Two Different Approaches to Automated Mark Up of Emotions in Text

Virginia Francisco, Raquel Hervás, Pablo Gervás
Dep. Sistemas Informáticos y Programación
Universidad Complutense de Madrid
Madrid, Spain
email: {virginia, raquelhb}@fdi.ucm.es, pgervas@sip.ucm.es

Abstract

This paper presents two different approaches to automated marking up of texts with emotional labels. For the first approach a corpus of example texts previously annotated by human evaluators is mined for an initial assignment of emotional features to words. This results in a List of Emotional Words (LEW) which becomes a useful resource for later automated mark up. The mark up algorithm in this first approach mirrors closely the steps taken during feature extraction, employing for the actual assignment of emotional features a combination of the LEW resource and WordNet for knowledge-based expansion of words not occurring in LEW. The algorithm for automated mark up is tested against new text samples to test its coverage. The second approach mark up texts during their generation. We have a knowledge base which contains the necessary information for marking up the text. This information is related to actions and characters. The algorithm in this case employ the information of the knowledge database and decides the correct emotion for every sentence. The algorithm for automated mark up is tested against four different texts. The results of the two approaches are compared and discussed with respect to three main issues: relative adequacy of each one of the representations used, correctness and coverage of the proposed algorithms, and additional techniques and solutions that may be employed to improve the results.

1 Introduction

The task of annotating text with specific labels indicating its emotional content or inclination is fundamental for any attempt to make computer interfaces respond in some way to the affective nature of the content they are handling. This is particularly true for research attempts to produce synthesised voice with different emotional states, but it may also be applicable in other contexts, such as multimodal presentation, where colors, typography or similar means can be used to convey emotion.

A comprehensive definition of emotion must take into account the conscious feeling of the emotion, the processes that appear in the nervous system and in the brain and the expressive models of the emotion [6]. Two issues must be addressed when experimenting in this field: to obtain a corpus of emotionally annotated texts to act as reference data, and to decide on a particular representation of emotion. For the study of emotional texts we need to decide which emotions we are going to model, and how we are going to represent them. There are different methods for representing emotions [2]:

- Emotional categories: The most common description of emotions is the use of emotion-denoting words, or category labels. Human languages have produced extremely powerful labels for emotional states.
- Descriptions based on psychology: The appraisal of a stimulus determines the significance of stimulus for the individual, and triggers and emotion as an appropriate response [1].
- Descriptions based on evaluation: These theories describe the emotions from the point of view of the evaluations involved.
- Circumflex models: Emotional concepts are represented by means of a circular structure [10] such that two emotional categories being close in the circle represents the conceptual similarity of these two categories.
- Emotional dimensions: Emotional dimensions represent the essential aspects of emotion concepts. Evaluation (positive/negative) and activation (active/passive) are the main dimensions; sometimes they are augmented with the power dimension (dominant/submissive). This approach is very useful because it allows measurement of the similarity between different emotional states. Another important property of this method is the relative arbitrarily in naming the dimensions.

For the particular purpose contemplated in this paper we have chosen emotional categories. One of the issues we want to discuss is how large a number of emotions is needed for adequate coverage. Too large a set of emotions makes the task very complex. Several approaches have been proposed in the literature for reducing the number of emotion-denoting adjectives:

- Basic emotions: There is general agreement that some full-blown emotions are more basic than others. From this point of view, the basic emotions appears in every person. The number of basic emotions is usually small (in early studies 10, in more recent ones between 10 and 20), so it is possible to characterize each emotional category in terms of its intrinsic properties [2].
- Super ordinate emotional categories: Some emotional categories have been proposed as more fundamental than others on the grounds that they include the others. Scherer [11] and Ortony suggest that an emotion A is more fundamental than other emotion B if the set of evaluation

components of the emotion A are a subset of the evaluation components of the emotion B. An example that may clarify the idea: five prototypes are proposed as underlying all emotional categories: anger, love, joy, fear and sadness. Joy, for example, would be subdivided into pride, contentment, and zest. Cowie and Cornellius [2] give a short overview of recent proposals of such lists.

- Essential everyday emotion terms: A pragmatic approach is to ask for the emotion terms that play an important role in everyday life. The approach is exemplified by the work of Cowie [3], who proposed a Basic English Emotion Vocabulary. Starting from lists of emotion terms from the literature, subjects were asked to select a subset which appropriately represents the emotions relevant in everyday life. A subset of 16 emotion terms emerged.

The aim of this work is to present two different approaches to emotional tagging and compare the results obtained with them when marking up texts of a particular domain - simple versions of children fairy tales. The last section discusses some ideas we are working on to improve these results.

2 Annotating Text with Labels for Emotional Content

This section presents a brief review of previous work on the labeling of texts with emotions. An important decision when annotating text with emotions is which particular approach should be used to relate emotions and textual elements.

Existing approaches can be grouped in four main categories [5]: keyword spotting - text is marked up with emotions based on the presence of affective words - , lexical affinity - not only detects affective words but also assigns arbitrary words a probability of indicating different emotions -, statistical natural language processing - involves feeding a machine learning algorithm a large training corpus of text marked-up with emotions - , approaches based on large-scale real-world knowledge - this method evaluates the affective qualities of the underlying semantic content of text - and hand-crafted methods - involves modeling emotional states in terms of hand-crafted models of affect based on psychological theories about human needs, goals, and desires -. The two approaches presented in this paper are based on lexical affinity and hand-crafted methods:

- Lexical affinity: This method not only detects affective words, as keyword spotting, but also assigns words a probability, obtained from a corpus, of indicating different emotions. The weaknesses of this approach are mainly two: it operates only at the word-level, so it can easily have problems when emotional words appear within the scope of negation; and lexical affinity is obtained from a corpus, which makes it difficult to develop a reusable, domain-independent model.

- Hand-Crafted method: This method involves modeling emotional states in terms of hand-crafted models of affect based on psychological theories about human needs, goals, and desires. This requires a deep understanding and analysis of the text. The difficulty with this approach is that it is very difficult to generalize.

On deciding the parts of the text which are going to be marked with emotions there are different options [5]: word, phrase, paragraph, chapter ... One of the simplest approaches is to have sentences as emotional structures. Another solution is to combine the sentences into large units using an algorithm to summarise the affect of text over multi-sentence regions (winner-take-all scheme, Bayesian networks ...). Our approaches mark up emotions at the sentence level.

3 EmoTag

The first method for annotating text relies on a dictionary of word to emotion assignments. This is obtained from a corpus of human evaluated texts by applying language analysis techniques. Similar techniques are later applied to assign emotions to sentences from the assignments for the words that compose them.

3.1 Construction of the Dictionary

This section deals with the process of building two basic resources for emotional mark up: a corpus of fairy tale sentences annotated with emotional information, and a list of emotional words (LEW). Both the corpus and the list of emotional words are annotated with emotional categories (happy, sad, angry ...). In this first approach all the emotional categories are available, we have not used any of the approaches for reducing the number of labels.

The method we are going to use for the mark up follows an approach based on lexical affinity. Based on a large corpus of marked up emotional text we have obtained a list of words and their relation with emotions (LEW). When we mark up a text we look for every word in this first list. If the word is not in our list we try to obtain from an ontology (WordNet) a word related to it which is in one of our emotional word lists. In the following sections we describe in detail how we have obtained the list of emotional words (LEW) and how our approach works.

3.1.1 How to Annotate the Corpus

If we want to obtain a program that marks up texts with emotions, as a human would, we first need a corpus of marked-up texts in order to analyze and obtain a set of key words which we will use in the mark up process. These texts must be marked up by different people. Each of the texts which forms part of the corpus may be marked by more than one person because assignment of emotions is a subjective task so we have to avoid "subjective extremes". In order to do that we obtain the emotion assigned to a phrase as the average of the mark-up provided by fifteen evaluators. Therefore the process of obtaining the list of emotional words involves two different phases:

- Evaluation method: Several people mark up some texts from our corpus.
- Extraction method: From the mark-up texts of the previous phase we obtain the list of emotional words.

First we had to decide which texts are going to be part of our corpus. We decided to focus the effort on a very specific domain: fairy tales. This decision was taken mainly because generally fairy tales are intended to help children understand better their feelings, and they usually involve instances of the emotions that most children experiment on their way to maturity: happiness, sadness, anger, fear ... The emotions in tales have mainly three functions: to communicate to the listener the personality of a given character to induce a certain emotional response in the listener and to communicate to the listener the emotions which given characters feels at given moments in the tale

Once the domain of the corpus' texts is established, the set of specific tales that we are going to work with must be selected. We have selected eight tales, every one of them popular tales with different lengths(altogether they result in 10.331 words and 1.084 sentences), in English. The eight tales are marked up with emotional categories: happy, sad, anger, surprise... In order to help the evaluators in the assignment of emotional categories we provide a list of different emotions. This list is only a guide, and they can add every category they need.

3.1.2 How to Extract the List of Emotional Words

Based on the tales marked up by the evaluators we obtain a data base of words and their relation to emotional categories.

Firstly we split the text into phrases and we obtain for every phrase the emotion assigned to it by most of the people, that is if one phrase has been marked as sad by 1 evaluator and as happy by 10 evaluators we considered that the phrase is happy. Phrases are processed with the qtag part-of-speech(POS) tagger, qtag [1], which assigns a part-of-speech tag (e.g. noun, verb, punctuation, etc.) to each word in a text. Every phrase is divided into words and with every word and its label we carry out the following process:

- Check if the label is in the list of stop POS tags, if it is we leave it out. Our stop list is composed of the following labels: verbs *"to be"*, *"to do"*, *"to have"* and all their conjugations, conjunctions, numbers, determiners, existential there, prepositions, modal auxiliary (*might*, *will*), possessive particles, pronouns, infinitive marker (*to*), interjections, adverbs, negative markers (*not*, *n't*), quotation mark, apostrophe...

[1] http://www.english.bham.ac.uk/staff/omason/software/qtag.html

- If the label is not in the stop list we proceed to extract the stem of the word. In order to do that we employ the Porter stemming algorithm with some changes. The Porter stemming algorithm (or "Porter stemmer") is a process for removing the most common morphological and inflexional endings from words in English [9].

- Once we have the stem of the word it is inserted into our word data base with the value 1 in the field of the emotion assigned to the phrase in which the word was; if the word was already in our list we add 1 to the field of the phrase's emotion.

- Once we have all the words in our list we carry out a normalization. We divide the numeric value we have for each of the emotions, by the number of appearances of the word in the text. In this way we have the probability of the presence of the word in a text indicating the emotions we are studying.

- Once all the tales have been processed we carry out an expansion process of our list of words. We extend our list with synonyms and antonyms. Synonyms and antonyms of every word are looked up in WordNet [7]. This process looks up all the synonyms and antonyms for every word in the list, and all of them are inserted into our data base. For inserting related words into the database, the same probabilities of the original word are used in the case of synonyms and the opposite probability is used in the case of antonyms (1- original probability).

3.2 A Method for Automated Mark Up of Emotions

Our process classifies sentences into emotions, the first step is to perform sentence detection and tokenization in order to carry out our process based in the relation between words and different emotions. In the following sections we describe how we carry out this process.

- By means of the tagger qtag, mentioned in the previous section, we obtain the POS tag for every word in the sentence. Based on these tags and words we decide the emotion of the sentence.

- If the tag associated with the word is in our label stop list we leave it out.

- If the tag is not in our stop list we get the stem of the word by means of the modified Porter stemming algorithm mentioned before.

- Once we have the stem of the word that we want to classify, we look it up in the lists of emotional words (LEW). If the word is present we assign to it the probability of carrying the emotions we are studying. Based on these probabilities of the words we obtain the final emotion of the sentence.

- If the word is not in the list we look up the hypernyms of the word in WordNet, and look them up in the LEW list. The first appearance of a hypernym is taken and the emotional content associated to our original word and the new word is inserted in the LEW list for subsequent occurrences of these words in our tales.
- If none of the hypernyms appear in the LEW list we leave out the word and it does not take part in the mark up process.
- Once all the words of the sentences have been evaluated, we add up the probability of each emotion of the different words and assign to the sentence the emotion which has a bigger probability.

A sample part of a marked tale:

...
<**anxiety**>*The knight faced the lioness.* </**anxiety**>
<**neutral**>*He fought she.* </**neutral**>
<**neutral**>*The knight threw the spear.* </**neutral**>
<**sad**>*It killed the fierce lioness.* </**sad**>
<**neutral**>*The knight drew the knife.* </**neutral**>
<**neutral**>*The knight opened the lioness.* </**neutral**>
<**happy**>*The knight resurrected the pretty blonde princess.* </**happy**>
<**delight**>*She returned to the strong castle.* </**delight**>
<**happy**>*The knight and the princess lived happy ever afterward.* </**happy**>

4 cFROGS Tagger

The cFROGS Tagger marks up text using emotional categories as does EmoTag, but not all the possible categories are used. This tagger uses one of the existing approaches for reducing the number of categories seen in Section 1: Basic Emotions. The following five basic emotions have been selected for this tagging: "happy", "sad", "angry", "fear" and "surprise". The emotion "neutral" is also used in the absence of any other emotion.

With cFROGS Tagger texts are marked up while they are being generated. In order to get tales tagged at the same time as we generate them, an existing module for automatic story generation [4] has been modified. This module generated a conceptual representation of fairy tales and its corresponding text by means of natural language generation techniques. The input of the module are the actions which take part in the story plot and the semantic information about characters, locations, attributes and relations involved in the actions. From this input the story is generated automatically.

The semantic information about the elements involved in the story is stored in a knowledge base of conceptual information about the discourse elements that appear in the input [8]. It is organized as a tree, including individuals, locations, objects, relations between them and their attributes. An extract of the knowledge base is given in Table 1.

```
character:
   human:
      no-magical:
         character(ch0,prince)
         character(ch23,hunter)
         [...]
   animal:
      can-fly:
         character(ch17,dragon)
         [...]
      cannot-fly:
         character(ch55,snake)
         character(ch28,lion)
         [...]
location:
   natural:
      location(l7,cave)
      location(l15,forest)
      [...]
   artificial:
      location(l1,palace)
      [...]
```

Table 1: Partial view of the knowledge base

The marking up of tales in our generator is carried out in the *lexicalization* stage of the natural language generation process, where it is decided which specific words and phrases should be chosen to express the domain concepts and relations which appear in the messages. Given the basic linguistic structures used by the generation module, the mark up is done by phrases. The result of the *lexicalization* stage is a list of messages with their correspondent lexical forms and the emotion they are going to be marked up with. A final stage of *surface realization* assembles all the relevant pieces into linguistically and typographically correct text.

Two elements of the tales are taken into account when deciding the emotion associated to each sentence:

- Characters.
- Actions in which the characters are involved.

4.1 Emotions Associated to Characters

Using the traditional distinction between good and evil, the characters in our stories are supposed to be involved in good, bad or neutral situations. For each case, one of the basic emotions is associated to the character. For the tale "Cinderella" the emotions in Table 2 have been considered for the main characters.

As they are the villains of the tale, for the "stepmother" and "stepsisters" the emotion assigned for the good situations is `angry`. For the hero and victim, the assigment is just the opposite.

	Good	Bad	Neutral
Cinderella	Happy	Sad	Neutral
prince	Happy	Sad	Neutral
father	Neutral	Neutral	Neutral
mother	Neutral	Neutral	Neutral
stepmother	Angry	Happy	Neutral
stepsisters	Angry	Happy	Neutral

Table 2: Emotions associated to characters

4.2 Emotions Associated to Actions

The actions are considered as good, bad or neutral situations. When choosing the emotion associated to the message representing the action, the characters involved in it are taken into account. There is a type of action that must be treated in a special way. These are the surprising actions, that are always assigned the `surprise` emotion, not taking into account its arguments. The information about the type of action is specified in the story plan received by the generation module as input.

4.3 An Example

An example of the resulting mark up with the emotions in Table 2 is:

...
<*sad*>*Cinderella lost her mother.*</*sad*>
<*angry*>*The father married a stepmother.*</*angry*>
<*sad*>*The stern stepmother made Cinderella work very hard.*</*sad*>
<*angry*>*The stepmother let the stepsisters do no work.* </*angry*>
...
<*surprised*>*The prince recognized Cinderella.* </*surprised*>
<*happy*>*The handsome prince married Cinderella.* </*happy*>

In this example the action "marry" stands out. Being considered a happy situation, it is associated with different emotions `angry` or `happy` depending on the characters involved.

5 Evaluation

Both taggers are evaluated separately, and then they are both compared in a back-to-back test for the same two stories.

5.1 EmoTag

In order to evaluate our work we carried out a test. In this test four tales are going to take part, two of them had been in our original corpus, the corpus we used to obtain our LEW list and the other two are new tales, which did not take part in our extraction method. In this way we will measure on the one hand how our process marks the tales from which we have obtained our LEW list and on the other hand how our approach works with tales that have not been involved in our extraction process. The tales which take part in this test are English popular tales with different number of words (from 153 words and 20 lines to 1404 words and 136 lines).

Each of our four tales will be marked with the emotional categories by different evaluators. Two aspects must be discussed: the results of the tagging by the human evaluators and the results of tagging by EmoTag. The results obtained are explained below.

- Evaluator's results: We have noticed that the percentage of sentences on which the majority of the human evaluators - half of their number plus one - agrees on the assignment of an emotion is very low, around 45%. This is an important data when it comes to interpreting the results obtained by our tagger. A reference value for the emotion of each phrase is obtained by choosing the emotion most often assigned to that sentence by the human evaluators.

- EmoTag's results: The reference value obtained in the evaluator's tales is used to compare with the results generated by EmoTag. The graph in Figure 1 shows the percentages of success obtained for each tale. Each sentence has been considered successfully tagged if the emotion assigned by the tagger matched the reference value.

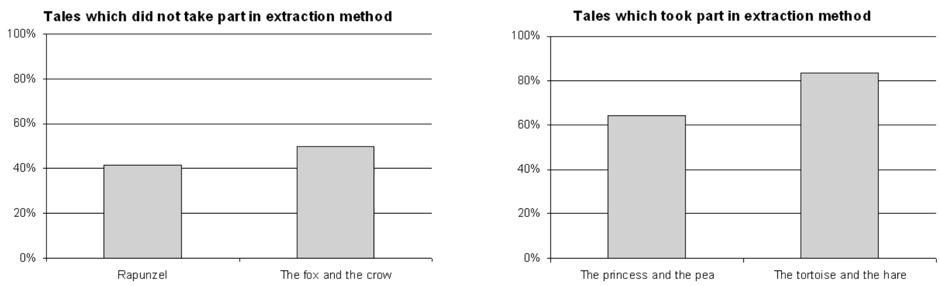

Figure 1: Percentage of success with EmoTag

Figure 2 shows the relationship between the percentage of success and the percentage of sentences whose reference value is supported by one more than half the number of evaluators.

With respect to the percentage of success we can conclude that the best results are obtained with the tales which took part in our extraction method ("The tortoise and the hare" and "The princess and the pea").

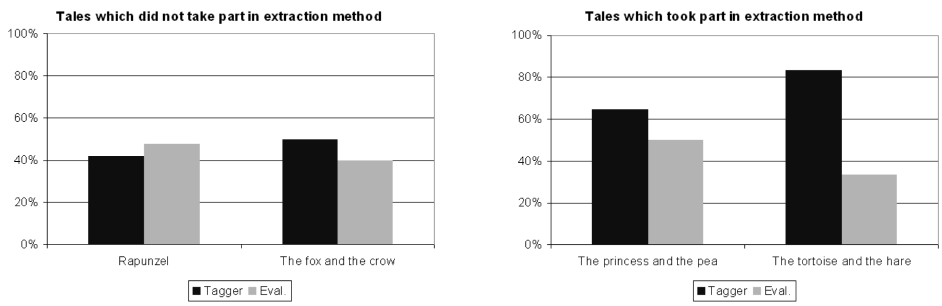

Figure 2: Success percentage and majority-supported evaluators with EmoTag

Analysis of the sentences that were tagged incorrectly indicates that most of them are either very long, include negations, or correspond to situations where human evaluators do not agree. These correspond to sentences where the reference value is not supported by at least one more than half the number of evaluators.

5.2 cFROGS Tagger

In order to evaluate our work we carried out a test with ten evaluators, in this test two tales are going to take part. These tales have been produced by cFROGS. Evaluators are asked to mark up the two tales with the six different emotions which we considered in this tagger: happy, sad, anger, surprise, fear and neutral. The results of this tests are the two tales marked with the basic emotional categories by different evaluators. Then we will obtain how we mark up every tale with our process.

- Evaluator's results: We have noticed that the percentage of sentences on which the majority of the human evaluators - half of their number plus one - agrees on the assignment of an emotion are higher than in the previous tagger, around 65%. This is an important data when it comes to comparing the results obtained by this tagger with the ones obtained in the previous tagger. A reference value for the emotion of each phrase is obtained by choosing the emotion most often assigned to that sentence by the human evaluators.

- cFROGS Tagger's results: The reference value obtained in the evaluator's tales is used to compare with the results generated by our tagger. The graph in Figure 3 shows the percentages of success obtained for each tale and the percentage of sentences whose reference value is supported by one more than half the number of evaluators. Each sentence has been considered successfully tagged if the emotion assigned by the tagger matched the reference value.

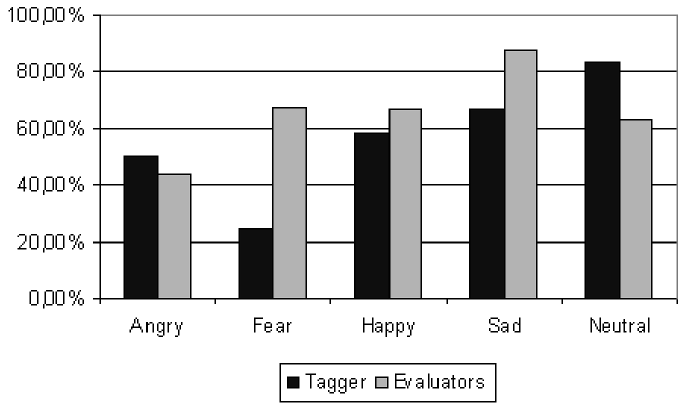

Figure 3: Success percentage and majority-supported evaluators with cFROGSTagger

Figure 4 shows the relationship between the percentage of success and the percentage of sentences whose reference value is supported by one more than half the number of evaluators in each of the tales studied.

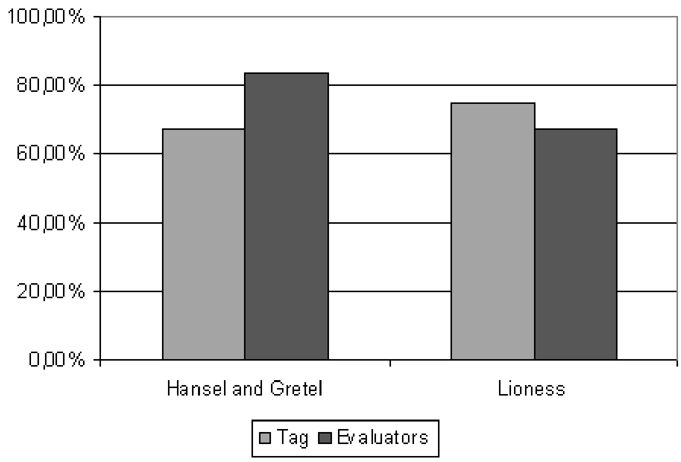

Figure 4: Percentage of success and majority-supported evaluators for the different tales with cFROGSTagger

5.3 Evaluation conclusions

With respect to the percentage of success we can conclude that in the case of EmoTag the best results are obtained with the tales which took part in our extraction method ("The princess and the pea" and "The tortoise and the hare"). If we compare the results of the two approaches we can see that if all the results are includes the best results are obtained with the cFROGS Tagger. If we compare the results of cFROGS Tagger with the subset of the results of EmoTag which correspond to tales used in the extraction method, EmoTag has a higher success rate.

Figure 5 shows the percentage of sentences in which EmoTag and cFROGS Tagger have selected the same emotion in order to mark up a sentence.

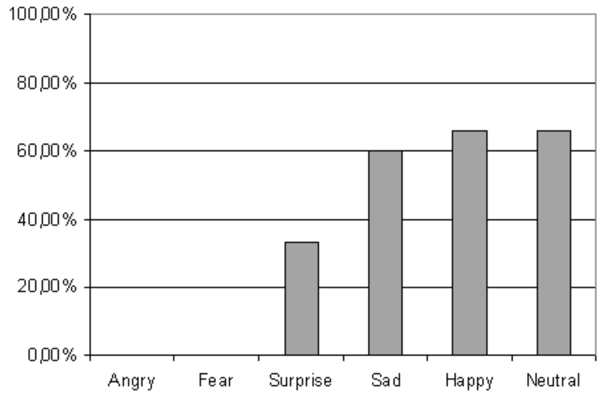

Figure 5: Percentage of sentences in which EmoTag and cFROGS Tagger mark up sentences with the same emotion

6 Conclusions

EmoTag reduced one of the disadvantages of methods based on lexical affinity. We have reduced the dependency on a given corpus by complementing our data base of emotional words with synonyms, antonyms, and hypernyms. Nonetheless, we still get better results for the tales used to obtain the LEW corpus than for new tales, so we consider necessary to continue exploring better solutions for this problem. Some issues related to context still need further work. Negation, for instance, may have the effect of inverting the polarity of the emotional content of words under its scope. We are considering the use of shallow parsing techniques to determine the scope of negations appearing in the sentences, in order to take their effect into account, both when computing word emotion from sentence emotion and viceversa.

Aside from these issues requiring improvement, we have observed that very long sentences lead to confusion when assigning emotions. In future versions we will consider a finer granularity for representing sentences.

Another problem was the large observable disagreement between human evaluators. This disagreement is lower in the cFROGS Tagger which indicates that the lower the number of emotions is, the lower the disagreement between evaluators is. We are considering reducing the number of emotions in EmoTag using one of the existing approaches mentioned in the introduction. The one which is better suited to our goal is the essential everyday emotion terms.

cFROGS Tagger obtains the better results in sentences marked by evaluators as *neutral*, *sad*, *happy* and *fear*. It is necessary to improve the results in the case of sentences marked as *angry* and *surprise*.

In future versions we will consider including some aspects of cFROGS Tagger in EmoTag, such as the reduced number of emotions and the idea of the emotion associated to characters. We can include an extra database in which we store the different characters that take part in a story and the emotion related to them.

References

[1] K. Alter, E. Rank, S. Kotz, U. Toepel, M. Besson, A. Schirmer, and A. Friederici. Accentuation and emotions - two different systems? In *Proceedings of the ISCA Workshop on Speech and Emotion*, pages 138–142, Northern Ireland, 2000.

[2] R. Cowie and R. Cornelius. Describing the emotional states that are expressed in speech. In *Speech Communication Special Issue on Speech and Emotion*, 2003.

[3] R. Cowie, E. Douglas-Cowie, and A. Romano. Changing emotional tone in dialogue and its prosodic correlates. In *Proc ESCA International Workshop on Dialogue and prosody*, Veldhoven, The Netherlands, 1999.

[4] P. Gervás, B. Díaz-Agudo, F. Peinado, and R. Hervás. Story plot generation based on CBR. In A. Macintosh, R. Ellis, and T. Allen, editors, *12th Conference on Applications and Innovations in Intelligent Systems*, Cambridge, UK, 2004. Springer, WICS series.

[5] H.Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of IUI*, Miami, Florida, 2003.

[6] C. Izard. *The face of emotion*. Appleton-Century-Crofts, New York, 1971.

[7] G. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38:39–41, 1995.

[8] F. Peinado, P. Gervas, and B. Daz-Agudo. A description logic ontology for fairy tale generation. In F. P. P. T.Veale, A.Cardoso, editor, *Procs. of the Workshop on Language Resources for Linguistic Creativity*, pages 56–61, Lisboa, Portugal, 29th May, 2004. LREA.

[9] Porter. An algorithm for suffix stripping, program. 14:130–137, 1980.

[10] J. Russell. A circumflex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[11] K. Scherer. *On the nature and function of emotion: A component process approach*. Scherer and K.R. and Ekman P and editors, Erlbaum, Hillsdale, NJ, 1984.