

Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text

Alexandra Balahur¹, Jesús M. Hermida²

¹ European Commission Joint Research Centre
Institute for the Protection and Security of the Citizen
GlobeSec - OPTIMA (OPensource Text Information Mining and Analysis)
Via Fermi 2749, T.P. 267
I-21027 Ispra (VA), Italy

² University of Alicante
Department of Software and Computing Systems
Ap. de Correos 99
E-03080 Alicante, Spain

alexandra.balahur@jrc.ec.europa.eu, jhermida@dlsi.ua.es

Abstract

Sentiment analysis is one of the recent, highly dynamic fields in Natural Language Processing. Although much research has been performed in this area, most existing approaches are based on word-level analysis of texts and are mostly able to detect only explicit expressions of sentiment. However, in many cases, emotions are not expressed by using words with an affective meaning (e.g. happy), but by describing real-life situations, which readers (based on their commonsense knowledge) detect as being related to a specific emotion. Given the challenges of detecting emotions from contexts in which no lexical clue is present, in this article we present a comparative analysis between the performance of well-established methods for emotion detection (supervised and lexical knowledge-based) and a method we extend, which is based on commonsense knowledge stored in the EmotiNet knowledge base. Our extensive comparative evaluations show that, in the context of this task, the approach based on EmotiNet is the most appropriate.

Keywords: EmotiNet, emotion detection, implicit emotion expressions, commonsense knowledge, sentiment analysis.

1. Introduction

Sentiment analysis is a recent task in Natural Language Processing (NLP) that aims at detecting and classifying sentiment expressions in text according to their polarity (semantic orientation) into different categories (usually, positive and negative). A related, more difficult task is emotion detection, which aims at labelling texts with an emotion category (e.g., “joy”, “anger”, “sadness”).

Sentiments can be expressed directly (e.g. “*I like this movie.*”), indirectly (e.g., “*It's the size of a button.*”) or implicitly, by describing a situation which points the reader towards a specific emotion (e.g., “*It took them 3 years to fix the leaky pipe.*” – pointing to “anger”; “*The bankrupt company spent 3 million on a new headquarters.*” – pointing to “anger”, as well).

Most of the research performed in the field of sentiment analysis and the related task of emotion detection has aimed at detecting explicit expressions of sentiment (i.e. situations where specific words or word combinations are found in texts).

In a first effort to overcome the issue of emotion detection from texts in which no or little lexical clues exist to mark the presence of a specific emotion (i.e., presence of words such as “joy”, “happy”, “sad”, etc.), we proposed a method to build a commonsense knowledge base

(EmotiNet; see Balahur et al., 2011) storing situations that trigger emotions, based on the principles of the Appraisal Theories (Scherer, 1989). The main idea behind our Psychology-inspired appraisal-based approach is that situations trigger emotions based on the result of the individual evaluation of their components, in accordance to “appraisal criteria” (Scherer, 1993). In order to detect the values of such criteria, each situation is represented in EmotiNet as a chain of actions, with their corresponding actors, objects, their properties and the associated emotion.

In the present article, we analyze the peculiarities of the data employed in our previous evaluation of EmotiNet (Balahur et al., 2011) and comparatively evaluate the performance of approaches that use established supervised and lexical knowledge-based methods for emotion detection versus the use of EmotiNet as emotion detection resource. Subsequently, we propose and evaluate two approaches to extend the knowledge contained in EmotiNet and show that such a method is appropriate for implicit emotion classification.

2. Related work

The approach on which we built the present research was initially put forward by Balahur et al. (2011) and is based on commonsense knowledge stored in a knowledge base and on a process of emotion detection built upon the

Appraisal Theories (Johnson-Laird and Oatley, 1989; Frijda, 1986; De Rivera, 1977). These theories have been successfully employed for emotion detection in other Artificial Intelligence areas (Gratch et al., 2009; Marsella et al., 2010).

With regard to previous approaches to spot affect in text, they include the use of models simulating human reactions according to their needs and desires (Dyer, 1987), fuzzy logic (Subasic and Huettner, 2000), lexical affinity based on similarity of contexts - WordNet Affect (Strapparava and Valitutti, 2004) or SentiWordNet (Esuli and Sebastiani, 2005), detection of affective keywords (Riloff et al., 2003) and machine learning using term frequency (Pang et al., 2002; Wiebe and Riloff, 2005). Other approaches were proposed within the SemEval 2007 Task 14: Affective Text (Strapparava and Mihalcea, 2007) – (Katz et al., 2007; Kozareva et al., 2007; Chaumartin, 2007). Here, the authors used both unsupervised, lexical knowledge-based approaches, as well as statistical unigram-based approaches. Additionally, related work on the ISEAR corpus includes the use of vectorial models (Danisman and Alpkocak, 2008).

Commonsense knowledge-based approaches were put forward by Liu et al. (2003) and within the framework of “Sentic Computing” (Cambria et al., 2009).

Finally, additional references to related work as far as knowledge bases and appraisal theories are concerned are presented in Balahur et al. (2011).

3. Motivation and contribution

In order to illustrate the difficulty of detecting emotion from text, let us consider the following example:

“*The man killed the mosquito.*” This sentence is different, at the lexical level, only by a word from the sentence “*The man killed his wife.*” However, at the conceptual level, based on the knowledge that is common to most humans (world knowledge, as well as moral, social, cultural criteria), the action of killing a human being (i.e. the wife) is highly blamable, while the one of killing an insect (in this example, a mosquito) is not.

In the light of these considerations, we proposed and implemented EmotiNet - a knowledge base (KB) for modelling affect based on the appraisal theories (Scherer, 1989). The analysis of the results obtained motivated the contributions brought by the present work.

A first contribution of the present work is to analyze the characteristics of the ISEAR corpus employed in previous experiments, with respect to those of the existing lexical resources that are used for emotion detection in NLP - WordNet Affect (WNA) and the emotion categories (anger, anxiety, sadness) in the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2007).

Our second contribution resides in testing two widely-used methods for emotion detection: a) one that is supervised, using Support Vector Machines Sequential Minimal Optimization (Platt, 1999) learning with uni-, bi- and trigrams and similarity of the examples among themselves); and b) another that is lexicon knowledge-based, which uses SVM SMO, but taking into account only the emotion-related words found in WNA and LIWC.

Our third contribution consists in extending the

knowledge in EmotiNet with two types of information: a) the first one is the information on additional situations that based on commonsense knowledge trigger emotion (i.e., we require more commonsense knowledge from existing repositories); b) the second source of knowledge is related to the surface realization of the textual presentation (i.e., because the same situation can be described using different linguistic expressions).

Finally, we comparatively analyze the performance of the methods presented and discuss their advantages and limitations.

4. ISEAR - a Corpus of Self-reported Affect: Dataset Analysis

4.1 Redefining the task

Self-reported affect is the most commonly used paradigm in Psychology to study the relationship between the emotional reaction and the appraisal preceding it (Scherer, 2001). ISEAR¹ (Scherer and Wallbott, 1997), a corpus of self-reported affect, contains examples of situations in which their participants had experienced all of 7 major emotions (joy, fear, anger, sadness, disgust, shame, and guilt), without mentioning the emotion explicitly. An example of entry in the ISEAR databank is: “*I lent my car to my brother and I had to pay the fine for the speeding ticket he got.*”. Each example is attached to one single emotion (e.g. “anger” in the case of the previous example).

For our experiments, we employed the 1081 examples used in our previous work (Balahur et al. 2011) that relate to family situations. As 175 were used to construct the core knowledge in EmotiNet, we will only use for testing the remaining 895 examples² to test the approach by Balahur et al. (2011). In order to study to what extent existing lexical knowledge-based and statistical methods can successfully be employed for this task, we have analyzed the corpus characteristics: number of examples per emotion, number of tokens and number of unique tokens and, additionally, the number of words found in two of the most relevant resources for emotion detection - WNA and LIWC.

Table 1 presents the characteristics of the 1081 examples we previously employed.

| Emotion | #ex | #tok | #utk | #wW | #wL | #uW | #uL |
|---------|-----|-------|------|-----|-----|-----|-----|
| anger | 174 | 5,074 | 879 | 141 | 70 | 35 | 31 |
| disgust | 87 | 2,291 | 554 | 61 | 32 | 24 | 14 |
| fear | 110 | 3,525 | 669 | 96 | 52 | 33 | 18 |
| guilt | 223 | 6,903 | 967 | 184 | 79 | 49 | 36 |
| joy | 76 | 1,894 | 437 | 51 | 3 | 20 | 3 |
| sadness | 292 | 6,360 | 847 | 181 | 117 | 46 | 28 |
| shame | 119 | 3,299 | 640 | 67 | 30 | 34 | 18 |

Table 1. Characteristics of the ISEAR examples used in our experiments.

¹ <http://www.unige.ch/fapse/emotion/databanks/isear.html>

² For 11 examples, the Semantic Role Labeling system employed - proposed by Moreda et al. (2007) had a void output.

Where:

- #ex = Number of examples
- #tok = Number of tokens in all examples
- #utk = Number of unique tokens in all examples
- #wW = Number of words in examples found in WordNet Affect
- #wL = Number of words in examples found in LIWC
- #uW = Number of unique words in examples found in WNA
- #uL = Number of unique words in examples found in LIWC

As can be seen from Table 1, the number of words in the examples previously employed by Balahur et al. (2011) that can be found in WordNet Affect or LIWC is very small compared to the total number of words in the ISEAR dataset employed. In the following section, we briefly present the EmotiNet knowledge base.

5. An Overview of EmotiNet

EmotiNet is a KB aiming to be a resource for detecting emotions in text. EmotiNet captures and stores emotional reaction to real-world situations in which commonsense knowledge plays a significant role in the affective interpretation, such as the ones presented in ISEAR. Within the KB, each situation is specified as chains of actions and their corresponding emotional labels from several situations in such a way that it facilitates the extraction of general patterns of appraisal. Action chains are sequences of action links, or simply actions, that trigger an emotion on one or more subjects. Each specific action link is described with a tuple (*actor, action type, patient, emotional reaction*). For example, for the situation “*I failed my exam because I did not study enough*”, the action chains are (*I, fail, exam, anger*), (*I, study, ?, guilt*){*not, enough*} and the final emotion label of the situation is “*guilt*”.

The process followed in the development of EmotiNet, as explained by Balahur et al. (2011), comprised the next stages:

1. The design of the EmotiNet ontology, which specifies the main concepts, properties and relations managed by the KB, capturing, combining and managing knowledge from three domains: a) kinship relations, based on a *family ontology*; b) emotions and their relations, modeled in the *emotion ontology*, which describes emotions and their relationships according to Robert Plutchik's wheel of emotion (Plutchik, 2001) and Parrot's tree-structured list of emotions (Parrot, 2001); and c) actions (characteristics and relations between them, using the *ReiAction ontology*). These three knowledge cores were combined into the EmotiNet ontology by means of a set of new classes and relations that interconnect the

components reused from them (Figure 1).

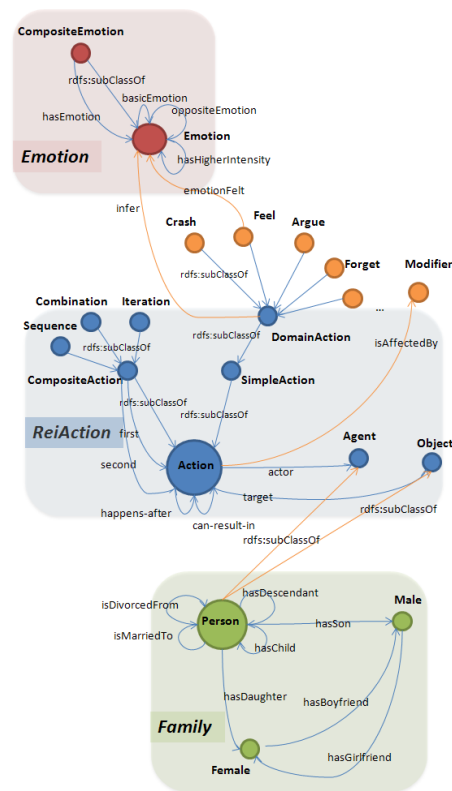


Figure 1. Main concepts and relations of EmotiNet (RDF-like schema).

2. The extension and population of this ontology using the situations stored in ISEAR database, carried out using the situations contained in the ISEAR database as examples (test set T, described in Section 6.1). These examples were transformed into 175 action chains of 4-tuples (actor, action, object, emotion) using the manual correction of the output of *Semrol* (the Semantic Role Labeling, SRL, system introduced by Moreda et al., 2007), a process of shallow anaphora resolution and a temporal sorting of actions based on a set of patterns based on adverbial expressions (e.g. “although”, “because” or “when”), establishing which action happens prior to or after the current context. The actions contained in the chains were mapped (if they existed) or added as concepts in the KB. All the action chains that represented a situation were grouped using instances of the class *Sequence* ended by an instance of the class *Feel*, which determines the final emotion felt by the main actor(s) of the chain.
3. The expansion of the EmotiNet KB using existing commonsense KBs – ConceptNet – and other resources – VerbOcean (Chlovski and Pantel, 2004) and SentiWordNet.

In the following section, we present a set of experiments we performed on ISEAR using well-established supervised and lexical knowledge-based approaches.

6. Experiments on ISEAR using supervised and lexical knowledge-based methods

6.1 Data sets

In order to test the performance of alternative methods for emotion detection, we will consider, on the one hand, the whole set of 1081 examples initially chosen by Balahur et al. (which we denote by *set A*), as well as the reduced set of 895 examples which has been employed to test EmotiNet (test *set B*). The 175 examples used to build the initial core of knowledge in EmotiNet will be denoted as *set T*.

6.2 Emotion Detection in Text Using Lexical Similarity

The first experiment we performed, we aimed at assessing if the similarity of the lexica used in the examples is high enough in order to produce a correct classification of the emotions described. In order to assess the similarity, we computed the Lesk distance between all examples (with one another) in test set A using Ted Pedersen's Statistics package³.

Subsequently, each of the examples in this set was represented as a vector, whose components were the similarities with all texts in test set A. We applied SVM SMO and performed a ten-fold cross-validation. The results are presented in Table 2.

| Emotion | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| anger | 0.353 | 0.414 | 0.381 |
| disgust | 0.292 | 0.241 | 0.264 |
| fear | 0.482 | 0.491 | 0.486 |
| guilt | 0.462 | 0.386 | 0.421 |
| joy | 0.439 | 0.474 | 0.456 |
| sadness | 0.707 | 0.76 | 0.733 |
| shame | 0.441 | 0.412 | 0.426 |

Table 2. Results of 10-fold cross validation using SVM SMO and inter-example similarity features on test set A.

Comparing these results with the ones previously obtained in the approach using EmotiNet (Balahur et al., 2011), we can see that this approach has a similar performance. However, the knowledge contained in EmotiNet is only the one extracted by modelling the initial core - i.e. test set T. Therefore, the only just comparison that can be done is by repeating the previous experiment, but computing the similarity of examples only with the ones in test set T, using test set T for training and classifying the 895 examples in test set B. The results of these experiments are reported in Table 3.

| Emotion | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| anger | 0.259 | 0.282 | 0.27 |
| disgust | 0.132 | 0.061 | 0.083 |
| fear | 0.265 | 0.086 | 0.129 |

³ <http://www.d.umn.edu/~tpederse/text-similarity.html>

| Emotion | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| guilt | 0.272 | 0.335 | 0.3 |
| joy | 0.143 | 0.203 | 0.168 |
| sadness | 0.512 | 0.583 | 0.545 |
| shame | 0.263 | 0.238 | 0.25 |

Table 3. Results of classifying test set B using SVM SMO and inter-example similarity with test set T.

As we can see from the results in Table 3, the performance when training only on the examples which in fact are used as initial knowledge in EmotiNet drop dramatically.

6.3 Emotion Detection in Text Using Affect Lexica

In order to test the appropriateness of using existing lexical resources for this task (i.e. WordNet Affect – WNA - and LIWC), we subsequently performed a series of experiments in which we represented the examples in test set A, B and T as vectors whose features accounted for the presence of words from the two lexical resources and then applied SVM SMO. Due to space limitations, we only present the results obtained when combining the two vocabularies. Table 4 presents the results obtained when performing a ten-fold cross-validation on test set A. Table 5 presents the results obtained when training on set T and testing on set B.

| Emotion | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| anger | 0.610 | 0.284 | 0.388 |
| fear | 0.712 | 0.330 | 0.451 |
| disgust | 0.692 | 0.202 | 0.313 |
| guilt | 0.559 | 0.293 | 0.385 |
| joy | 0.895 | 0.218 | 0.351 |
| sadness | 0.336 | 0.895 | 0.489 |
| shame | 0.500 | 0.066 | 0.117 |

Table 4. Results of ten-fold cross-validation on test set A using SVM SMO and words in WNA & LIWC.

| Emotion | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| anger | 0.405 | 0.201 | 0.269 |
| fear | 0.457 | 0.165 | 0.242 |
| disgust | 0.933 | 0.175 | 0.295 |
| guilt | 0.207 | 0.772 | 0.326 |
| joy | 0.204 | 0.172 | 0.135 |
| sadness | 0.667 | 0.188 | 0.293 |
| shame | 0.243 | 0.085 | 0.126 |

Table 5. Results of classifying test set B using SVM SMO and the words in WNA & LIWC using set T as training.

In this case, there is a significant drop in performance and the results are lower than the ones obtained by EmotiNet.

6.4 Emotion Detection in Text Using Supervised Learning with N-gram Features

Finally, in the following set of experiments we performed, we represented each example as feature vector, whose values (0 or 1) accounted for the presence of unigrams,

bigrams, trigrams (separately) and jointly (unigrams and bigrams - u+b; unigrams, bigrams and trigrams - u+b+t). We extracted these five different representations for test set A and performed a ten-fold cross-validation in each case (Table 6).

| Emotion | Unigrams | | Bigrams | | Trigrams | |
|---------|----------|------|---------|------|----------|------|
| | P | R | P | R | P | R |
| anger | 0.38 | 0.42 | 0.37 | 0.28 | 0.38 | 0.16 |
| disgust | 0.50 | 0.49 | 0.49 | 0.32 | 0.59 | 0.17 |
| fear | 0.67 | 0.75 | 0.45 | 0.77 | 0.39 | 0.85 |
| guilt | 0.47 | 0.53 | 0.47 | 0.55 | 0.43 | 0.48 |
| joy | 0.42 | 0.35 | 0.38 | 0.23 | 0.39 | 0.16 |
| sadness | 0.6 | 0.38 | 0.54 | 0.26 | 0.54 | 0.17 |
| shame | 0.28 | 0.16 | 0.40 | 0.07 | 0.38 | 0.03 |

| Emotion | Unigrams + Bigrams | | Unigrams + Bigrams + Trigrams | |
|---------|--------------------|------|-------------------------------|------|
| | P | R | P | R |
| anger | 0.41 | 0.37 | 0.44 | 0.38 |
| disgust | 0.54 | 0.41 | 0.57 | 0.09 |
| fear | 0.55 | 0.80 | 0.62 | 0.36 |
| guilt | 0.50 | 0.59 | 0.48 | 0.59 |
| joy | 0.49 | 0.37 | 0.71 | 0.26 |
| sadness | 0.70 | 0.30 | 0.50 | 0.82 |
| shame | 0.41 | 0.13 | 0.51 | 0.29 |

Table 6. Results of classifying test set A using 10-fold cross-validation with SVM SMO and n-grams.

Subsequently, we extracted these five different representations for set T and B, using T as training and B as test set (i.e. the presence of n-grams was computed based on the vocabulary in T). Results of these evaluations are presented in Table 7.

| Emotion | Unigrams | | Bigrams | | Trigrams | |
|---------|----------|------|---------|------|----------|------|
| | P | R | P | R | P | R |
| anger | 0.30 | 0.37 | 0.33 | 0.13 | 0.12 | 0.01 |
| disgust | 0.42 | 0.25 | 0.40 | 0.04 | 0.40 | 0.02 |
| fear | 0.36 | 0.33 | 0.29 | 0.96 | 0.27 | 0.98 |
| guilt | 0.53 | 0.16 | 0.54 | 0.07 | 0.5 | 0.06 |
| joy | 0.50 | 0.01 | 0 | 0 | 0 | 0 |
| sadness | 0.30 | 0.21 | 0.20 | 0.05 | 0 | 0 |
| shame | 0.09 | 0.04 | 0.17 | 0.01 | 0.50 | 0.01 |

| Emotion | Unigrams + Bigrams | | Unigrams + Bigrams + Trigrams | |
|---------|--------------------|------|-------------------------------|------|
| | P | R | P | R |
| anger | 0.32 | 0.26 | 0.34 | 0.15 |
| disgust | 0.50 | 0.11 | 0.44 | 0.04 |
| fear | 0.32 | 0.93 | 0.29 | 0.98 |
| guilt | 0.57 | 0.12 | 0.70 | 0.07 |
| joy | 0 | 0 | 0 | 0 |
| sadness | 0.33 | 0.10 | 0.27 | 0.06 |
| shame | 0.08 | 0.01 | 0 | 0 |

Table 7. Results of classification of test set B using SVM SMO and n-grams as features with set T as training.

As we can see from the results obtained using the different methods presented (ten-fold cross-validation using test set A and classification using set T as training and set B as test, respectively) versus the method employing EmotiNet (Balahur et al., 2011), the approach based on commonsense performs at a comparable level to the one using knowledge extracted from the entire set A. In the cases where the knowledge employed for training is equal to the one in the EmotiNet core, the difference in performance is significant, all the other methods performing much below EmotiNet.

The results of these evaluations show that the approaches working at the word level are not capable of accurately detecting and classifying emotions from examples as the ones described in the ISEAR corpus.

7. Emotion Detection Using Extensions of EmotiNet

In order to extend the coverage of the resource, the EmotiNet ontology needs to be iteratively expanded with new types of actions and relations between actions from existing resources.

Subsequent to the extensions proposed in our previous work, we extended the EmotiNet ontology by adding new actions to EmotiNet similar to the ones included in the core. The new set of actions was obtained from three existing resources: VerbOcean, “Core” WordNet⁴ and WNA. In order to effectively carry out the task, it was considered that verbs represent the essence of actions, so that the verbs contained in these resources can be mapped into EmotiNet actions. New actions were included in EmotiNet as subconcepts of the class DomainAction and related to the initial EmotiNet action set by means of a new ontology relationship: *similarAction*.

Each resource defines the similarity between actions using different mechanisms. VerbOcean explicitly contains and manages the relationship of similarity (called *similar*) between verbs. “Core” WordNet and WordNet Affect follow the same structure as WordNet, i.e., extracting similar verbs is reduced to obtaining those verbs that are in the same synset. Given this, the mapping between the *similarAction* EmotiNet relationship and the mechanisms employed in the rest of resources is direct. The reason for using two different versions of WordNet is that each of them is aimed for a specific application and, therefore, they contain different collections of verbs. Instead of using the whole WordNet, with its known problems of ambiguity and granularity, these reduced versions can provide a simplified view of the most used verbs with their usual semantics for different tasks.

Table 8 shows a comparison between the resources used to expand the EmotiNet ontology and the ontology itself. It also illustrates the degree of overlapping existing between each resource in order to clarify the contribution of each resource to the resulting ontology. Note that the

⁴ <http://wordnet.princeton.edu/wordnet/download/>

column *Unique* contains the number of actions that are uniquely present in that specific resource and not included in the rest.

| Resource | #Act. | EN | VO | CWN | WNA | Unique |
|----------|-------|-----|-----|-----|-----|--------|
| EN | 143 | * | 83 | 100 | 7 | 28 |
| VO | 782 | 83 | * | 466 | 35 | 288 |
| CWN | 2230 | 100 | 466 | * | 51 | 1702 |
| WNA | 174 | 7 | 35 | 51 | * | 109 |

Table 8. Degree of overlapping between resources measured in terms of number of action types.

Where:

- #Act.= Number of actions in resource
- EN = EmotiNet
- VO = VerbOcean
- CWN = "Core" WordNet
- WNA = WordNet Affect
- Unique = Number of actions only contained in a resource

7.1 Experiments with EmotiNet

In the set of experiments carried out with EmotiNet, we assessed the performance of the task of emotion detection in text using EmotiNet as a resource for emotion detection in text and we analyzed the impact of the different resources used in its expansion on the final results. These experiments were divided into two collections and were aimed at improving the performance of the results we previously obtained using EmotiNet (Balahur et al., 2011):

- Experiments using the EmotiNet *action* chains. In the first collection of experiments, once the action chains are extracted from the input texts, we compute their similarity with those contained in EmotiNet. The resulting emotion has the same label as the EmotiNet action chain with the highest similarity score. When an action found in the text is not contained in EmotiNet, we use the ontology relationships to the actions imported from VerbOcean (VO), "Core" WordNet (CWN) and WNA.
- Experiments using the *emotion* component of the EmotiNet action chains. This second set of experiments is based on the use of the *infer* relationship, which associates an action to the possible emotions felt by the agents of that action. We have performed different experiments in which we used the Emotion ontology and this component to obtain the emotions associated to a chain regarding each of its individual action links.

The following subsections describe each of the experiments and the obtained results.

7.1.1 Assessing the Impact of Extending EmotiNet with Other Resources

In the first collection of experiments, we calculated the similarity between the action chains extracted from the

ISEAR corpus and the action chains contained in EmotiNet. Each experiment used different EmotiNet relationships to obtain similar actions in case the exact action was not contained in the initial version of EmotiNet. Each type of EmotiNet relationship links the original EmotiNet actions to the actions imported from one or more specific resources, i.e. VerbOcean (*similar* relation), "Core" WordNet (*CWN_similar* relation) and WordNet Affect (*wna_relation* relation). Specifically, two experiments were designed and executed:

- use the initial core of EmotiNet, which establishes a baseline for the rest EmotiNet and
- use similar actions from all the resources (EN+V+C+W).

The results obtained in this set of experiments over the ISEAR corpus, described in previous sections, are illustrated in Table 9 in terms of precision and recall .

| Emotion | EmotiNet | | EN+V+C+W | |
|---------|----------|-------|----------|-------|
| | P | R | P | R |
| Anger | 54.54 | 41.37 | 54.08 | 49.42 |
| Disgust | 38.35 | 32.55 | 48.78 | 46.51 |
| Fear | 30.00 | 21.81 | 22.22 | 18.18 |
| Guilt | 30.65 | 27.47 | 28.30 | 27.02 |
| Joy | 60.00 | 43.42 | 54.23 | 42.10 |
| Sadness | 39.32 | 23.97 | 33.98 | 23.97 |
| Shame | 51.51 | 42.85 | 46.01 | 43.69 |

Table 9. Precision and recall for each run of the first collection of experiments (1a+1b).

7.1.2 Assessing the Impact of Annotating the Action Links of EmotiNet with Existing Resources

In the second collection of experiments, we applied different methods for detecting emotions from text based on EmotiNet. These methods obtained the emotions associated to each action link and subsequently combined them by means of the Emotion core of the EmotiNet ontology through a voting process. The emotion associated to each action link is initially retrieved using the *infer* relationship from EmotiNet (see Fig. 1). In order to carry out this collection of experiments, we previously generated different versions of EmotiNet. In each of these versions, the *infer* relationship was automatically populated using two well-known resources:

- the LIWC dictionary*, and more specifically, three word categories from it, i.e. Anx (LIWC code 128), Anger (LIWC code 129), Sad (LIWC code 130); and
- WNA*.

However, these do not cover all the emotions considered by ISEAR. LIWC only contains words associated to to anxiety (as a subtype of fear), anger and sadness, and the elements of WNA are only related to five emotions: anger, disgust, fear, joy and sadness. As in the first collection of evaluations, these experiments were carried out using the initial EmotiNet core and the relations of action similarity, in this case, for VerbOcean and "Core" WordNet. We designed and executed the following experiments:

- 2a) use the initial core of EmotiNet annotated with LIWC (LIWC);
 2b) use the initial core of EmotiNet annotated with WNA (WNA);
 2c) use the initial core of EmotiNet annotated with LIWC and WNA (LIWC+WNA);
 2d) use the initial core of EmotiNet annotated with LIWC and WNA and similar actions from VerbOcean and “Core” WordNet (L+WA+V+CW).

The results for this second collection of experiments are shown in Table 10.

| Emotion | LIWC | | WNA | | LIWC+WNA | |
|---------|-----------|-------|-------|-------|----------|-------|
| | P (%) | R (%) | P | R | P | R |
| Anger | 64.70 | 12.64 | 16.21 | 3.44 | 36.2 | 12.06 |
| Disgust | 0 | 0 | 37.50 | 10.46 | 32.14 | 10.46 |
| Fear | 34.78 | 7.27 | 0 | 0 | 25.80 | 7.27 |
| Joy | 0 | 0 | 50.00 | 5.26 | 44.44 | 5.26 |
| Sadness | 78.04 | 10.95 | 11.42 | 1.36 | 48.52 | 11.3 |
| Emotion | L+WA+V+CW | | | | | |
| | P | R | | | | |
| Anger | 36.80 | 26.43 | | | | |
| Disgust | 8.82 | 6.97 | | | | |
| Fear | 10.71 | 8.18 | | | | |
| Joy | 75.67 | 36.84 | | | | |
| Sadness | 30.21 | 14.38 | | | | |

Table 10. Precision and recall (%) for each run of the second collection of experiments (2a-d).

7.1.3 Combining the Best-performing Approaches

Finally, we decided to perform another experiment which combines the two methods with the best performance (in terms of average F-measure) from the first and second collection of experiments, i.e. 1b) EN+V+C+W and 2d) L+WA+V+CW. For the cases in which the methods obtained different values, the final value that was assigned was that from experiment 1b). The results for this last experiment are represented in Table 11.

| Emotion | T (#) | C (#) | R (#) | P (%) | R (%) | F (%) |
|---------|-------|-------|-------|-------|-------|-------|
| Anger | 174 | 100 | 159 | 62.89 | 57.47 | 60.06 |
| Disgust | 86 | 43 | 83 | 51.80 | 50.00 | 50.88 |
| Fear | 110 | 27 | 95 | 28.42 | 24.54 | 26.34 |
| Guilt | 222 | 60 | 214 | 28.03 | 27.02 | 27.52 |
| Joy | 76 | 46 | 59 | 77.96 | 60.52 | 68.14 |
| Sadness | 292 | 98 | 206 | 47.57 | 33.56 | 39.36 |
| Shame | 119 | 52 | 114 | 45.61 | 43.69 | 44.63 |
| Average | 154 | 61 | 133 | 48.90 | 42.40 | 45.27 |

Table 11. Results from the combination of the best performing approaches (1b and 2d).

Where:

- T = Total; C = Correct R = Number of examples with a Result; P = Precision;

R = Recall; F = F-Measure.

8. Conclusions and Future Work

From the results obtained in the experiments with EmotiNet versus well-established methods that we have presented herein, we can conclude that the task of emotion detection from texts such as the one in the ISEAR corpus (where little or no lexical clues of affect are present) can be best tackled using approaches based on commonsense knowledge. In this sense, EmotiNet, apart from being a precise resource for classifying emotions in such examples, has the advantage of being extendable with external sources, thus increasing the recall of the methods employing it. As such, we have shown that by adding knowledge from lexical resources (WordNet Affect, LIWC), we were able to further increase the performance of the approach using EmotiNet.

With the extensive evaluations we have performed, we have shown that by using EmotiNet, even with a small quantity of knowledge, we obtain comparable results to the methods that employ supervised learning or lexical knowledge on a much greater training set.

From the comparisons among the different settings and experiments, we can conclude that the approach using EmotiNet is valid and appropriate for the detection of emotions from contexts where no affect-related words are present.

A further source of errors remained the lack of knowledge on specific actions and the need to include modifiers in the heuristics used. The knowledge in EmotiNet must be even further extended using existing knowledge bases or applying automatic methods that have been proven successful in other approaches for knowledge base population.

Finally, other errors remained as a result of the NLP processes, which propagated at various steps of the processing chain. In this sense, we contemplate the use of alternative tools and methods (e.g., syntactic parsing instead of SRL) and additional usage of alternative evaluation (i.e., an assessment of the quality of the knowledge acquired).

Future work aims at extending the model by new knowledge from sources such as CYC and using patterns extracted from high quantities of online subjective texts. Additionally, we intend to expand the knowledge in EmotiNet to other languages and domains, making it a reliable resource for emotion detection from any type of text.

Acknowledgments

This paper has been supported by the Spanish Ministry of Economics and Competitiveness (grant no. TIN2009-13391-C04-01), by the Spanish Ministry of Education under the FPU Program (AP2007-03076), and by the Valencian Ministry of Education (grant no. PROMETEO/2009/119 and ACOMP/ 2010/288).

9. References

- Andreevskaia, A. & Bergler, S. (2007). Clac and Clac-NB: knowledge-based and corpus-based approaches to sentiment tagging. In Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07, pages 117–120, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Balahur, A., Hermida, J. M. & Montoyo, A. (2011). Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE Transactions on Affective Computing*.
- Calvo, R. A. & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods and their applications. *IEEE Transactions on Affective Computing*, 1(1).
- Cambria, E., Hussain, E., Havasi, C. & Eckl, C. (2009). Affective space: Blending common sense and affective knowledge to perform emotive reasoning. In Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA).
- Chaumartin, F.-R. (2007). Upar7: a knowledge-based system for headline sentiment tagging. In Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07, pages 422–425.
- Chklovski, T. & Pantel, P. (2004). Verbocean: Mining the web for fine-grained semantic verb relations. In Proceedings of EMNLP 2004, pages 33–40.
- Danisman, T. & Alpkocak, A. (2008). Feeler: Emotion classification of text using vector space model. In Proceedings of the AISB 2008 Convention "Communication, Interaction and Social Intelligence".
- De Rivera, J. (1977). A structural theory of the emotions. *Psychological Issues, Monograph 40*, 10(4). M. Dyer. 1987. Emotions and their computations: three computer models. *Cognition and Emotion*, 1(1):323–347.
- Esuli, A. & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss analysis. In Proceedings of CIKM 2005, pages 617–624.
- Frijda, N. (1986). *The emotions*. Cambridge University Press.
- Gratch, J., Marsella, S., Wang, N. & Stankovic, B. (2009). Assessing the validity of appraisal-based models of emotion. In Proceedings of ACII, Amsterdam, Netherlands.
- Johnson-Laird, P. N. & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, (3):81–123.
- Katz, P., Singleton, M. & Wicentowski, R. (2007). Swat-mp: the Semeval-2007 systems for task 5 and task 14. In Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval'07, pages 308–313, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kozareva, Z., Navarro, B., Vazquez, S. & Montoyo, A. (2007). Ua-zbsa: a headline emotion classification through web information. In Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07, pages 334–337, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Liu, H., Lieberman, H. & Selker, T. (2003). A model of textual affect sensing using real-world knowledge. In Proceedings of IUI 2003.
- Marsella, S., Gratch, J. & Petta, P. (2010). Computational models of emotion. In *A blueprint for a affective computing: A sourcebook and manual* (Scherer, K.R., Bnziger, T., and Roesch, E. (Eds.)). Oxford: Oxford University Press.
- Mei Lee, S.Y. , Chen, Y. & Huang, C.R. (2009). Cause event representations of happiness and surprise. In Proceedings of PACLIC 2009.
- Moreda, P., Navarro, B. & Palomar, M. (2007). Corpus-based semantic role approach in information retrieval. *Data and Knowledge Engineering (DKE)*, 61(3):467– 483.
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In Proceedings of EMNLP 2002, pages 79–86.
- Parrott, W. (2001). *Emotions in Social Psychology*. Psychology Press.
- Pennebaker, J.W. , Booth, R.J. & Francias, M.E. (2007). *Linguistic Inquiry and Word Count. LIWC 2007*.
- Platt, J.C. (1999). Fast training of support vector machines using sequential minimal optimization, pages 185–208. MIT Press, Cambridge, MA, USA.
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89(4):344.
- Riloff, E., Wiebe, J. & Wilson, T. (2003). Learning subjective nouns using extraction pattern bootstrapping. In Proceedings of the Conference on Natural Language Learning (CoNLL), pages 25–32.
- Scherer, K. & Wallbott, H. (1997). The ISEAR questionnaire and codebook.
- Scherer, K. (1989). *Handbook of Cognition and Emotion, Chapter Appraisal Theory*. John Wiley & Sons Ltd.
- Scherer, K. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion*, 7(3–4).
- Scherer, K. (2001). Toward a dynamic theory of emotion. The component process of affective states. *Cognition and Emotion*, 1(1).
- Strapparava, C. & Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 70–74, Prague, Czech Republic, June. ACL.
- Strapparava, C. & Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. In Proceedings of LREC, volume 4, pages 1083–1086.
- Subasic, P. & Huettner, A. (2000). Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, (9):483–496.
- Wiebe, J. & Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), pages 73–99.