Procesamiento del Lenguaje Natural, Revista nº 52 marzo de 2014, pp 61-68 recibido 15-11-13 revisado 04-02-14 aceptado 05-02-14

Automatic prediction of emotions from text in Spanish for expressive speech synthesis in the chat domain

Predicción automática de emociones a partir de texto en español para síntesis de voz expresiva en el dominio del chat

Benjamin Kolz, Juan María Garrido, Yesika Laplaza
Universitat Pompeu Fabra
Roc Boronat 138,
08018 Barcelona
{benjamin.kolz, juanmaria.garrido}@upf.edu, yesika.laplaza@gmail.com

Resumen: El presente artículo describe un módulo para predecir emociones en textos de chats en castellano que se usará en sistemas de conversión texto-habla para dominios específicos. Tanto el funcionamiento del sistema como los resultados de diferentes evaluaciones realizadas a través de dos corpora de mensajes reales de chat están descritos detalladamente. Los resultados parecen indicar que el rendimiento del sistema es similar a otros sistemas del estado del arte, pero para una tarea más compleja que la que realizan otros sistemas (identificación de emociones e intensidad emocional en el dominio del chat).
Palabras clave: procesamiento de texto, detección de emociones, texto a voz, habla expresiva

Abstract: This paper describes a module for the prediction of emotions in text chats in Spanish, oriented to its use in specific-domain text-to-speech systems. A general overview of the system is given, and the results of some evaluations carried out with two corpora of real chat messages are described. These results seem to indicate that this system offers a performance similar to other systems described in the literature, for a more complex task than other systems (identification of emotions and emotional intensity in the chat domain).
Keywords: text processing, emotion detection, text-to-speech, expressive speech

1 Introduction

The generation of synthetic expressive speech for specific domains is currently a key topic in the speech synthesis field. It involves several research problems, such as the prediction of F0 and duration parameters, or the extraction of the necessary linguistic and paralinguistic information, such as emotions, from the input text.

The automatic prediction of the underlying emotions associated to the production of the input text of a text-to-speech (TTS) system is not an easy task: in some cases, meaning (at word or sentence levels) can help, but, in many other, there is no information in the utterance to establish if it was produced expressing a given emotion: only context is able to provide some clues. For this reason, most existing specific-domain TTS systems accept tags in the input text to provide this information to the linguistic processing module. However, the use of TTS in specific applications, such as reading aloud chat messages, does not allow previous tagging of texts; in those cases, automatic detection of emotions seems to be an interesting challenge.

Emotional classification of texts is a task that has been extensively attempted for different purposes, such as information extraction, text classification, sentiment analysis, and also TTS applications (García and Alías, 2008, for example). More specifically, emotion detection in online informal text, such as blogs, SMS, chat or social media texts, has also been attempted previously, especially in the field of sentiment analysis (Holzman and Pottenger, 2003; Thelwall et al., 2010; Paltoglou et al, 2013, among others). These works were mainly oriented to the classification of texts into

ISSN 1135-5948

'positive' or 'negative' categories, the location of the text in the valence/arousal space, or the identification of a limited set of emotions (generally the basic emotions inventory), and not to larger sets, which would be the task in domains such as chats. Several approaches and techniques have also been applied -for example, Emotional Keyword Spotting (EKS) or SO-PMI-IR (Semantic Orientation from Pointwise Mutual Information and Information Retrieval; Turney, 2002)-, mixing knowledge-based and machine learning (Alm, Roth and Sproat, 2005, among many others) solutions. Most work on emotion classification using linguistic information is based on the use of emotional dictionaries, which provide lists of words associated with a given emotional label or parameter (valence or arousal). Some of these works use preexisting general emotional lists, such as ANEW (Affective Norms for English Words; Bradley and Lang, 1999) or WordNet-Affect (Strapparava and Valitutti, 2004) for English, or ANSW (Affective Norms for Spanish Words; Redondo et al., 2007) for Spanish, which are the result of a manual classification by experts of generic words.

This paper describes EmotionFinder, a module for the detection of emotions in Spanish chat texts which has been implemented in TexAFon, the Python-based linguistic processing module for TTS applications developed at GLiCom (Garrido et al., 2012b). It has been developed to detect the emotional labels most represented in an annotated corpus of chat texts in Spanish, which has been used as base ('training') material for this work. It uses lexicon-based techniques similar to the ones applied in previous works (Francisco, Hervás and Gervás, 2005, for example) to identify emotions, but includes also a set of knowledge-based heuristic rules derived from the analysis of the training corpus. The emotional dictionary used in this case has been derived from a corpus of chat material, the same communicative situation in which the TTS system is expected to be used.

In the following pages, a brief description of the base material used for this work is given, the system is described, and the results of several evaluations are presented.

2 Training corpus: an annotated database of chat messages in Spanish

The work presented here is based on the analysis of a set of 4207 utterances of real chat messages in Spanish, annotated with emotional tags, which is called here the 'training corpus'. This training corpus is a subset of a more general corpus of chat conversations collected for an ongoing project on expressive synthesis in the chat domain (45 generic -without a specific topic- chat conversations, 8780 interventions). This general corpus was labeled with emotional tags by a single human annotator, using the inventory of emotions described in Garrido et al. (2012a), and then partially revised by two people different from the main annotator. The training corpus contains only the interventions representing the most frequent emotional labels found in the general corpus (16 out of 37, those showing a relative frequency beyond 1%). Table 1 presents the list of these 16 emotional tags, and the number of appearances of each one in the training corpus.

This training corpus was used for three different tasks during the development of EmotionFinder:

- the definition of the set of 8 emotions currently detected by the module, which is a subset of the emotions included in the training corpus (16);
- the development of the emotional dictionary;
- the development of the heuristic rules.

Emotion	Number of appearances
Rejection	1185
Derision	547
Happiness	495
Interest	407
Anger	371
Affection	220
Disturbance	194
Surprise	122
Pride	124
Sadness	111
Negative surprise	95
Fun	90
Admiration	90
Resignation	64
Doubt	63
Disappointment	59

Table 1: List of the 16 emotion labels covered by the training corpus.

3 EmotionFinder Overview

The current implementation of EmotionFinder is able to detect eight different emotions in the input text: admiration, affection, disappointment, interest, happiness, surprise, rejection, sadness. These labels are a subset of the most frequent emotions found in the training corpus.

It works at sentence level: it tries to assign a single emotional label (or none, if the text is considered to be 'neutral') to the sentences detected by TexAFon in the input text. It assumes a previous step of lemmatization of the words making up the input sentence (both the emotional dictionary and the rules include only lemmatized words, to improve its generalization power), which is carried out by a separate module (Lemmatizer) which has also been integrated in TexAFon as part of this project.

The EmotionFinder module includes a set of functions, one per emotion, which combine searching for key words (taken from the emotional dictionary) and regular expressions with rule based emotion inference. All these functions are applied to the input sentence one by one to check for possible cues related to the considered emotions. If a function detects one or several cues for the corresponding emotion in the input sentence, it adds the following information to the list of 'emotion candidates' of the sentence:

- the label of the candidate emotion;
- a number indicating the predicted intensity of the emotion (1, 2 or 3);
- an associated weight indicating how reliable is the cue for the detection of that emotion.

If the function finds several different hits for the same emotion in one sentence, the final weight is the sum of all of them. The final intensity value corresponds to the highest one within the found intensities in the set of detected hits. So for example, the output of the function corresponding to 'happiness' for the sentence "*Estoy feliz y encantado con el plan*" would be 'ALEGRIA(3):70' (happiness with intensity level 3, and weight 70), which would be the result of the combination of the information of two different cues detected in the sentence: 'ALEGRIA(2):40' and ' 'ALEGRIA(3):30'.

At the end of the process, the emotion label with the highest weight is selected as the sentence emotion. For example, in the case of the sample sentence "Es un buen amigo", the final list of candidate emotions would be 'ADMIRACION(1):20' and 'ALEGRIA(1):40', and the final output label would be 'ALEGRIA(1)', which is the one with the highest weight.

Figure 1 illustrates the workflow of the emotion labeling procedure in EmotionFinder.

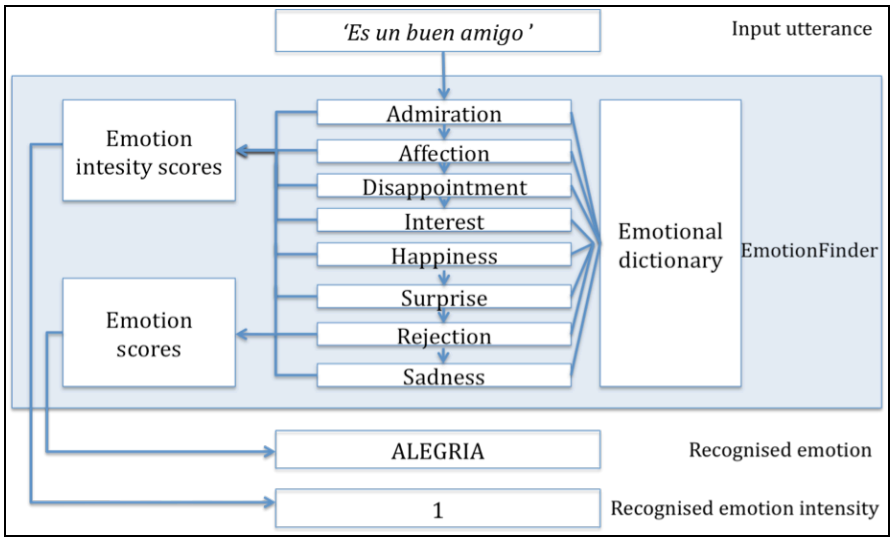

Figure 1: EmotionFinder workflow.

3.1 Emotional dictionary

In its current implementation, the emotional dictionary contains 454 entries (lemmatized isolated words and fixed expressions). Each entry contains: the lemma of the word or expression; its associated emotional label; a number (1, 2 or 3) expressing the intensity of the associated emotion; and the weight of the entry. These entries were chosen after a manual analysis of the utterances labeled with the 8 considered emotions in the training corpus, and were manually annotated with emotion, intensity and weight associated information. Table 2 shows a summary of the contents of the dictionary and table 3 gives some sample entries.

Emotion	Lemmas	Fixed Expressions/ Collocations	Total entries
Admiration	68	10	78
Affection	45	40	85
Happiness	55	25	80
Disappointment	8	5	13
Interest	7	4	11
Rejection	112	19	131
Surprise	5	7	12
Sadness	40	2	42
Anger	2	0	2
Total	**342**	**112**	**454**

Table 2: Summary of the contents of the emotional dictionary.

Entry	Intensity	Emotion	Weight
estupendo	2	admiration	50
excepcional	2	admiration	50
extraordinario	2	admiration	60
fascinar	3	admiration	70
fascinación	3	admiration	70
fenómeno	2	admiration	70
formidable	2	admiration	60
forrarse	2	admiration	60
fuerte	2	admiration	60
genial	3	admiration	60

Table 3: Sample entries of the emotional dictionary.

3.2 Emotion prediction rules

Emotion prediction rules have been implemented in the emotion recognition procedures to incorporate to the detection process some other additional information related to the identification of emotions, such as negation, comparative forms or the use of specific punctuation marks, which is not derivable by detecting single lexical items, but which can be relevant for the prediction of emotions and their intensity. For example, negation can make an emotion word not having the effect of creating an emotion like in "*No me interesa*" (emotion 'interest' is not evoked here) or even create a contrary emotion like in "*No eres una buena persona*" (not 'admiration' but 'rejection'). Then some rules to deal with polarity effects have been included. These rules have been developed using negation identifiers in regular

expressions which can already identify and treat correctly an important part of cases where negation is implied. However, other cases in which a larger scope linguistic analysis (at sentence level, for example) is needed cannot be correctly handled yet, because morphosyntactic analysis of the input sentence is not currently available.

4 Evaluation

The system was submitted to two different evaluations: the first one was carried out using a subset of the training corpus already described in section 2, and a second one with a smaller corpus of chat messages, different from the training corpus (the 'evaluation corpus'). The evaluation data presented in García and Alías (2008) have been taken as reference: this work describes a system similar to the one presented here (oriented also to the identification of emotions in a TTS task), but considers a smaller set of emotion labels, all of them contained in the basic emotions inventory ('anger', 'happiness', 'fear', 'surprise' and 'sadness'), plus the 'neutral' label. Also, the evaluation was carried out on a different domain to the one chosen for this work: 250 headlines of English newspapers. The results of that evaluation are reproduced in table 4.

Label	Precision
Neutral	0.84
Happiness	0.25
Anger	0.04
Surprise	0
Fear	0.28
Sadness	0

Table 4: Evaluation results of the system described in (García and Alías, 2008).

4.1 Training corpus evaluation

4.1.1 Procedure

The training evaluation corpus contained the subset of utterances labeled with the 8 implemented emotions (admiration, affection, disappointment, interest, happiness, surprise, rejection and sadness) within the training corpus, plus a set of 1756 neutral sentences coming from the same general corpus. The inclusion of this set of neutral sentences was motivated by to facts: first, the training corpus contains a large amount of neutral sentences, and it has been considered that their correct identification as neutral is as important as the recognition of the different considered emotions; second, the evaluation task described in García and Alías (2008) included also neutral sentences, so neutral sentences should also be considered for the evaluation of EmotionFinder in order to make evaluation results comparable. In addition, the set of sentences labeled as 'rejection' included in the evaluation corpus was reduced to 731, instead of the 1185 of the original training corpus. Then, the total number of evaluated sentences was 3991, distributed as specified in table 5.

Label	Number of sentences
Neutral	1756
Rejection	731
Happiness	495
Interest	407
Affection	220
Surprise	122
Sadness	111
Admiration	90
Disappointment	59
TOTAL	**3991**

Table 5: Contents of the training evaluation corpus.

This corpus was processed with EmotionFinder to obtain a prediction of labels, which were then compared with the emotion labels of the human annotator of the training corpus. Precision and recall values were then calculated.

4.1.2 Results

Table 6 presents the results obtained with the training evaluation corpus. A mean precision of 0.54 was obtained, with a recall of 0.49, but strong differences among emotional labels can be observed. Best results are obtained in the case of the 'interest' label (0.67), followed by the 'neutral' label (0.65). Labels showing the worst results are 'disappointment' (0.05) and 'surprise' (0.04). These results can be considered acceptable, but it has to be taken into account that they have been obtained from the same corpus from which data have been extracted to build the system.

Label	True positive	False positive	False negative	Recall	Precision	F1
Neutral	1216	762	540	0.69	0.61	0.65
Happiness	71	141	424	0.14	0.34	0.2
Admiration	49	128	41	0.54	0.28	0.37
Affection	96	205	124	0.44	0.32	0.37
Rejection	214	175	517	0.29	0.55	0.38
Surprise	4	54	118	0.03	0.07	0.04
Interest	276	143	131	0.68	0.66	0.67
Sadness	22	30	89	0.2	0.42	0.27
Disappointment	2	20	57	0.03	0.09	0.05
TOTAL	**1950**	**1658**	**2041**	**0.49**	**0.54**	**0.51**

Table 6: Results obtained with the training evaluation corpus.

4.2 Evaluation corpus

4.2.1 Procedure

The evaluation corpus was collected to test the performance of the system with a set of data different from the one used for its development. It was made of a set of 609 sentences, coming from the same source of the general corpus (real messages from chats in Spanish), but not included in the training corpus. This corpus was also annotated with emotional labels by a human annotator, different from the one who annotated the training corpus, using the same label inventory. The resulting annotation was partially revised by a second annotator, the same who labeled the training corpus, in order to check consistency in the use of the emotion labels. As the case of the training corpus, this corpus included a high amount of neutral sentences, which were used also in the evaluation task for the same reasons as in the previous evaluation. Table 7 shows the distribution of the sentences according to their label in this corpus.

Label	Number of sentences
Neutral	380
Happiness	11
Admiration	10
Affection	67
Rejection	54
Surprise	34
Interest	25
Sadness	22
Disappointment	6
TOTAL	**609**

Table 7: Contents of the evaluation corpus.

As before, the corpus was processed with EmotionFinder to obtain the prediction of labels, which were then compared with the labels added by the human annotator. Precision and recall values were again calculated.

4.2.2 Results

Table 8 presents the results obtained with the evaluation corpus. A mean precision of 0.6 was obtained, with a recall of 0.58, even better results than those obtained with the training corpus. However, a closer look to the data allows to observe that this value is mainly due to the very good results obtained in the case of the 'neutral' label (precision 0.81); the emotional labels shows clearly lower values than in the previous evaluation, with two labels ('happiness' and 'surprise') having a precision score of 0, and a maximum of 0.33 in the case of 'rejection'. These results reveal a dependency on the training corpus of the dictionary and the rules.

Label	True positive	False positive	False negative	Recall	Precision	F1
Neutral	308	90	72	0.81	0.77	0.79
Happiness	0	9	11	0	0	0
Admiration	2	28	8	0.07	0.2	0.1
Affection	20	43	47	0.30	0.32	0.31
Rejection	6	12	48	0.11	0.33	0.17
Surprise	0	2	34	0	0	0
Interest	14	38	11	0.56	0.27	0.36
Sadness	1	5	21	0.05	0.17	0.08
Disappointment	1	6	5	0.17	0.14	0.15
TOTAL	**352**	**233**	**257**	**0.58**	**0.60**	**0.59**

Table 8: Results obtained with the evaluation corpus

5 Discussion and conclusions

In this paper a new module for the prediction of emotions in chat text, oriented to the generation of emotional speech in the chat domain, has been presented. It makes use of a combination of lexical information (in the form of an emotional dictionary especially built for the system from a reference corpus) and hand-made expert rules, to attempt the identification of some of the most frequent emotions, as well as the intensity of the emotion, appearing in the emotional annotation of a corpus of chat messages. Both aspects (detection of emotions beyond the inventory of basic emotions and detection of the emotion intensity) are novel with respect to other previous systems.

The results obtained in the performed evaluations are encouraging: they are slightly better than those of the system chosen as reference, for a more complex identification task (nine emotional labels instead of the six labels of the reference system). Also, the system shows a good performance in the correct discrimination of neutral from emotional sentences, an important task in the generation of synthetic expressive speech in a specific domain situation, where neutral and emotional sentences appear mixed and they have been properly handled.

The observed differences in the results of both evaluations (better scores in emotion detection task with the training corpus than with the evaluation corpus) seem to indicate that the performance of the system is still quite dependent on the corpus used to develop the rules and the emotional dictionary. Further research should be done to enlarge the dictionary and to improve the rules to consider phenomena not included in the used training corpus. The use of morphosyntactic information could also improve the performance of the current rules, and allow the development of new ones.

6 References

Alm, C. O., Roth, D. and Sproat, R. 2005. ”Emotions from text: machine learning for text based emotion prediction”, Proceedings of HLT/EMNLP.

Bradley, M. and Lang, P. 1999. Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Technical Report C-1, Gainesville, FL, The Center for Research in Psychophysiology, University of Florida.

Francisco, V., Hervás, R. and Gervás, P. 2005. “Expresión de emociones en la síntesis de voz en contextos narrativos”. Simposio de Computación Ubicua e Inteligencia Ambiental.

García, D. and Alías, F. 2008. “Identificación de emociones a partir de texto usando desambiguación semántica”, Procesamiento del Lenguaje Natural, 40: 75-82.

Garrido, J. M., Laplaza, Y., Marquina, M, Pearman, A., Escalada, J. G., Rodríguez, M. A. and Armenta, A. 2012a. “The I3MEDIA speech database: a trilingual annotated corpus for the analysis and synthesis of emotional speech”, LREC 2012 Proceedings: 1197-1202. Online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/865_Paper.pdf, accessed on 13 November 2013.

Garrido, J. M., Laplaza, Y., Marquina, M., Schoenfelder, C. and Rustullet, S. 2012b. "TexAFon: a multilingual text processing tool for text-to- speech applications", Proceedings of IberSpeech 2012, Madrid, Spain, November 21-23, 2012: 281-289. Online: http://iberspeech2012.ii.uam.es/index.php/onlineproceedings, accessed on 13 November 2013.

Holzman, L,. and Pottenger, W. 2003. Classification Of Emotions in Internet Chat: An Application of Machine Learning Using Speech Phonemes. Technical Report LU-CSE-03-002, Lehigh University,

Paltoglou, G. Theunis, M., Kappas, A. and Thelwall, M. 2013. "Predicting Emotional Responses to Long Informal Text," IEEE Transactions on Affective Computing, 4, 1: 106-115.

Redondo, J., Fraga, I., Padrón, I. and Comesaña, M. 2007. "The Spanish adaptation of ANEW (Affective Norms for English Words)". Behavior Research Methods, 39(3): 600–605.

Strapparava, C., and Valitutti, A. 2004. "Wordnet-affect: An affective extension of wordnet". Proceedings of the Fourth International Conference on Language Resources and Evaluation: 1083–1086.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. and Kappas, A. 2010. "Sentiment Strength Detection in Short Informal Text," Journal of the American Society for Information Science and Technology, 61: 2544-2558.

Turney, P. D. 2002. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, Pennsylvania, USA: 417-424.