

# EmoSenticSpace: A Novel Framework for Affective Common-Sense Reasoning

Soujanya Poria<sup>1</sup>, Alexander Gelbukh<sup>2</sup>, Erik Cambria<sup>3</sup>, Amir Hussain<sup>4</sup>, Guang-Bin Huang<sup>1</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

<sup>2</sup> Center for Computing Research, National Polytechnic Institute, Mexico

<sup>3</sup> School of Computer Engineering, Nanyang Technological University, Singapore

<sup>4</sup> Department of Computing Science and Mathematics, University of Stirling, UK

**Abstract**—Emotions play a key role in natural language understanding and sensemaking. Pure machine learning usually fails to recognize and interpret emotions in text. The need for knowledge bases that give access to semantics and sentics (the conceptual and affective information) associated with natural language is growing exponentially in the context of big social data analysis. To this end, this paper proposes EmoSenticSpace, a new framework for affective common-sense reasoning that extends WordNet-Affect and SenticNet by providing both emotion labels and polarity scores for a large set of natural language concepts. The framework is built by means of fuzzy c-means clustering and support-vector-machine classification, and takes into account different similarity measures, such as point-wise mutual information and emotional affinity. EmoSenticSpace was tested on three emotion-related natural language processing tasks, namely sentiment analysis, emotion recognition, and personality detection. In all cases, the proposed framework outperforms the state of the art. In particular, the direct evaluation of EmoSenticSpace against the psychological features provided in the ISEAR dataset shows a 92.15% agreement.

**Keywords**—Sentic computing, opinion mining, sentiment analysis, emotion detection, personality detection, fuzzy clustering.

## I. INTRODUCTION

Opinion mining is a rapidly developing area of natural language processing research that has recently received increasing attention from both academia and industry. It helps companies know what customers feel about their products, helps political parties or governments understand what voters think about their actions and proposals; on the other hand, it can help customers or voters choose wisely and in an informed way, by knowing the opinions of their peers regarding a specific product or a political candidate. Therefore, opinion mining is of great importance for improving economy, democracy, and quality of life.

The basic “feeling” about something can be described on a scale of approval or disapproval, good or bad, positive or negative, termed *polarity*. The basic question of whether customers or voters are satisfied with a product, service or action can be answered by detecting the average polarity of what they express in blogs and reviews.

For such a purpose, many online resources have been developed in recent years. SenticNet, in particular, is a semantic resource for concept-level sentiment analysis built by means of dimensionality reduction (Cambria *et al.*, 2014). It provides polarity scores for 5,700 multi-word concepts, and in its most recent release the set is expanded to 30,000 concepts. The score ranges between  $-1$  (bad) to  $+1$  (good), with neutral scores being around zero; e.g., *aggravation*:  $-0.925$ , *accomplish goal*:  $+0.967$ , *December*:  $+0.111$ . For example, this dictionary allows us to detect whether a customer review is 0.567 positive or 0.876 negative.

However, more detailed information is often desirable (Plutchik, 2001). Do citizens disliking the governing party specifically feel anger, fear, disgust, or shame? Do customers like a product due to the joy a product gives them or because it surprised them, and if both, are they more surprised than joyful, or vice versa?

One of the main lexical resources employed to detect emotions in text is WordNet-Affect (WNA) (Strapparava and Valitutti, 2004). For a relatively small set of words, WNA indicates whether a given term is related to one of six basic emotions: anger, fear, disgust, sadness, surprise, or joy. However, it does not provide information on how strong the emotion is: e.g., *angered* and *infuriated* appear as synonymous in WNA, while they evidently express a different emotional intensity. As shown later in the paper, augmenting emotion labels with quantitative scores that go beyond the current yes/no labeling is very useful to accurately and quantitatively answer many opinion-related questions.

On the other hand, for each concept SenticNet gives a quantitative measure of some unspecified emotion or unspecified mixture of emotions. However, for better opinion mining and sentiment analysis, both a quantitative measure and a specific emotion label are needed for each concept.

This paper starts with a description of an empirical method for automatically merging WNA and SenticNet, thus creating a resource, termed EmoSenticNet<sup>1</sup> (ESN), with both qualitative emotion labels and quantitative polarity scores assigned to a large number of concepts (Poria *et al.*, 2013b). Then, we present EmoSenticSpace, obtained by blending ESN with an existing resource, ConceptNet. We show that this new resource outperforms state-of-the-art approaches on a number of important cognitive-related applications, specifically, sentiment analysis, emotion detection, and personality recognition.

The paper is organized as follows: Section II discusses related works; Section III presents a general overview of the proposed method. Sections IV to IX describe in detail the process of assigning emotion labels to SenticNet concepts, and Sections X and XI describe the EmoSenticSpace resource properly. Namely, Section IV describes the lexical resources used in assigning emotion labels to SenticNet concepts; Section V presents the features used for classification in this process; Section VI describes the fuzzy c-means clustering scheme and Section VII the final hard clustering; Section VIII gives details on data selection for building our specific resource; finally, Section IX presents some evaluation results. Next, Section X introduces EmoSenticSpace, and Section XI describes three applications of the developed resources. For convenience of the reader, in Section XII we summarize the various lexical resources and features mentioned in the paper and their relationships with each other; we encourage the reader to use this section as a cheat sheet for the names of lexical resources. Finally, Section XIII concludes the paper and discusses some directions for future work.

---

<sup>1</sup> <http://gelbukh.com/emosenticnet>

<sup>2</sup> <http://sentic.net>, visited on May 14, 2014.

## II. RELATED WORK

Our work lies in the intersection of two large and interrelated research fields: opinion mining and sentiment analysis.

### A. Opinion Mining

Opinion mining is a recent and rapidly growing field of interdisciplinary research. As the Web plays an ever-increasing role in people's social lives, it also starts to store more information concerning their opinions and sentiments. There are three main categories of methods that can be applied to effectively mine such opinions and detect their emotional content or polarity:

- keyword spotting: the text is classified into categories based on the presence of fairly unambiguous affect words (Elliott, 1992; Wiebe *et al.*, 2005);
- lexical affinity: arbitrary words are assigned a probabilistic affinity to a particular topic or emotion (Wilson *et al.*, 2005; Somasundaran *et al.*, 2008; Rao and Ravichandran, 2009; Stevenson *et al.* 2007; Bradley and Lang, 1999); and
- statistical methods: the valences of keywords, punctuation, and word co-occurrence frequencies automatically calculated on a large training corpora are used (Turney and Littman, 2003; Hu and Liu, 2004; Pang and Lee, 2005; Abbasi *et al.*, 2008; Velikovich *et al.*, 2010).

The above approaches mainly rely on parts of text in which opinions and sentiment are explicitly expressed, such as polarity terms, e.g., *good*, *bad*, *nice*, *nasty*, *excellent*, and *poor*, as well as affect words, such as *happy*, *sad*, *calm*, *angry*, *interested*, and *bored*.

### B. Sentic Computing and SenticNet

Currently available lexical resources for opinion polarity and affect recognition such as SentiWordNet (Esuli and Sebastiani, 2006) or WNA are known to be rather noisy and limited. These resources mainly provide opinion polarity and affective information 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 texts to express viewpoints and affect.

In contrast, some recent approaches deal with concepts for mining the opinions (Poria *et al.*, 2014). Sentic computing (Cambria and Hussain, 2012) is a multidisciplinary approach to concept-level sentiment analysis that combines computer-science and social-science approaches to better recognize, interpret, and process opinions and sentiments found in the Web. The core module of sentic computing is SenticNet, a lexical resource that lists several thousand common-sense knowledge concepts along with their polarity. For example, for evaluation of medical patient data, this lexicon based on concepts performed better than other available lexicons such as SentiWordNet or WNA (Cambria *et al.*, 2010b).

### C. *Affect and Emotions*

While emotions are not linguistic entities, the most convenient access to them is through language (Strapparava and Valitutti, 2004). Natural language texts not only contain informative contents, but also attitudinal personal information including emotions, opinions, and attitudes. Recently, research activities related to emotions expressed in natural language texts and other media are gaining ground under the umbrella of subjectivity analysis and affective computing (Das, 2011).

The majority of subjectivity analysis methods related to emotions are based on textual keyword spotting using specially tailored lexical resources. A number of techniques for developing dictionaries of sentiment-related words have been proposed (Pang *et al.*, 2002; Wiebe and Mihalcea, 2006). The Affective lexicon (Strapparava and Valitutti, 2004) is one of the most important resources for detecting emotions in text, despite its small size.

The aspects that govern the lexical level semantic orientation depend on natural language context (Pang *et al.*, 2002), language properties (Wiebe and Mihalcea, 2006), domain pragmatic knowledge (Aue and Gamon, 2005), time dimension (Read, 2005), colors and culture (Strapparava and Ozbal, 2010) and

probably many other aspects that still remain unrevealed. Combining all such regulating aspects of emotion orientation lies in comprehending the human psychology and thus is a multifaceted, interdisciplinary problem (Liu, 2010). Although a word may evoke different emotions in different contexts, an emotion lexicon is a useful component for any sophisticated emotion detection algorithm (Mohammad and Turney, 2010) and is one of the main resources to start from.

### III. OVERVIEW OF THE PROPOSED METHOD

First (Sections IV to IX) we present and evaluate an empirical method for automatically assigning emotion labels to each concept of SenticNet, in addition to the polarity score already present in SenticNet. Associating affective information with concepts improves opinion mining results and enables mining for more detailed affective information. Then (Sections X and XI) we present the EmoSenticSpace resource and show how our new resources, in combination with existing resources, offer superior accuracy on sentiment analysis, emotion detection, and personality detection tasks.

Specifically, for the first task—assigning emotion labels to concepts of SenticNet, we extended emotion labels from the seed concepts (for which associated labels were known from the WNA), to all concepts in SenticNet. Thus, our task was to classify the concepts present in SenticNet into six categories given in WNA: anger, fear, disgust, sadness, surprise, and joy (Poria *et al.*, 2012).

For this, we used a supervised learning approach. As a source of features, we used various lexical resources, most importantly the International Survey of Emotion Antecedents and Reactions (ISEAR) dataset (Scherer, 2005), as well as WordNet. We constructed two kinds of feature: features directly assigned to a single concept and similarity measures between concepts. Each similarity measure was converted into a set of features: for a given concept, we considered its similarity to every other concept as its independent features (Poria *et al.*, 2013a).

A concept, e.g., *succumb*, can trigger a mix of multiple emotions, such as shame, fear, and sadness. This suggests that the problem of concept emotion identification is intrinsically a multi-label classification

problem. We accounted for this intuitively by using fuzzy c-means clustering (Bezdek, 1981) to assign each concept a vector of its membership values in each of the six categories.

However, our goal was single-label classification. This was partly justified by the simplicity of use of the created resource. More importantly, the obtained multi-label classification was still not reliable enough to be used in applications. Thus, we did not directly use the obtained membership values. Instead, we used them to improve the accuracy of the conventional, hard classification problem.

A trivial approach would be to choose the label with the maximum membership function as the final label for a concept: *succumb* invokes more shame than sadness or fear. However, since the obtained membership functions were not very accurate, such a choice would not be reliable when a concept was associated with more than one category in comparable degree. Instead, we restricted the choice to the best and the second-best prediction of the fuzzy clustering (we also tried  $K$  best options, and  $K = 2$  gave the best results).

To disambiguate between these options, we used the well-known support vector machine (SVM) classifier (Joachims, 1998). It operated on the same set of features as the fuzzy clustering algorithm, plus the obtained fuzzy membership values as additional features. These additional features supposedly made the choice easy when the difference between the best and the second-best membership values was big enough, and when it was not, the algorithm resorted to other features to make the choice.

In addition, we improved the objective function of fuzzy c-means clustering by incorporating additional functions such as point-wise mutual information (PMI) and the emotional affinity between two concepts.

#### A. *The Algorithm for Assigning Emotion Labels to SenticNet Concepts*

In summary, our algorithm for assigning emotion labels to SenticNet concepts can be outlined as below:

1. Select a repertoire of concepts (as described in Section IV).
2. Compute features for each concept (Section V).
3. Use these features to cluster the concepts into six fuzzy clusters (Sections VI.A, B).

4. Associate these clusters with specific emotion labels (Section VI.C).
5. For each concept, use the fuzzy clustering results to:
  - a. restrict the confusion set to top two labels (Section VII.A);
  - b. extend the feature vector by the six membership values.
6. Use a hard classifier (SVM) on these extended feature vectors to disambiguate between these two labels (Section VII.B).

#### B. Role of Fuzzy Clustering in Assigning Emotion Labels

For each concept, fuzzy clustering provides a vector of the membership values of a concept, i.e., affinity of a concept to each of the six emotion clusters. This is similar to the Hourglass model introduced by Cambria and Hussain (2012).

While the fuzzy clustering algorithm produces fuzzy clusters containing the concepts, our purpose was to identify the definite emotion class labels for each concept. The concepts that belong to more than one cluster to a significant degree, require an effective separation algorithm in order to be classified accurately. We employed the SVM-based classifier to identify the final class for each concept.

Yu *et al.* (2003), Awad *et al.* (2004), and Boley and Cao (2004), amongst others, have shown that clustering techniques can help to decrease the complexity of SVM training. However, these techniques consume significant computational resources to build the hierarchical structure. Cervantes *et al.* (2006) introduced SVM classification based on fuzzy clustering. In this paper we follow a similar approach for emotion classification. As outlined in Section III.A, the fuzzy clustering helps the final classification task in two ways:

First, it reduces the confusion set for the SVM-based classification from 6 to 2 labels associated with the highest membership values, given that we identify the emotion labels of the corresponding clusters. Thus we reduce the task to binary classification, at which SVM is particularly effective.



Second, we employ the vector of the fuzzy membership values of a particular concept as an additional feature for SVM. For example, if the fuzzy membership vector for a concept is 0.45 for anger, 0.34 for sadness, 0.03 for surprise, 0.05 for joy, 0.01 for disgust, and 0.127 for fear, then these six numeric values are used as independent features in the feature vector for this concept, along with all other features.

Specific class names were obtained by employing the majority voting method described in Section VI.C. As demonstrated in Section IX.A, the reduction of the confusion set and the additional fuzzy vector features increased the classification accuracy.

#### IV. LEXICAL RESOURCES USED

In this section, we describe the lexical resources used to build the set of concepts, construct the features of the concepts along with the similarity measures, and to evaluate the obtained resource.

##### A. *SenticNet*

As the target lexicon and the source of polarity information for our polarity-based concept similarity measure, we used SenticNet,<sup>2</sup> a freely available knowledge base that assigns polarity values to words or multi-word concepts.

Specifically, we employed the beta version of SenticNet 3.0.<sup>3</sup> It contains 13,741 concepts,<sup>4</sup> of which 7,626 are multi-word expressions, e.g., *prevent pregnancy*, *high pay job*, *feel happy*. Of the concepts in SenticNet, 6,452 are found in WordNet 3.0 and 7,289 are not. Of the latter, most are multi-word concepts such as *access internet* or *make mistake*, except for 82 single-word concepts, such as *against* or *telemarketer*.

---

<sup>2</sup> <http://sentic.net>, visited on May 14, 2014.

<sup>3</sup> <http://sentic.net/senticnet-3.0.zip>, downloaded on May 14, 2014.

<sup>4</sup> SenticNet 3.0 is currently under development; it will contain 30,000 concepts. Applying our method to this new version will automatically result in a resource of the corresponding size.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://sentic.net/api/en/concept/celebrate_special_occasion">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net/api">celebrate special occasion</text>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_holiday"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_occasion"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_birthday"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/celebrate_wed"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/express_appreciation"/>
    <pleasantness xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.93</pleasantness>
    <attention xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.724</attention>
    <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</sensitivity>
    <aptitude xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</aptitude>
    <polarity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.551</polarity>
  </rdf:Description>
</rdf:RDF>

```

Figure 1. A sample of SenticNet

The resource is distributed in RDF XML format (Figure 1) and it is also accessible through an API.<sup>7</sup> The first 20 SenticNet concepts in alphabetical order along with their corresponding polarities are shown in Table I.

TABLE I.  
A sample of SenticNet data

<i>a lot</i>	+0.258	<i>Abhorrent</i>	-0.443
<i>a lot sex</i>	+0.858	<i>able read</i>	+0.865
<i>a little</i>	+0.032	<i>able run</i>	+0.775
<i>abandon</i>	-0.566	<i>able use</i>	+0.856
<i>Abase</i>	-0.153	<i>abominably</i>	-0.443
<i>Abash</i>	-0.174	<i>Abominate</i>	-0.391
<i>Abashed</i>	-0.174	<i>abomination</i>	-0.391
<i>abashment</i>	-0.186	<i>Abortion</i>	-0.27
<i>Abhor</i>	-0.391	<i>Abroad</i>	+0.255
<i>abhorrence</i>	-0.391	<i>Absolute</i>	+0.277

The distribution of the polarity values in the dictionary is given in Figure 2. In this figure, each bar corresponds to the number of concepts in SenticNet 3.0 that have the given digit after the dot: for example, +0.0 corresponds to the interval from +0.000 to +0.099.

<sup>7</sup> <http://sentic.net/api>

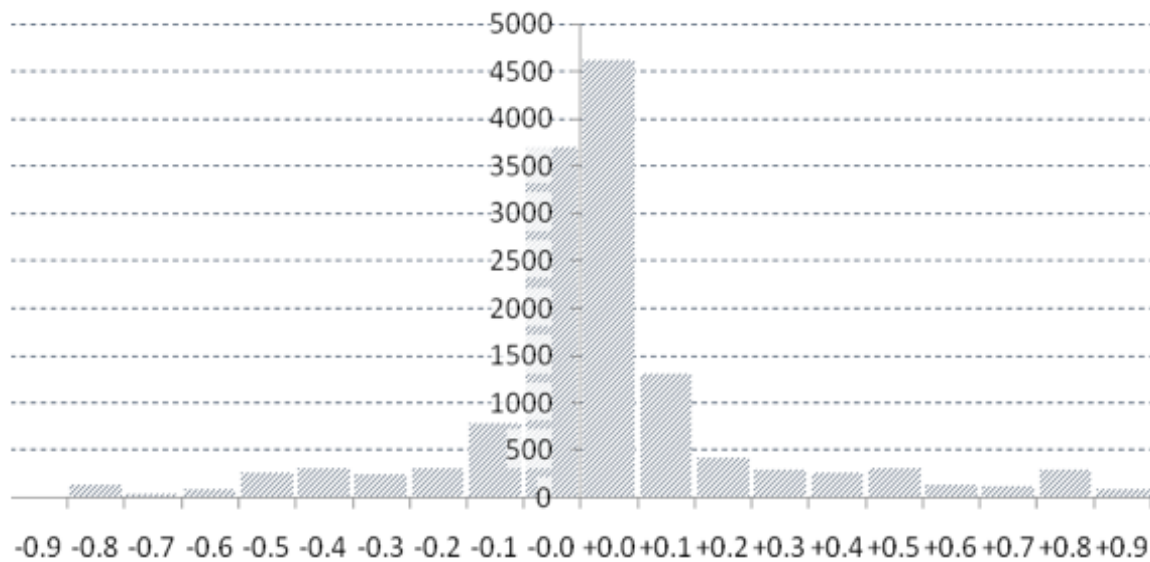


Figure 2. Distribution of polarity values in SenticNet

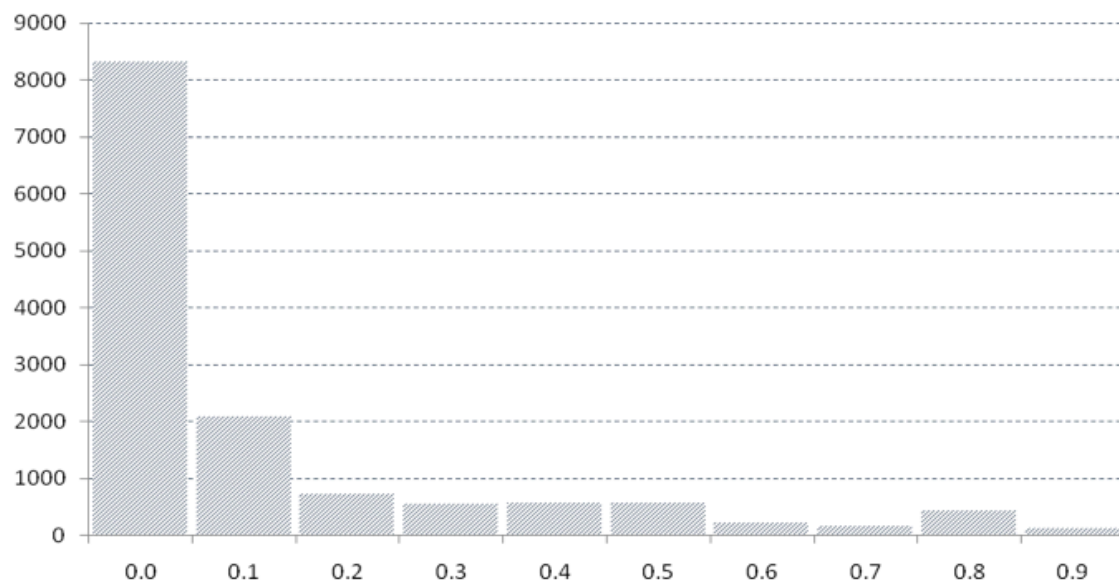


Figure 3. Distribution of polarity intensity in SenticNet

The distribution of the polarity intensity irrespective of its sign (positive or negative) is given in Figure 3 and Figure 4. Again, each bar corresponds to the number of concepts in SenticNet that have the given digit after the dot in their intensity: for example, 0.0 corresponds to the interval from 0.000 to 0.099.

As it can be observed from Figure 3, the majority of concepts have low intensity, and for intensities lower than 0.5, the lower the intensity the greater number of concepts possessing it. This gives the idea

that concepts with very low polarity intensity were not included in the dataset, and that concepts not listed in the dictionary can be assumed to have almost null polarity.

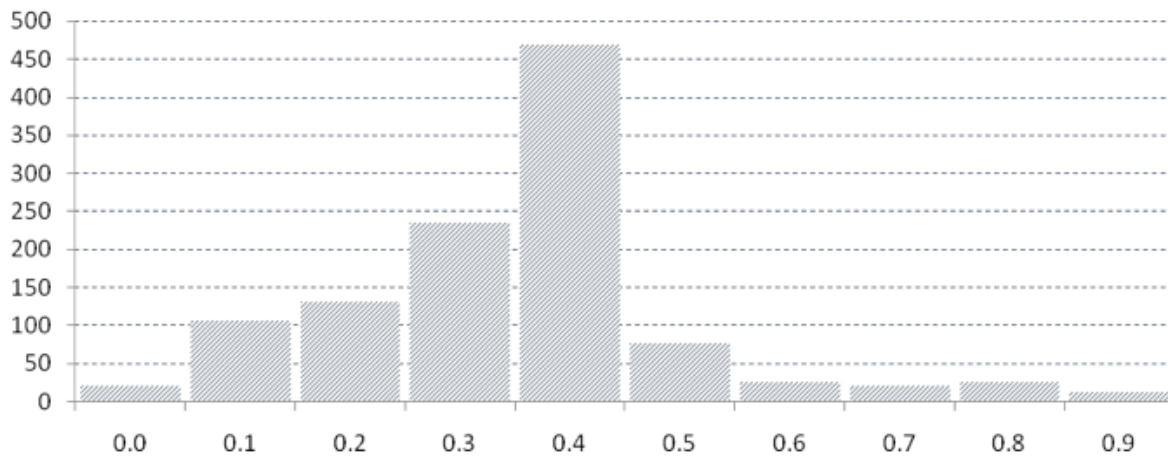


Figure 4. Distribution of polarity intensity of the words from WNA lists in SenticNet

It can also be observed that the words with a polarity intensity of about 0.5 form the smallest group. This might be an artifact of the labeling procedure or indeed a real linguistic phenomenon. Either way it does suggest that this is a natural division point between high and low polarity: the concepts with high polarity are those that have significant emotional charge.

#### B. WNA Emotion Lists

As an inventory of target labels and a source of training examples for the supervised classification, we used the emotion lists<sup>8</sup> provided for the SemEval 2007 task 14: Affective text.<sup>9</sup> According to the organizers of this task, the lists were extracted from WNA (Strapparava and Valitutti, 2004). There are six lists corresponding to the six basic emotions: anger, fear, disgust, sadness, surprise, and joy. 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. However, we ignored the synonymy information contained in the data and used the labels for individual words or

<sup>8</sup> <http://www.cse.unt.edu/~rada/affectivetext/data/WordNetAffectEmotionLists.tar.gz>, downloaded on July 12, 2012

<sup>9</sup> <http://www.cse.unt.edu/~rada/affectivetext>, visited on July 13, 2012

concepts, i.e., *puppy love*→JOY, *calf love*→JOY, *crush*→JOY, *infatuation*→JOY. Statistics of the synsets and concepts by label is given in Table II.

The dataset contains a total of 532 different synsets, of which 2 are assigned two distinct labels each:  $\{cliff-hanging, suspenseful, suspensive\} \rightarrow \{FEAR, JOY\}$  and  $\{persecute, oppress, harass\} \rightarrow \{ANGER, SADNESS\}$ . Thus, the numbers of synsets in Table II sum up to 534.

TABLE II.  
The number of synsets and concepts by emotion label in WNA lists

	Synsets	Concepts
JOY	204	400
ANGER	115	255
SADNESS	95	202
FEAR	76	147
SURPRISE	27	71
DISGUST	17	53
total:	534	1128

If 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 1128 in Table II). Apart from the words from the two ambiguous synsets, these are words that belong to different unambiguous synsets (in different senses), e.g., *awful*→ {FEAR, SURPRISE}: when breaking synsets into individual concepts, we lose the sense disambiguation information.

Of the obtained 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 11 (99.02%) are contained in SenticNet.

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

where we count a concept as ambiguous if it has two labels with contradicting polarity, e.g., *suspensive* → {JOY, FEAR}.

Of the 1,121 concepts present both in SenticNet and in WNA lists, 5 were found to have ambiguous WNA polarity, and the rest distributed according to the following confusion matrix:

In WNA	In SenticNet	
	positive	negative
positive	467	3
negative	5	646

This shows very good (99.28%) agreement between WNA lists and SenticNet as to the sign of concept polarity.

In some cases of disagreement, the polarity assigned by 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’).

### C. The ISEAR Dataset

As a source of various features and similarity measures between concepts, we employed the International Survey of Emotion Antecedents and Reactions (ISEAR)<sup>10</sup> dataset (Scherer, 2005). The survey was conducted in the 1990s across 37 countries involving 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:

---

<sup>10</sup> <http://www.affective-sciences.org/system/files/page/2636/ISEAR.zip>, downloaded on July 14, 2012.

*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 full dataset contains 7,666 such statements, which include 18,146 sentences and 449,060 running words. Of the 13,741 concepts contained in SenticNet, 3,312 were found in ISEAR. For these concepts important features were extracted from ISEAR and emotion labels were assigned.

Each statement in the ISEAR dataset is supplied with 40 numeric or categorical values, which give various kinds of information on the given situation and the respondent. Thus, the dataset is arranged in a table, with a statement column and 40 data columns. Some of these columns are not informative for our goals, such as the statement ID, the respondent ID, etc. We used only 16 data columns; they are presented in five groups in Table III.

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

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

The majority of 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; and 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. Seven values are used: ANGER, DISGUST, FEAR, GUILT, JOY, SADNESS, and SHAME; in the example above the label was FEAR. This set of seven emotions is slightly different from our target set of six basic emotions used in WNA lists: ANGER, FEAR, DISGUST, SADNESS, SURPRISE, and JOY. A more comprehensive overview of different sets of basic emotions can be found in (Ortony and Turner, 1990). We do not directly use the ISEAR emotion labels to assign our target emotion label, but we do use them as one of the feature types for our classification, alongside many other features.

The choice of ISEAR as the source of corpus-based information is motivated by the following considerations:

- This corpus is particularly rich in emotion-related words, as compared with more standard corpora used in natural language processing;
- Its statements are consistent with the emotion content. Thus, the “one emotion per discourse” principle, (Gale *et al.*, 1992), can be applied: if two expressions co-occur in the same statement, then they are related to 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 USED FOR CLASSIFICATION

The data units that we classified, and to which we assigned the emotion labels (classes) as our final result were concepts: single- or multi-word expressions present in SenticNet. In type-token terms, when gathering information from corpora such as ISEAR, we aggregated the information on multiple occurrences of the concept as a token in the text into one feature vector for the given concept as a type.



For our classification we used two kinds of features for the concepts: (1) those based on the data columns directly provided in the ISEAR dataset, and (2) those based on various similarity measures between concepts.

#### A. Features Based on ISEAR Data

We used the 16 ISEAR data columns listed in Table III as independent features. We treated all the features used, with the exception described below, as categorical features. For example, the *country* column has 16 different numerical codes, so we used 16 different features corresponding to each specific country. As 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.

The only exception was the *age* column. We tested two different approaches: to treat all values (integer number of years) as different categorical values, or to group them in ranges—the latter was found to give better results. According to (Peersman *et al.*, 2011), a psychologically motivated grouping of ages is: 18–23 years (all respondents of the ISEAR datasets were at least 18 years old), 23–28 years, and older than 28 years. So we used these three intervals as categorical features. This gave us about 100 categorical features, which were used as different dimensions of the feature space.

#### B. Features Based on Similarity Measures

Another kind of feature was given by similarity measures between concepts. Given  $N$  concepts ( $N = 3,312$  in our case), measuring similarity between pairs of concepts provides  $N$  distinct dimensions of the feature vector: the similarity between the given concept and each other concept in the vocabulary. This

corresponds to the intuition that if, for two data points, their Euclidian distances to a number of other points are similar, then these two points are probably close to one another.

To define these similarity-based features for classification, we used the following 13 similarity measures: 10 lexical resources-based measures (one SenticNet score-based similarity and nine WordNet-based similarity measures) and three co-occurrence-based measures (text distance-based similarity, point-wise mutual information, and emotional affinity) as described below. This contributed  $13N$  dimensions to the feature space.

The three 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. Therefore, we included all of them in the feature vectors as independent features.

#### 1) *SenticNet score-based similarity*

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

Since all concepts that we considered were present in SenticNet, they had valid SenticNet scores.

#### 2) *WordNet distance-based similarity*

The open-source package WordNet::Similarity<sup>11</sup> (Pedersen *et al.*, 2004) based on English WordNet 3.0 was used to measure the WordNet-based distance between concepts. This package provides nine similarity measures based on the analysis of the WordNet hierarchy, glosses, and other data present in WordNet. Namely, we used these following measures:

---

<sup>11</sup> <http://www.d.umn.edu/tpederse/similarity.html>

- A simple edge counting approach
- Hirst and St-Onge (1998)
- Leacock and Chodorow (1998)
- Extended Gloss Overlaps (Banerjee and Pedersen, 2003)
- Lin (1998)
- Jiang and Conrath (1997)
- Resnik (1995)
- Gloss Vector (Patwardhan *et al.*, 2003)
- Wu and Palmer (1994)

In this work, we used all the above nine measures as independent sources of information, each one corresponding to its own  $N$  dimensions in the feature vectors.

As mentioned in Section IV, not all concepts from SenticNet were present in WordNet 3.0. Of the 3,312 SenticNet concepts that were found in the ISEAR database, 1,436 were directly found in WordNet. Those concepts not found in WordNet, were examined manually and rephrased. For example, if a multi-word concept not present in WordNet, such as *make mistake*, contained a word that carried its main semantics and was not a stop word, the concept was manually reduced to this main word, *mistake* in our example. After this, 169 more concepts proved to map to WordNet, giving in total 1,605 concepts with meaningful WordNet pairwise similarity scores.

For the remaining 1,124 concepts not mapped to WordNet, their  $N$  similarity values to all other concepts were set to random values in the interval  $[0, 1]$ . We did not set those values to 0 or some other fixed value because this made all concepts not found in WordNet, very far from other concepts and very similar to each other: indeed, 70% of their features ( $9N$ ) would be identical. This made them form one large cluster, which deteriorated the final results. In contrast, using random values better expressed the idea of unknown similarity, placing all such concepts in the feature space far from all others and also from each other.

All nine similarity scores are defined for specific senses and not for just character strings. For ambiguous concepts, we defined similarity as the maximum similarity overall senses of the first and the second concept.

### 3) ISEAR text distance-based similarity

The positional information of concept tokens in the ISEAR statements was used to measure the ISEAR-based similarity between them. For this, we calculated the average distance between the concepts in the ISEAR dataset statements. Namely, if the tokens  $a$  and  $b$  occur in a statement  $S$  at the positions (in words)  $a_1, \dots, a_n$  and  $b_1, \dots, b_n$ , correspondingly, then the distance between  $a$  and  $b$  in this statement was defined as  $D_{ISEAR}(a, b, S) = \min(|a_i - b_j|)$ , and the distance over the entire ISEAR dataset was defined by averaging over individual statements where both concepts co-occur  $S_k$ :  $D_{ISEAR}(a, b) = \text{avg} D_{ISEAR}(a, b, S_k)$ .

The similarity was defined as the inverse of the distance:  $\text{Sim}_{ISEAR}(a, b) = 1 / D_{ISEAR}(a, b)$ . Note that if the two tokens appear next to each other (as a bigram), then  $D_{ISEAR}(a, b, S) = 1$ ; in particular,  $D_{ISEAR}(a, b, S) \geq 1$ . If the concepts did not co-occur in any statement, then we considered  $\text{Sim}_{ISEAR}(a, b)$  to be a random number between 0 and 1.

### 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

$$\text{Sim}_{PMI} = \log \frac{P(a, b)}{P(a)P(b)}, \quad (1)$$

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  $b$ .

Ten concept pairs with the greatest  $\text{Sim}_{PMI}$  are given in Table IV.

TABLE IV.  
Top 10  $Sim_{PMI}$  pairs

Concept pair		$Sim_{PMI}$
<i>Weekend</i>	<i>December</i>	3.864
<i>Birthday</i>	<i>Celebration</i>	3.839
<i>Happy</i>	<i>December</i>	3.821
<i>Angry</i>	<i>Friend</i>	3.813
<i>Learn</i>	<i>school student</i>	3.813
<i>Party</i>	<i>Friend</i>	3.813
<i>graduate</i>	<i>school examination</i>	3.811
<i>Disgust</i>	<i>Behavior</i>	3.807
<i>Money</i>	<i>Important</i>	3.807
<i>Mistake</i>	<i>Realize</i>	3.806

##### 5) Emotional affinity

We defined the emotional affinity between two concepts  $a$  and  $b$  in the same way as  $Sim_{PMI}$  but at the level of entire statements and not sentences, i.e.,  $P(\cdot)$  in (1) was defined as the corresponding number of *statements* instead of sentences, normalized by the total number of *statements*. Similar to Table IV, the top ten emotional affinity pairs are given in Table V.

TABLE V.  
Top 10 emotional affinity pairs

Concept pair		Affinity
<i>Weekend</i>	<i>December</i>	3.827
<i>unemployed</i>	<i>Sad</i>	3.800
<i>tired of</i>	<i>Headache</i>	3.768
<i>Happy</i>	<i>December</i>	3.755
<i>Worry</i>	<i>computer virus</i>	3.721
<i>need relax</i>	<i>Annoyed</i>	3.707
<i>Mistake</i>	<i>Realize</i>	3.691

Concept pair		Affinity
<i>Disgusting</i>	<i>to person</i>	3.604
<i>Serious</i>	<i>feel guilty</i>	3.518
<i>Birthday</i>	<i>Thought</i>	3.499

While PMI often reflects syntactic association between the concepts—for example, it is high for a verb and its typical object, or for components 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. 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.

## VI. FUZZY CLUSTERING

The first step in our process was unsupervised: we did not use the emotion labels known from WNA list. At this step, we grouped together similar concepts. Later on, these clusters were identified with specific emotion labels to construct the six clusters: one for each of the six WNA basic emotions.

We considered *fuzzy* clusters so that a concept could belong to multiple clusters with different degrees of membership. Fuzzy clustering was applied to determine the preliminary estimation for a concept to be related with each specific group; later in the process only one cluster was finally selected for each concept; see Section VII.

The features used for the fuzzy clustering were described in Section V, and the result of the clustering, for each concept as well as for each of the six groups, was the membership function of a concept in a class defined by the given emotion label. A number ranging between 0 and 1 represented the estimation of the association between the token and a particular emotion.

### A. Fuzzy C-means Clustering Algorithm

For fuzzy clustering, we used the fuzzy c-means clustering algorithm (Bezdek, 1981) with a modified objective function as described in Section VI.B below.

The well-known fuzzy c-means clustering algorithm takes as input, a set of  $N$  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 found (where  $c$  is a given parameter), and a set of  $c * N$  *membership values*  $\mu_{ik}$ ,  $i = 1, \dots, c$ ;  $k = 1, \dots, N$ , 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, so it is assumed that  $0 \leq \mu_{ik} \leq 1$  and such shares sum up to unity for each point, that is:

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

To find the optimal distribution of points by clusters and optimal placement of the centroids, the algorithm uses an *objective function*  $J$ , which achieves its minimum 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  $\mu_0, v_0$  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 parameter that controls the degree of fuzziness of the obtained clusters (for our case,  $p = 2$ ), and

$$\|x - y\| = \sum_{l=1}^P (x_l - y_l)^2 \quad (4)$$

is the Euclidean distance in the feature space.

The optimal solution of a constraint optimization problem defined by equations (2) and (3) is given by the following (see Section VI.B below for derivation):

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^p x_k}{\sum_{k=1}^N \mu_{ik}^p} \quad (5)$$

and

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{p-1}}}; \quad (6)$$

Now, a stationary point  $(\mu_0, v_0)$  of the system (5), (6) is found iteratively using the steps below:

- (i) Assign random values to all  $\mu_{ik}$ ; the values are normalized to satisfy the constraints (2).
- (ii) Iteratively re-calculate the values for all  $v_i$  and then all  $\mu_{ik}$  according to (5) and (6), respectively;
- (iii) Stop when the objective function  $J$  changes from the previous iteration less than by a small number  $\varepsilon$  (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 into the original objective function (3):

$$J_p(\mu, v) = \sum_{i=1}^c \sum_{k=1}^N \mu_{ik}^p \left( \|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right), \quad (7)$$

where the parameter  $\rho$  is intended to control the effect of the new term (for our case,  $\rho = 1$ ) and  $N_k$  is the set constructed in the following way:

- For each data point  $x$ , we identified the nearest centroid

$$v(x) = \arg \min_i \|v_i - x\| \quad (8)$$

(in case of a tie, an arbitrary centroid was chosen);

- $N_k = \{x \mid v(x) = v(x_k)\}$  was thus set of all data points with the same nearest centroid as  $x_k$ .



This additional term provides a ‘hint’ to the algorithm to increase the membership function of a data point in the cluster with the nearest centroid, thus better grouping the similar points together. Our experiments (see Table VII) show that this modification of the objective function improved the accuracy of the results.

In our implementation, we constructed the sets  $N_k$  on the fly while re-calculating the positions of the centroids according to (5) (rather its modified version, (10)), i.e., when re-calculating  $v_2$ , we considered in (8) the already re-calculated value for  $v_1$ .

The change of the expression for the objective function required modification of (5) and (6). For the sake of completeness, a derivation of the modified equations is provided. A note on the discontinuity of  $J_p(\mu, v)$  is also given below.

A necessary condition for the optimum of function (7), subject to constraint (2), to be achieved at a point  $(\mu_0, v_0)$  is the vanishing of all partial derivatives  $\frac{\partial L}{\partial \mu_{ik}}, \frac{\partial L}{\partial v_{im}}$  ( $v_{im}$  are coordinates of the centroid  $v_i$  in the  $P$ -dimensional feature space), and  $\frac{\partial L}{\partial \lambda_k}$  of the Lagrangian of the system:

$$L = \sum_{k=1}^c \sum_{i=1}^N \mu_{ik}^p \left( \|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right) + \sum_{k=1}^N \lambda_k \left( 1 - \sum_{i=1}^c \mu_{ik} \right). \quad (9)$$

The first term of  $L$  is  $J_p(\mu, v)$  from (7), and the second term, according to the method of Lagrange multipliers, is derived from the constraint (2), so that the conditions  $\frac{\partial L}{\partial \lambda_k} = 0$  reduce to equation (2).

Given (7), we have the following:

$$\begin{aligned} \frac{\partial L}{\partial v_{im}} &= \frac{\partial}{\partial v_{im}} \sum_{k=1}^N \mu_{ik}^p \left( \sum_{l=1}^P (x_{kl} - v_{il})^2 + \rho \sum_{x \in N_k} \sum_{l=1}^P (x_l - v_{il})^2 \right) \\ &= \sum_{k=1}^N \mu_{ik}^p \left( -2(x_{km} - v_{im}) + \rho_k \sum_{x \in N_k} -2(x_m - v_{im}) \right) = 0 \end{aligned}$$

and, taking into account that  $\sum_{x \in N_k} v_{im} = v_{im} \sum_{x \in N_k} 1 = v_{im} |N_k|$ , we have:

$$v_{im} = \frac{\sum_{k=1}^N \mu_{ik}^p \left( x_{km} + \rho \sum_{x \in N_k} x_m \right)}{\sum_{k=1}^N \mu_{ik}^p (1 + \rho |N_k|)}$$

or, in a vector form,

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^p \left( x_k + \rho \sum_{x \in N_k} x \right)}{\sum_{k=1}^N \mu_{ik}^p (1 + \rho |N_k|)}, \quad (10)$$

which substitutes (5) for our modified objective function. Similarly,

$$\frac{\partial L}{\partial \mu_{ik}} = p \mu_{ik}^{p-1} \left( \|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right) - \lambda_k = 0$$

gives

$$\mu_{ik} = \frac{\lambda_k^{\frac{1}{p-1}}}{p^{\frac{1}{p-1}} \left( \|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right)^{\frac{1}{p-1}}},$$

which together with (2) gives

$$\lambda_k^{\frac{1}{p-1}} = \frac{1}{\sum_{i=1}^c \frac{1}{p^{\frac{1}{p-1}} \left( \|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2 \right)^{\frac{1}{p-1}}}}$$

and thus:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_i\|^2 + \rho \sum_{x \in N_k} \|x - v_i\|^2}{\|x_k - v_j\|^2 + \rho \sum_{x \in N_k} \|x - v_j\|^2} \right)^{\frac{1}{p-1}}}, \quad (11)$$

which substitutes (6) for our modified objective function. If  $\rho = 0$ , then (10) and (11) reduce to (5) and (6), respectively.

This derivation only applies to the regions where  $J_p(\mu, v)$  is differentiable. In fact it is not everywhere continuous, since the sets  $N_k$  change sharply when the data points jump from one  $N_k$  to another as  $v_i$  changes. Therefore, our analytic solution will only find a local optimum within a region of continuity of  $J_p(\mu, v)$ .

This is, however, common practice, as the original solution (5), (6) does not guarantee a global optimum, either. In such cases, estimation of the performance of the resulting iterative algorithm is left to empirical evaluation, which is given in Section IX.

### C. Mapping Fuzzy Classes to Emotion Labels

Following the completion of the unsupervised fuzzy clustering process, we identified which of the  $c = 6$  classes appropriately corresponded to one of the six emotion labels. For this, we employed a simple supervised approach.

First, we converted the fuzzy clustering into hard clustering; in our implementation we chose for each data point  $x_k$  a cluster  $C(x_k) = \arg\max_i \mu_{ik}$  to which it was most strongly associated (in case of a tie, an arbitrary class was chosen).

Then the emotion label for each hard cluster was chosen by majority voting. Namely, considering all concepts in the cluster that happened to be in WNA emotion lists and thus had known emotion labels, we 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 then selected a label that occurred the highest number of times.

This procedure did not guarantee that there be six hard clusters, or that the majority voting would not result in a tie, nor that two clusters would share the same emotion label. In the latter case, some labels would not be assigned to any cluster. However, such low probable events did not occur 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 in our final results, which would not be possible with incorrectly mapped labels.

## VII. HARD CLUSTERING

While several concepts appear in more than one WNA list (for example, *harass* is listed under SADNESS and ANGER, *suspensive* under JOY and FEAR), most of them have only one emotion label. Predicting whether a word is ambiguous, is out of the scope of the present paper, and it cannot be said with certainty if such ambiguity might be the result of errors in WNA lists. Therefore, to simplify things, we chose to assign only one emotion label to each concept.

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

A two-step process was used for choosing a single class for each token classification.

### A. Reduction of the Confusion Set

Given a parameter  $K$  discussed below, for each concept, we chose top  $K$  classes, that is,  $K$  classes for which the fuzzy clustering assigned this concept the highest value of the membership function. For example, if  $K = 2$  and the six membership functions for a given concept were  $\{0.1, 0.8, 0.3, 0.4, 0.6, 0.2\}$ , then classes 2 and 4 were selected and the classes 1, 3, 5, and 6 discarded as having lower membership functions for this concept than the  $K = 2$  selected classes. The hard clustering technique used afterwards was only allowed to choose between those top  $K$  labels pre-selected for a given concept.

In case of  $K = 1$ , only one class per concept is chosen, and thus no further processing is needed: in this case, the final result is given by choosing the greatest membership function obtained from the fuzzy clustering. In case of  $K = 7$ , there is no reduction of the confusion set, i.e., the reduction step is in effect omitted. In case of  $K = 2$  or  $K = 3$ —the values we experimented with—the confusion set is reduced to 2 or

3 options, correspondingly. We show in Section IX.B that reducing the confusion set to only two candidates increases the accuracy.

#### B. Final hard classification

Given the  $K$  options left after reducing the confusion set, we trained separate 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., which gives  $6 \times 5 / 2 = 15$  different classifiers. To assign a unique label to a concept, the  $K$  emotion labels for it were selected as explained above, e.g., FEAR and SADNESS, and then the corresponding classifier was used.

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

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

As a hard clustering 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.

## VIII. IMPLEMENTATION

While in Sections III to VII we have described an abstract method applicable to any datasets, in this section we give the details of the specific dataset to which we applied the method described above and present examples of the resulting labels.

### A. Data Preparation

We used a number of standard pre-processing techniques, such as tokenizing and lemmatizing, as described below. For this, we used the tools provided by Rapidminer's text plug-in, except for lemmatizing (a lemmatizer differs from a stemmer in that it provides a complete form: for example, for *feet*, it provides *foot*), for which we used the WordNet lemmatizer (Miller, 1995).

For each SenticNet concept, we identified all its occurrences in the text of the ISEAR statements. All words in the statements were lemmatized before matching, because the concepts could appear in the text in a different form, e.g., *made mistake* in the text vs. *make mistake* in the vocabulary (this also generated some small number of false matches). For multi-word SenticNet concepts, such as *after summer*, *to person*, etc., we allowed any number of stop-words to appear in the position of the space, so that, in SenticNet, they were readily matched, e.g., *to a person* or *to the person* in the text.

Total of 3,312 SenticNet concepts appear at least once in the ISEAR dataset. Only these concepts participated in further processing and were finally assigned the emotion labels. For each occurrence, we extracted the corresponding data fields from the ISEAR dataset, and the data for multiple occurrences of the same concept in the corpus (as a token) were aggregated in a feature vector for that concept (as a type). Thus, this gave us a dataset with a total of 3,312 feature vectors.

All 3,312 concepts participated in the unsupervised fuzzy clustering phase, though not all of them participated in the supervised final hard clustering; see Section VII.

## B. Examples

Consider the following example: *When I make mistake that I has accused someone else of, this is obvious to the person.* The concepts found in this example and their corresponding emotion labels assigned by our classifier are *make mistake* (sadness), *obvious* (joy), *to person* (surprise).

For the example *I am sad when friends try to put me down or hurt me*, the corresponding concepts and emotion labels are *sad* (sadness), *friend* (joy), *put down* (sadness), *hurt* (fear).

## IX. DIRECT EVALUATION OF THE ASSIGNED EMOTION LABELS

For evaluation, we used a standard tenfold cross-validation procedure. Namely, we excluded from training a tenth part of WNA emotion labels. Then we normally constructed the resource using this reduced WNA set. In the constructed resource, we compared the labels assigned to the concepts that were temporarily excluded from WNA (but left in SenticNet, so that they were automatically assigned some labels) with the true labels present for these concepts (but unavailable during training); percent of coincidence gave accuracy. This accuracy was averaged by ten experiments where one of ten different tenth parts of WNA was excluded from training and used for evaluation.

We tested several classifiers for the final classification task. Of them, SVM produced the best result, with 92.15% accuracy; see Table VI.

Apart from different classifiers, we compared the effect of selection of different parameters and different subsets of features, in order to study which features are most important for assigning the emotion labels. Below we first describe these experiments in detail.

### A. Impact of the Fuzzy Clustering and Hard Classification

We compared the results of the pipeline of fuzzy clustering and hard classification as described above versus directly applying hard classification without the fuzzy clustering phase; in the latter case, the hard classifier did not have access to the extra features obtained by the fuzzy classifier. The results are presented in Sections A and B below.

We also experimented with different values of  $K$  introduced in Section VII: the size of the confusion set after reduction based on the result of fuzzy clustering. Namely:

- $K = 1$  means that the final classification directly results from the fuzzy clustering and no further hard clustering is necessary;
- $K = 2$  means that the final hard classification is based on binary choices between the two top labels for each concept;
- $K = 3$  means that reduction of the confusion set for the hard classification to a choice of three options;
- $K = 6$  means no reduction of the confusion set. However, it is not the same as not using the fuzzy clustering phase at all as described above, because the fuzzy clustering results are still used as additional features for the hard classifier.

In addition, we conducted experiments by SVM alone without prior fuzzy clustering. We found that the fuzzy clustering step does help SVM to achieve better classification. We show that SVM performed best when for each concept we selected two ( $K = 2$ ) clusters with the highest membership values. As mentioned earlier, this selection reduces the size of the confusion set for the SVM-based classification.

Table VI presents a comparison between respectively: reduction to two top classes, reduction to three top classes, and no reduction of the confusion set ( $K = 6$ ). In these experiments, all features were used (later in this section we will study the impact of different feature combinations on the classification; see Table VII below). In case of no reduction of the confusion set, the results of the fuzzy clustering step were still used by the SVM classifier in the form of additional features: specifically a vector of the membership functions of the given concept in all six clusters.

When we chose only one highest membership value, i.e., the strongest cluster as the final classification result for the given concept ( $K = 1$ ), then no SVM step was needed. The results for this case with different feature combinations are shown in Table VII in the leftmost numerical column. In this table, WordNet stands for the set of WordNet-based features (Section V.B.2), SenticNet stands for SenticNet similarity



feature (Section V.B.1), Lexical stands for other lexical features (Section V.B. 3, 4, 5), Background, General, Physiological, and Behavioral stands for various ISEAR data-based features (Table III and Section V.A), and Membership stands for the vector of the fuzzy clustering results: membership functions in each of the six clusters. The figures are presented based on tenfold cross-validation.

TABLE VI.  
Impact of the selection of most likely fuzzy cluster

$K$	Accuracy
$K = 1$ : Fuzzy clustering only, no SVM	83.41%
$K = 2$ : SVM applied to top two fuzzy clusters	<b>92.15%</b>
$K = 3$ : SVM applied to top three fuzzy clusters	67.45%
$K = 6$ : no reduction of confusion set	65.43%

TABLE VII.  
Precision with different feature combinations and different classifier combinations

Feature Combination								Classifier combination		
WordNet	SenticNet	Lexical	Background	General	Physiological	Behavioral	Membership	Fuzzy only (K = 1)	SVM only	Fuzzy + SVM
+								47.66%	50.59%	52.02%
	+							25.27%	29.07%	32.23%
+	+							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%

Feature Combination								Classifier combination		
WordNet	SenticNet	Lexical	Background	General	Physiological	Behavioral	Membership	Fuzzy only (K = 1)	SVM only	Fuzzy + SVM
+	+	+			+			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%	—	<b>92.15%</b>

Table VII shows that using the fuzzy membership vector as a feature for SVM increased accuracy. This fuzzy membership feature, along with all other features, gave the highest accuracy of 92.15% on the test set.

As training and test data, we used the intersection between the sets of concepts found in the WNA lists (for which we had the gold standard emotion labels) and those 3,312 SenticNet concepts found in the ISEAR texts (for which we had ISEAR-specific features); this intersection consisted of 1,202 concepts. Specifically, the system was evaluated on the 362 WNA concepts from the six WNA lists.

#### B. Impact of Psychological Factors

We experimented with different combinations of the psychological features extracted from the statements of the ISEAR dataset.

The impact of the psychological features is demonstrated in Table VI. It can be observed that all ISEAR data-based features improved the classification accuracy, and the best accuracy was obtained with all the features.

The individual ISEAR data-based features results based on the general and physiological variables show interesting insights into the effect of the features from the perspective of emotion. For example, low intensity for emotion classes of *shame* and *guilt* and high intensity for *joy*, *fear*, and *sadness* (Das and Bandyopadhyay, 2011). It should be noted that Das and Bandyopadhyay (2011) did not use any other type of psychological variables. In contrast, we use the features related to background variables and expressive behavior in addition to the general and psychological variables. One can observe that the expressive behavior contributes less as compared with other groups of variables.

Furthermore, we evaluated our results using the polarity sign. We considered *anger*, *fear*, *sadness*, and *disgust* as negative, and *joy* and *surprise* as positive emotions. This allowed us to evaluate the method on all 3,312 concepts against SenticNet data, because the polarity sign (positive or negative emotion) of all concepts present in SenticNet can be inferred from the SenticNet data. This evaluation gave 96.11% precision and 92.24% recall, which is considered to be very satisfactory result.

### C. Error Analysis

We observed a few typical situations leading to errors. One of them is detection of emotion words that are present in the same ISEAR statement but in two separate sentences. The lexical affinity between such remote occurrences does not always imply similarity between their lexical patterns, and therefore sometimes leads to wrong results.

In fuzzy c-means clustering, some concepts, such as *humiliate*, have very similar membership values in more than one cluster (in this case, those corresponding to FEAR and SADNESS). And, sometimes the SVM classifier chose the correct final emotion class for such a concept (SADNESS in this case). However, for

certain some ambiguous concepts such as *faint*, *sick*, *humble*, etc., the SVM classifier assigned incorrect emotion label.

## X. CONSTRUCTION OF EMOSENTICSPACE

Cambria *et al.* (2010a) have combined WNA with ConceptNet (Speer and Havasi, 2012), a large publicly available common-sense knowledge base that contains simple statements such as “you use a spoon for eating” or “a book is made of paper”. The combined resource, termed AffectiveSpace, contains both types of information: semantics and world knowledge on the one hand, and sentics on the other hand, which makes it a powerful resource for reasoning about real-world situations and behavior involving emotions.

Since our ESN is a superset of WNA, we extended AffectiveSpace to a combination of ConceptNet with ESN instead of WNA and, hence, obtain EmoSenticSpace. This extended resource contains the same semantic and real-world information but much more sentic information, since more words in it have affective labels.

### A. Building EmoSenticSpace

To build EmoSenticSpace, we followed the procedure proposed by Cambria *et al.* (2010a), which is briefly described below. ConceptNet is represented in the form of a labeled direct graph, with nodes being concepts such as, for example, *spoon*, *eating*, *book*, *paper*, and arcs being relations such as *UsedFor* (*spoon*–*UsedFor*→*eating*) and *MadeOf* (*book*–*MadeOf*→*paper*).

Technically, a graph can be thought of as a matrix. To perform inference on multiple matrices, blending is the most widely used technique. It allows multiple matrices to be combined in a single matrix, basing on the overlap between these matrices. The new matrix is rich in information and contains much of the information shared by the two original matrices. By means of the singular value decomposition on the new matrix, new connections are formed in source matrices based on the shared information and overlap

between them. This method enables creation of a new resource, which is a combination of multiple resources representing different kinds of knowledge.

In order to build a suitable knowledge base for affective reasoning, we applied the blending technique to ConceptNet and ESN.

First, we represented ESN as a directed graph, similarly to ConceptNet. For example, the concept *birthday party* has the associated emotion *joy*; we considered *birthday party* and *joy* as two nodes, and added an assertion *HasProperty* on the edge directed from the node *birthday party* to the node *joy*.

Next, we converted the two graphs, ConceptNet and ESN, to sparse matrices to blend them. After blending the two matrices, we performed the Truncated Singular Value Decomposition (TSVD) on the resulting matrix, to discard those components that represent relatively small variations in the data. We kept only 100 components of the blended matrix to obtain a good approximation of the original matrix. The number 100 was selected empirically: it gave the best results.

## B. Features Used in the Applications

### 1) Features Based on ESN's Emotion Labels and EmoSenticSpace

In order to demonstrate the effectiveness of both resources on such tasks as sentiment analysis, emotion detection, and personality detection, we used them to extract the following features from documents:

*Features based on ESN.* For each concept found in the document, we obtained its corresponding affective label from ESN, if any. We aggregated the individual concept labels into a six-dimensional vector for the document by counting the frequency of occurrence of each label in the document: say, if the document had five concepts labeled as JOY, the coordinate in the document vector corresponding to JOY was set to 5.

This gave us a six-dimensional integer-valued feature vector for each text document.

*Features based on AffectiveSpace and EmoSenticSpace.* For each concept found in the document, we extracted its 100-dimensional feature vector from EmoSenticSpace. We aggregated the individual concept vectors into one document vector by coordinate-wise summation:

$$x_i = \sum_{j=0}^N x_{ij} ,$$

where  $x_i$  is the  $i$ -th coordinate of the document's feature vector,  $x_{ij}$  is the  $i$ -th coordinate of the  $j$ -th concept's vector of this document, and  $N$  is the number of concepts found in the document.

This gave us a 100-dimensional real-valued feature vector for each text document. For comparison, we also computed these 100 features using AffectiveSpace instead of EmoSenticSpace.

In the experiments described in the next section, we observed that the 100-dimensional EmoSenticSpace feature vectors gave very good results. Six-dimensional ESN feature vectors performed very poor. However, combined 106-dimensional feature vectors showed very slight improvement over the 100-dimensional EmoSenticSpace vectors. Below we report the results obtained with these 106-dimensional feature vectors.

## 2) Other Features

For better results, we used a few other features for the documents in our experiments. For simplicity and given that all three tasks we considered have a common cognitive nature, we used the same set of features for all these tasks. These additional features are as follows:

- Sentic feature: the polarity scores of each concept in the text, obtained from ESN, were summed up to produce a single scalar feature.
- Part of speech features: three features defined as the number of adjectives, adverbs, and nouns in the text.

- Modification feature: the number of sentences in the text that have a word modified by a noun, adjective, or adverb in its dependency tree (we used the Stanford dependency parser).
- Negation feature: the number of sentences in the text that contain negation.

The latter feature is important because the negation can invert the polarity of the sentence.

## XI. PERFORMANCE ON APPLICATIONS

We used ESN and EmoSenticSpace in a number of affective applications. We found that these resources give significantly higher accuracy than that reported so far in the literature for these applications, specifically, sentiment analysis, emotion detection, and personality detection.

Identifying emotions, sentiment, and personality in text is a challenging task because of the ambiguity of words in the text, complexity of meaning, and interplay of various factors such as irony, politeness, writing style, as well as variability of language from person to person and from culture to culture.

Surprisingly, the same feature set, namely, the 106 features described in the previous section, worked excellently for all three tasks. For comparison, we also give the results obtained with 100-dimensional feature vectors extracted from the original AffectiveSpace. In all three experiments, we used tenfold cross-validation for evaluation.

### A. Sentiment Analysis of Text

For experiments on detecting positive and negative sentiment in texts, we used Stanford Twitter dataset (Go *et al.* 2009). This resource gives binary polarity labels (POSITIVE / NEGATIVE) for a large number of tweets. We cast this task as a binary classification task. For sentiment analysis experiment, this was binary classification. We report the results obtained with the SVM as the classifier. The accuracy obtained with SVM was superior to that obtained with other supervised classifiers we tried, such as other state-of-the-art Artificial Neural Network (ANN) and Naïve Bayes classifiers.

Table VIII shows the experimental results and presents a comparison between our approach and the highest state-of-the-art accuracy reported so far in literature. While the original AffectiveSpace performed quite poorly, our EmoSenticSpace outperformed the best state-of-the art approach.

TABLE VIII.  
Sentiment analysis on Stanford Twitter dataset

Method	Accuracy
ESN & AffectiveSpace	72.35%
Go <i>et al.</i> (2009)	83.10%
ESN & EmoSenticSpace	<b>85.05%</b>

#### B. Emotion Detection from Text

As a dataset for the emotion detection experiment, we used the ISEAR dataset. We cast the task as a seven-way classification, where the seven classes were the emotion labels used in the ISEAR dataset: ANGER, DISGUST, FEAR, GUILT, JOY, SADNESS, and SHAME. Note that these classes differ from the six emotion labels that we used in ESN.

Again, SVM was found to give comparatively better accuracy than other supervised classifiers. Table IX shows a very significant improvement achieved by our approach over the highest state-of-the-art accuracy result we are aware of. As can be seen from Table IX, despite the original AffectiveSpace significantly outperforming the best state-of-the art approach, EmoSenticSpace was found to perform even much better.

TABLE IX.  
Emotion detection on the ISEAR dataset

Method	Accuracy
Kim <i>et al.</i> (2010)	52.80%
ESN & AffectiveSpace	61.20%
ESN & EmoSenticSpace	<b>67.25%</b>



For experiments on detection personality from text, we used five-way classification according to the five personality traits described by Mathews *et al.* (2009), which are *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*, sometimes abbreviated as OCEAN, by their first letters.

We used the dataset provided by Mairesse *et al.* (2007). This dataset provides student essays along with a binary vector of personality traits of the author of each essay. The vector indicates the presence or absence of each of the five OCEAN traits, such as *openness*: yes, *conscientiousness*: no, *extraversion*: yes, etc.

We cast the task as five independent binary classification tasks. In this case, SVM only slightly outperformed other supervised classifiers in terms of accuracy. Table X shows the results of this experiment and a comparison with two best state-of-the-art approaches we are aware of; the letters in the header correspond to the OCEAN label set. In these experiments, the original AffectiveSpace's performance was only slightly lower than that of EmoSentSpace. More details on these experiments can be found in (Poria *et al.*, 2013c).

TABLE X.  
Personality detection on the essays dataset for personality detection

Method	O	C	E	A	N
Mohammad and Kiritchenko, (2012)	60.4%	56.4%	54.6%	54.0%	55.7%
Mairesse <i>et al.</i> , (2007)	62.1%	55.2%	54.9%	55.7%	57.3%
ESN & EmoSentSpace	<b>66.1%</b>	<b>63.3%</b>	<b>63.4%</b>	<b>61.5%</b>	<b>63.7%</b>

Therefore, in all our experiments the resources that we developed, ESN and EmoSentSpace, significantly outperformed existing state-of-the-art techniques.

We built the EmoSenticSpace resource by first extending the emotion labels from WNA to SenticNet, and then blending the obtained resource, which we called ESN, with ConceptNet. Below we remind the reader the various resources mentioned in this paper and their relationships with each other.

In addition, the first of the two resources was obtained via supervised machine learning. As in any typical machine learning setting, our work consisted in mapping a set of data points (in our case *vocabulary*—a list of words and multiword expressions denoting concepts), to some categories (in our case, emotion *labels*) via a set of *features* (extracted from *lexical resources*), using a source of *ground truth labels* both for training (along with unlabelled training data) and testing by tenfold cross-validation. In this section, we summarize our use of lexical resources to extract the vocabulary, the features, and the ground truth labels for building ESN.

The lexical resources and the features we used for the development of ESN are summarized in Table XI. Specifically, for the development of ESN we used the following four lexical resources:

- **SenticNet** is a dictionary that provides average polarity for words and concepts. We used it to extract both vocabulary (which was the complete vocabulary of this dictionary) and the SenticNet score-based similarity feature, which measured how similar are the polarity values associated with two words.
- **WNA Emotion Lists** are a dictionary that provides emotion labels for a small number of words and concepts. We used it as a source of both training instances (to associate fuzzy classes with specific labels, as described in Section VI.C) and ground truth testing instances.
- **The ISEAR Dataset** is a collection of short texts (statements) describing a situation in which an emotion was felt, along with 40 numerical or categorical data items for each statement specifying the emotion and describing the person that felt the emotion and other circumstances accompanying the situation. We used many of those data items as features for words and

concepts from our target vocabulary that occurred in the statements. In addition, we used the text of the statements (ignoring the data items) as a text corpus for learning various co-occurrence measures between words that occur in these texts.

TABLE XI.

Summary of vocabulary, features, and training categories used in development of ESN, along with lexical resources that provided them

Features	SenticNet	WNA	ISEAR		WordNet
			Data	Text	
Vocabulary	+				
Background			+	+	
General			+	+	
Physiological			+	+	
Behavioral			+	+	
Emotion			+	+	
SenticNet score-based similarity	+				
WordNet distance-based similarity					+
ISEAR text distance-based similarity				+	
Point-wise Mutual Information				+	
Emotional affinity				+	
Mapping classes to labels (training)		+			
Ground truth labels (testing)		+			

For development of ESN, we used the following data points, features, and categories extracted from these lexical resources:

- Data points (vocabulary of concepts to be assigned labels) were words and concepts that we aimed to map to final emotion labels. We extracted them from SenticNet’s vocabulary.
- Features of those data points (words or concepts) were as follows:
  - Features accompanying a situation when an emotion was felt: the background of the person who felt it, general characteristics of the situation, psychological arousals,

behavioral data of the person, and the name of the emotion described by the text containing the word or concept. All these features were extracted from the ISEAR dataset.

- Similarity scores: the closeness of polarity (extracted from SenticNet), numerous variants of calculation of the distance in WordNet graph, and co-occurrence frequency at different granularity (as observed in the ISEAR texts). Each similarity score contributed  $N$  (the size of the vocabulary) individual features to the feature vectors of concepts: the similarity of the given concept to every other concept in the vocabulary.
- The target categories, or emotion labels to be assigned to the concepts, were used both for training and as ground truth for testing. They were extracted from the WordNet Affect lists.

The lexical resources we used in the three applications of EmoSenticSpace (and thus indirectly of ESN) are summarized in Table XII. Specifically, in our applications we used the following seven lexical resources:

- **EmoSenticNet** (ESN) is a dictionary built in this work, which provides average polarity and emotion labels for a large number of concepts. First, we used it to extract emotion information to be blended with ConceptNet to obtain EmoSenticSpace. Then, we used it to extract the polarity score and emotion category features for our three applications.
- **ConceptNet** is a semantic network dictionary that provides relations between concepts, such as *spoon*  $\xrightarrow{\text{UsedFor}}$  *eating*. We used it to build our combined resource, EmoSenticSpace.
- **AffectiveSpace** is a resource similar to our EmoSenticSpace but built with WNA instead of our much larger ESN. We do not use this resource in our proposed method, but compare the use of our EmoSenticSpace with the use of AffectiveSpace to show the advantages of our resource.
- **EmoSenticSpace** is the second lexical resource that we built. We obtained it by blending our ESN with ConceptNet and reducing the resulting graph to a matrix of concepts by features. For each concept from ESN or ConceptNet's vocabulary, it gives 100 unnamed features, which

result from an algebraic transformation of a much larger graph obtained from ConceptNet and ESN. This resource is similar to AffectiveSpace but much richer in information.

- **Stanford Twitter dataset** is a set of Twitter samples supplied with binary polarity labels (POSITIVE / NEGATIVE). We used it as a training and testing set for our sentiment analysis application.
- **The ISEAR Dataset** described above is a collection of short texts describing a situation in which a specific emotion was felt; this emotion is specified for each text. We used it as a training and testing set for our emotion detection application.
- **Personality dataset** (Mairesse *et al.*, 2007) is a collection of student essays along with indication of whether each of the five personality traits is present or absent in the author's personality profile. We used it as a training and testing set for our personality recognition application.

For our applications, we used the following data points, features, and categories extracted from these lexical resources:

- **Data points** were documents from the corresponding dataset for each of the three tasks that we aimed to map to the labels specified in this dataset.
- **Features** of those data points (documents) were as follows:
  - Sentic and common-sense features: the count of each emotion label for all concepts in the document (6 integer features), determined with ESN; the coordinate-wise sum of the EmoSenticSpace features for all concepts in the document (100 numeric features); and the sum of the polarity scores for all concepts in the document (1 numeric feature), determined with ESN.

- Syntactic features: the count of adjectives, adverbs, and nouns in the document (3 integer features) and the counts of sentences in the document with modification constructions and with negation construction (2 integer features).
- **The target categories** for classification were those specified by the resource used for each application:
  - For the sentiment analysis application, POSITIVE vs. NEGATIVE polarity.
  - For the emotion detection application, the seven ISEAR emotion labels.
  - For the personality recognition application, five independent binary features corresponding to the OCEAN set; that is, personality recognition application consisted in five independent binary classification tasks.

TABLE XII.

Summary of the use of lexical resources as sources of features and training categories used in our applications

Application of ESN	Sources of features					Testing and training datasets		
	ESN		Concept-Net	Affective-Space	Emo-Sentic-Space	Twitter dataset	ISEAR dataset	Personality dataset
	Labels	Scores						
Building EmoSenticSpace	+		+					
Sentiment analysis	+	+		±	+	+		
Emotion detection	+	+		±	+		+	
Personality Recognition	+	+			+			+

In Table XII, the sign  $\pm$  stands for the use of AffectiveSpace only for comparison with the use of EmoSenticSpace, but not as part of our proposed methods.

### XIII. CONCLUSIONS AND FUTURE WORK

We have proposed an empirical method for assigning emotion orientation labels to the concepts of SenticNet, or, in other words, for expanding WNA to a much larger vocabulary of concepts present in

SenticNet. Thus, the resulting resource, ESN, can be thought of as augmenting SenticNet with emotion labels, or as an extension of WNA to a much larger vocabulary. The resource is publicly available.

Given that emotion orientations are fuzzy in nature, we used the fuzzy c-means clustering algorithm to initially cluster the concepts into fuzzy clusters. The results of this fuzzy clustering were used in two ways to facilitate the final hard categorization into six emotion labels: first, to reduce the confusion set to the top two labels, and second, as additional six features used by the hard classifier. For the hard classification stage we employed a state-of-the-art SVM. We exploit some novel features, such as point-wise mutual information and similarity scores, to determine emotional affinity between two concepts. We also introduced effective modifications to the conventional fitness function for fuzzy clustering. Further, we investigated the impact of different features, including psychological features, and the performance of our method with different feature combinations. The method showed 92.15% accuracy on the best combination of features.

Since ESN is an extension of SenticNet and of WNA, it can be used instead of these resources to achieve better results. Consequently, we also employed it to improve the AffectiveSpace resource by substituting the WNA. The constructed resource, EmoSenticSpace, in combination with the direct use of ESN, allowed for achieving a significant improvement over the state-of-the-art on such tasks as sentiment analysis, emotion detection, and personality detection in texts.

There are a number of directions for future work. One is the incorporation of syntactic information. Another is to develop solutions for typical error cases. The textual clues related to psychology may be included to improve the performance of the method. In our current work, we were able to assign emotion labels only to 3,312 concepts of SenticNet that appeared in the ISEAR corpus. We plan to use additional monolingual or multilingual (Sidorov *et al.*, 2011) lexical resources to be able to assign emotion labels to all concepts of SenticNet, as well as to construct emotion lexica for other languages (Arora *et al.*, 2012; Das and Bandyopadhyay, 2012; Wawer, 2012).

We plan to apply our resource to opinion mining tasks, as well as cognitive applications such as identifying emotion in music or mining opinion in narrative financial disclosures. We will also employ ESN in tasks at which SenticNet or WNA have been previously applied, to evaluate comparative performance improvements.

In this work, we reduced fuzzy clusters to a single label for each concept. However, emotion labeling should be considered as a multi-label problem, because a word can either invoke different emotions in different contexts, or invoke more than one emotion at the same time. Thus assigning only one emotion label to each concept may be misleading. For example, the concept *succumb* invokes sadness and, depending on the context, shame or fear. If we only assign it shame, we lose the additional information that this concept invokes sadness, and will report misleading information if, in a particular context, the concept invokes fear and not shame (*succumb to temptation* vs. *succumb to disease*). In the future, we will explore the possibility to use fuzzy clustering results directly or to assign to concepts multiple labels, probably weighted, and preferably anchored in the context, which can be useful for multi-dimensional opinion mining in systems, such as movie recommenders. This would also open the way to contextual polarity classification, where the same textual content can be presented with different emotional slants.

#### REFERENCES

- [1] Abbasi, A.; Chen, H.; and Salem, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems* 26(3):1–34.
- [2] Arora, P.; Bakliwal, A.; and Varma, V. 2012. Hindi subjective lexicon generation using WordNet graph traversal. *International Journal of Computational Linguistics and Applications*, 3(1):25–39.
- [3] Aue, A.; and Gamon, M. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of RANLP*.
- [4] Awad, M.; Khan, L.; Bastani, F.; and Yen I. L. 2004. An Effective support vector machine (SVMs) performance using hierarchical clustering. *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04)*, 663–667.
- [5] Banerjee, S.; and Pedersen, T. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco*, 805–810.
- [6] Bezdek, J. C. 1981. *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 256 pp.
- [7] Boley, D.; and Cao, D. 2004. Training support vector machine using adaptive clustering. In *Proc. of SIAM Int. Conf. on Data Mining*, Lake Buena Vista, FL, USA, 126–137.



- [8] Bradley, M.; and Lang, P. 1999. Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida.
- [9] Cambria, E.; and Hussain, A. 2012. Sentic Computing: Techniques, Tools, and Applications. Berlin Heidelberg: Springer, 153 pp.
- [10] Cambria, E.; Olsher, D.; and Rajagopal, D. 2014. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In Proceedings of AAAI, Quebec City.
- [11] Cambria, E.; Hussain, A.; Havasi, C. and Eckl, C. 2010a. SenticSpace: Visualizing opinions and sentiments in a multi-dimensional vector space. In: LNAI, vol. 6279, pp. 385–393, Springer.
- [12] Cambria, E.; Hussain, A.; Durrani, T.; Havasi, C.; Eckl, C. and Munro, J. 2010b. Sentic computing for patient centered applications. In Proceedings of IEEE ICSP, pp. 1279-1282, Beijing.
- [13] Cervantes, J.; Li, X.; and Yu, W. 2006. Support vector machine classification based on fuzzy clustering for large data sets. MICAI 2006, Lecture Notes in Computer Science, 572–582.
- [14] Das, D., 2011. Studies on emotion analysis at word and sentence level. Lambert Academic Publishing, 96 pp.
- [15] Das, D.; and Bandyopadhyay, S. 2011. Analyzing emotional statements—roles of general and physiological variables. In the SAAIP Workshop, 5th IJCNLP, Thailand, 59–67.
- [16] Das, D.; and Bandyopadhyay, S. 2012. Tracking Emotions of Bloggers—A Case Study for Bengali. Polibits, vol. 45, pp. 53–59.
- [17] Elliott, C. 1992. The affective reasoner: A process model of emotions in a multi-agent system. Ph.D. thesis, Institute for the Learning Sciences, Northwestern University.
- [18] Esuli, A.; and Sebastiani, F. 2006. SentiWordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC, 417–422.
- [19] Gale, W.; Church, K.; Yarowsky, D. 1992. One sense per discourse. In: Proc. of 4th DARPA Speech and Natural Language Workshop. 233–237.
- [20] Go, A.; Bhayani, R.; and Huang, L. 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1–12.
- [21] Hirst, G.; and St-Onge, D. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In C. Fellbaum, editor, WordNet: An electronic lexical database, MIT Press, 305–332.
- [22] Hu, M.; and Liu, B. 2004. Mining opinion features in customer reviews. In Proceedings of AAAI, 755–760.
- [23] Jiang, J.J.; and Conrath, D.W. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In Proceedings of the International Conference on Research in Computational Linguistics. Taiwan, 19–33.
- [24] Joachims, T. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of the European Conference on Machine Learning (ECML), Lecture Notes in Computer Science 1398:137–142.
- [25] Kim, S. M.; Valitutti, A.; and Calvo, R. A. 2010. Evaluation of unsupervised emotion models to textual affect recognition. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics, 62–70.
- [26] Leacock, C.; and Chodorow, M. 1998. Combining local context and WordNet similarity for word sense identification. In C. Fellbaum, editor, WordNet: An electronic lexical database, pages 265–283. MIT Press.
- [27] Lin, D. 1998. An information-theoretic definition of similarity. In Proceedings of the 15th International Conference on Machine Learning. Morgan Kaufmann, San Francisco, CA, 296–304.
- [28] Liu, B. 2010. Sentiment Analysis: A Multi-Faceted Problem. IEEE Intelligent Systems.
- [29] Mairesse, F.; Walker, M. A.; Mehl, M. R.; and Moore, R. K. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30(1):457–500.

- [30] Matthews, G.; Deary, I. J.; and Whiteman, M. C. 2009. *Personality Traits*. Cambridge, UK, 23–26.
- [31] Miller, A. G. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38(11):39–41.
- [32] Mohammad S.M.; and Kiritchenko, S. 2012. Using nuances of emotion to identify personality. AAAI Technical Report WS-13-01. *Computational Personality Recognition (Shared Task)*.
- [33] Mohammad, S.; and Turney, P.D. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. *Proceedings of the NAACL-HLT, Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34.
- [34] Ortony, A.; and Turner, T. J., 1990. What's basic about basic emotions? *Psychological Review* 97:315–331.
- [35] Patwardhan, S.; Banerjee, S.; and Pedersen, T. 2003. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2003, Mexico City, February*, 241–257.
- [36] Pang, B.; and Lee, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL*, 115–124.
- [37] Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In the *Proc. of EMNLP*, 79–86.
- [38] Pedersen, T.; Patwardhan, S.; and Michelizzi, J. 2004. WordNet::Similarity: Measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004, Association for Computational Linguistics*, pp. 38–41.
- [39] Peersman, C.; Daelemans, W.; and Van Vaerenbergh, L. 2011. Predicting age and gender in online social networks. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents. ACM*, 37–44.
- [40] Picard, R. 1997. *Affective Computing*. The MIT Press.
- [41] Plutchik, R. 2001. The nature of emotions. *American Scientist* 89(4):344–350.
- [42] Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., and Howard, N. 2014. Dependency-Based Semantic Parsing for Concept-Level Text Analysis. In *Proceedings of the 15th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2014, Part I. Lecture Notes in Computer Science 8403*:113–127.
- [43] Poria, S.; Gelbukh, A.; Cambria, E.; Das, D.; and Bandyopadhyay, S. 2012. Enriching SenticNet Polarity Scores through Semi-Supervised Fuzzy Clustering. *Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction, SENTIRE 2012, at 12<sup>th</sup> IEEE International Conference on Data Mining Workshops (ICDMW), Belgium. IEEE CS Press*, pp. 709–716.
- [44] Poria, S.; Gelbukh, A.; Das, D.; and Bandyopadhyay, S. 2013a. Fuzzy clustering for semi-supervised learning—Case study: construction of an emotion lexicon. *Lecture Notes in Artificial Intelligence*, vol. 7629, pp. 73–86.
- [45] Poria, S.; Gelbukh, A.; Hussain, A.; Das, D.; and Bandyopadhyay, S. 2013b. Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining. *IEEE Intelligent Systems* 28(2):31–38.
- [46] Poria, S.; Gelbukh, A.; Agarwal, B.; Cambria, E.; and Howard, N. 2013c. Common Sense Knowledge Based Personality Recognition from Text. *Lecture Notes in Artificial Intelligence* 8266:484–496.
- [47] Rao, D.; and Ravichandran, D. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of EACL*, 675–682.
- [48] Read, J. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, 43–48.
- [49] Resnik P. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Francisco, CA*, 448–453.
- [50] Scherer K. R. 2005. What are emotions? And how can they be measured? *Social Science Information*, 44(4):693–727.

- [51] Sidorov, G.; Posadas-Durán, J.-P.; Jiménez Salazar, H; and 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.
- [52] Somasundaran, S.; Wiebe, J.; and Ruppenhofer, J. 2008. Discourse level opinion interpretation. In *Proceedings of COLING*, 801–808.
- [53] Speer, R.; and Havasi, C. 2012. ConceptNet 5: A large semantic network for relational knowledge. In *Theory and Applications of Natural Language Processing*. Springer, 161–176.
- [54] Stevenson, R.; Mikels, J.; and James, T. 2007. Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods* 39:1020–1024.
- [55] Strapparava, C.; and Ozbal, G. 2010. The color of emotions in texts. *Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon, CogALex 2010, Beijing*, 28–32.
- [56] Strapparava, C.; and Valitutti, A. 2004. WordNet-Affect: An affective extension of WordNet. In *Proceedings of LREC*, 1083–1086.
- [57] Turney, P. D., Littman, Michael L. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM TIS*, 21(4):315–346.
- [58] Velikovich, L.; Goldensohn, S.; Hannan, K.; and McDonald, R. 2010. The viability of web-derived polarity lexicons. In *Proceedings of NAACL*, 777–785.
- [59] Wawer, A. 2012. Extracting emotive patterns for languages with rich morphology. *International Journal of Computational Linguistics and Applications*, 3(1):11–24.
- [60] Wiebe, J.; and Mihalcea, R. 2006. Word sense and subjectivity. In *Proceedings of COLING/ACL, Sydney, Australia*, 1065–1072.
- [61] Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39(2):165–210.
- [62] Wilson, T.; Wiebe, J.; and Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP*, 347–354.
- [63] Wu, Z.; and Palmer, M. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico*, 133–138.
- [64] Yu, H.; Yang, J.; and Han, J. 2003. Classifying large data sets using SVMs with hierarchical clusters. In *Proc. of the 9th ACM SIGKDD*, 306–315.