clustering gave very similar membership functions in the top-2 clusters, e.g., *faint*, *sick*, *humble*.

XI. IMPACT OF THE NEW DEVELOPED RESOURCE

A preliminary evaluation of the developed resource has been carried out on 150 Amazon product reviews, of which 80 are positive and 70 are negative. We have treated the polarity score along with emotion of a concept as feature of a review (so, if one review contains 5 concepts that are also present in our knowledge base, we include the polarity score and emotion of those concepts in the feature vector of that review). We ran SVM on the feature file to predict the polarity(+/-) of each review. The same experiment was run by using SenticNet 1.0. Results show that the newly developed knowledge base achieves 8% better accuracy over SenticNet.

XII. CONCLUSION AND FUTURE WORKS

We have proposed a method for assigning emotion labels to a significant subset of the concepts in the SenticNet vocabulary. The proposed methodology is based on a semi-

supervised approach that uses a variety of features extracted from different lexical resources, notably the ISEAR dataset, and the SemEval 2007 emotion wordlists based on WNA, as a gold standard test and training data.

We used a novel semi-supervised method based on fuzzy clustering. We also used a novel objective function for the fuzzy clustering. When all improvements and features were used, we achieved 92.15% accuracy.

As future work, we plan to provide special treatment to the cases of most typical errors made by our algorithm. Moreover, we plan to incorporate syntactic information as additional features. Textual clues related to psychology may be included, too.

ACKNOWLEDGEMENTS

The work was partially supported by the Governments of India and Mexico under the CONACYT-DST India (CONACYT 122030) project “Answer Validation through Textual Entailment”, the Government of Mexico under the CONACYT 50206-H project and SIP-IPN 20121823 project through Instituto Politécnico Nacional, and the Seventh Framework Programme of European Union, project 269180 “Web Information Quality Evaluation Initiative (WIQ-EI).

REFERENCES

- [1] E. Cambria, R. Speer, C. Havasi, and A. Hussain, “SenticNet: A publicly available semantic resource for opinion mining,” Proceedings of AAAI CSK, 2010, pp. 14-18.
- [2] C. Strapparava and A. Valitutti, “WordNet-Affect: an affective extension of WordNet,” Proceedings of LREC, 2004, pp. 1083-1086.
- [3] C. Alm, D. Roth, and R. Sproat, “Emotions from text: Machine learning for text-based emotion prediction,” Proceedings of HLT/EMNLP, 2005, pp. 579-586.
- [4] C. Yang, H.Y. Lin, and H.H. Chen, “Building emotion lexicon from weblog corpora”, 45th Annual Meeting of ACL, 2007, pp. 133-136.
- [5] C. Banea, R. Mihalcea, and J. Wiebe, “A bootstrapping method for building subjectivity lexicons for languages with scarce resources,” Proceedings of LREC, 2008, pp. 2764-2767.
- [6] A. Esuli and F. Sebastiani, “SentiWordnet: A publicly available lexical resource for opinion mining,” Proceedings of LREC, 2006, pp. 417-422.
- [7] E. Cambria and A. Hussain, Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, 2012.
- [8] E. Cambria, D. Olsher, and K. Kwok, “Sentic activation: A two-level affective common sense reasoning framework,” Proceedings of AAAI, 2012, pp. 186-192.
- [9] K. Scherer, “What are emotions? And how can they be measured?” Social Science Information, vol 44(4), pp. 693-727, 2005.
- [10] W. Gale, K. Church, and D. Yarowsky, “One Sense per Discourse,” Proceedings of the DARPA Workshop on Speech and Natural Language, 1992.
- [11] T. Pedersen, S. Patwardhan, and J. Michelizzi, “Wordnet::Similarity - Measuring the relatedness of concepts,” Proceedings of AAAI, 2004, pp. 1024-1025.
- [12] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, “Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space,” Advances in Neural Networks, 2011, pp. 601-610.
- [13] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers: Norwell, 1981.
- [14] S. Poria, A. Gelbukh, D. Das, and S. Bandyopadhyay, “Fuzzy clustering for semi-supervised learning—Case study: Construction of an emotion lexicon,” Proceedings of MICAI, 2012.