Hybrid Text Affect Sensing System for Emotional Language Analysis

Rafael del-Hoyo, Isabelle Hupont, Francisco J. Lacueva and David Abadía
Instituto Tecnológico de Aragón
C/ María Luna nº6,
50005 Zaragoza, Spain
rdelhoyo@ita.es, ihupont@ita.es, fjlacueva@ita.es, dabadia@ita.es

ABSTRACT
It is argued that for a computer to be able to interact with humans it needs to have the communication skills of humans. One of the main skills of human computer intelligent interaction is the affective aspect of communication, and language is one of the main ways for humans to express emotions. In order to analyze the emotions in language, it is necessary to study the general tone of conversation and the semantic content. This paper explores the influence of affective information about words in sentiment analysis and presents a hybrid statistical-semantic system for opinion detection in Spanish language texts. Affect sensing analysis in language content is very important for achieving realistic interaction with intelligent virtual agents, however it is still an unexplored field nowadays.

Categories and Subject Descriptors
D.3.3 [**Models and Principles**]: User/Machine Systems – *human factors*, h*uman information processing;* H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Natural language;*

General Terms
Algorithms, Experimentation, Human Factors.

Keywords
Affective computing, natural language processing, sentiment analysis, machine-learning algorithms.

1. INTRODUCTION
Human computer intelligent interaction is an emerging field aimed at providing natural ways for humans to communicate with computers. It is argued that for a computer to be able to understand humans it needs to have the communication skills of humans. One of these skills is the affective aspect of communication. Recognition of emotions has been implemented in many kinds of media, such as speech, image, signals, etc. Among this, natural language affect sensing is an important mean of emotional expression which results can be used in a wide field of applications, e.g. in automatic answering systems where human-like robots or virtual agents dialogue with users, so that the user feels as the system is human. The interest in the automated recognition of affect from written and spoken language is increasing with the large amount of texts becoming available on the Internet and its wide field of applications.

In general, there are two main approaches that aim at solving emotional analysis of texts (a brief discussion on available approaches is given in [1]): statistical and semantic. Statistical approaches make use of data mining methods for affect sensing in texts, for instance, by using words and n-grams counts or by measuring distances for the automatic detection of the main words that provide affect. They allow automatic affect classification, but usually produce low classification results and need a wide number of texts for training. On the other hand, semantic approaches aim at classifying affect in texts using emotional information about a set of keywords, which implies the need of hand-crafted dictionaries [2, 3] and the incorporation of lexical analysis for detecting lexical modifiers of sense and negations in the phrase-level analysis.

This paper presents a hybrid statistical-semantic system for affect sensing in Spanish language, even though the results obtained can be extrapolated to other languages. The aim of the paper is to study the influence of affective information about words in sentiment analysis. For achieving this goal, a comparison of affect sensing performance using pure statistical methods, pure semantic methods and a hybrid approach is carried out. For simplicity, throughout the paper, we consider sentiment sensing (in this case opinion classification) as the prior task of “affect” or “emotion” sensing. The rest of this article continues as follows. In section 2, an overview of the pure statistical method is given. Section 3 explains the pure affective semantic approach. In section 4, the Spanish corpus used for the system’s training is presented. Finally, section 5 carries out a comparison of affect sensing using pure statistical methods, pure semantic methods and the hybrid approach, in order to analyze the influence of affective information about words in sentiment sensing.

2. STATISTICAL ARCHITECTURE OVERVIEW
The pure statistical system developed is composed of three main blocks (in black in Figure 1). The pre-processing linguistic module is the first element block. The second element is in charge of statistics feature extraction. Finally, the last block is the classification module. In the following sub-sections, a general description of the modules is given.

AFFINE '09, November 6, 2009 Boston, MA.

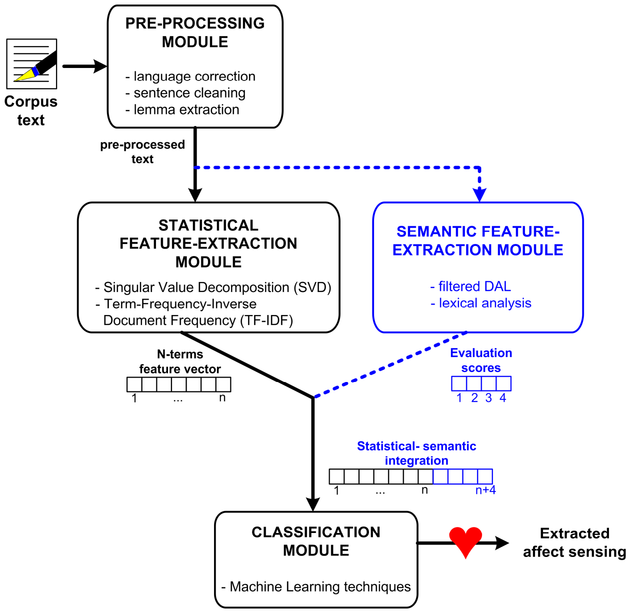

Figure 1. The hybrid system's architecture overview. The pure statistical architecture is shown in black (solid line). The added semantic module is shown in dotted blue.

2.1 Pre-processing Module

This module is responsible for language pre-processing, correction, sentence cleaning and lemma extraction. Inside this module, a sentence splitting is done for identifying sentence boundaries in each text. The second step corresponds to the tokenization which consists basically in labeling individual words, or sometimes word parts, to be clearly identified for future analysis or word cleaning. The third step corresponds to the Part-of-Speech (pos) Tagging: a pos-tag is assigned to each token. In this case, the word lemma is identified and swapped with the original word. The final step is "stop words" filtering, where very common words like "and" and "the" are filtered. Finally, the output of this module is the input for statistical feature extraction module.

2.2 Feature-Extraction Module

In this module, a statistical analysis of the texts for characteristics extraction is done. The result of this module is a n-dimensional numerical vector in which there is a representation of statistical information of the text.

The statistical features of the numeric vector are constructed in base of a term vector space model from the corpus. A document is represented as a vector. Each dimension corresponds to a separate term and a term is a word or an n-gram. If a term occurs in the document, its value in the vector is non-zero. The value is a statistical measure used to evaluate how important a term is and the importance increases proportionally to the number of times a term appears in the document but is offsetted by the frequency of the word in the corpus. Several different ways of computing these values exists, one of the best known schemes is Term-Frequency-Inverse Document Frequency (TF-IDF) which is used in our model. The dimension of the vector depends on the number of important terms of the corpus. One of the main problems of this technique is the high-dimension of this vector (normally thousand of terms are selected). In our case, the problem is solved by using a dimension reduction algorithm. The vectors are reduced from approx. 3200 to 20 dimensions. Special care has been taken to preserve the topology information relation among terms by using the Singular Value Decomposition (SVD) algorithm [4]. On one hand, this information reduction decreases the complexity of the classification task, on the other hand, it is reduced the information necessary for classification. Thus there is a direct trade-off between complexity and information reduction.

2.3 Classification Module

The extracted statistics are evaluated using machine learning techniques. From the state of the art in text mining, two algorithms have been chosen: Multilayer Perceptron [5] and Support Vector Machine (SVM) [6], both well-known and powerful tools for classification of vectors of real-valued features. Both algorithms were optimized, always looking for the smallest classification error. The classification accuracy of each algorithm was tested via supervised learning and hundreds of configurations of Neural Network and SVM algorithms were tried out. The output of the classification module is the extracted affect sensing information (classified into "positive" or "negative" labels).

3. AFFECTIVE INFORMATION

In this section, a method for the extraction of affective information in texts is presented. This information is an excellent complement for the statistical method since it bases text sentiment sensing on the finding of affective cues, in a similar way as human beings.

3.1 Dictionary of Affect in Language: a Semantic Tool for Detecting Affect in Texts

One of the most followed psychological representations of affect considers emotions as a continuous 2D space whose dimensions are "evaluation" and "activation" [7, 8, 9]. The "evaluation" dimension measures how a human feels, from positive to negative. The "activation" dimension measures whether humans are more or less likely to take an action under an emotional state, from active to passive. Taking this affective 2D representation into account, the work of Cynthia Whissell [7] provides a pair of <"activation", "evaluation"> values (ranged from 1 to 3) to each one of the approx. 9000 affective words that compose her "Dictionary of Affect in Language" (DAL). The use of the DAL turns out interesting for the present work since it gives an "evaluation" value to each word, i.e. a polarity, depending on its affective contents. That way, the sentimental sensing in terms of affective contents can be extracted.

3.2 Including Affective Information

To analyze the influence of "evaluation" dimension in texts' affect sensing, affective information is extracted from texts through the use of the Spanish translated DAL. Different syntactic categories of words can be found in the DAL: verbs, adjectives, pronouns, prepositions… However, many of them are context-dependent (e.g. the spanish word "cara" means either "face" or "expensive", depending on the text context) and others are neutral in terms of affect. In order to minimize noise in the system and avoid possible confusions, a DAL filtering has been carried out for this work. Firstly, the words with neutral affective contents, i.e. with "evaluation" values ranging from 1.5 and 2.5, have been removed. Secondly, the rest of words have been divided into several word-lists, depending on their syntactic category and their "evaluation" value (considered " very negative" if lower than 1.25, "negative" if ranging from 1.25 and 1.5, "positive" if comprised between 2.5 and 2.75, and "very positive" if greater than 2.5). In particular, only the adjectives and verbs were retrieved from the DAL, since they are the groups of words with more contribution to emotional contents and more context-independent. Examples of built word-lists are: *<very positive adjectives>, <negative adjectives>, <positive verbs>,* etc. According to this process, the DAL is filtered and therefore the dictionary to work with is obtained.

The objective now is to provide each text with a general "evaluation" annotation. In order to achieve it, a set of structures or expert rules with different priorities are defined for the lexical analysis of texts. Those rules include the detection of lexical modifiers of meaning and negations in the phrase-level analysis. Examples of implemented rules are:

> *If there is a <positive adjective> then "evaluation" is <positive>* (low priority rule)
>
> *If there is a <negative modifier> followed by a <very positive adjective> then "evaluation" is <very negative>* (high priority rule)

Where <*negative modifier*> and <*positive adjective*> are lists of negation structures and positives adjectives, correspondingly.

Once the rules have been established, the texts can be processed. Four labels are defined: "very negative evaluation", "negative evaluation", "positive evaluation" and "very positive evaluation". Every time a structure of the aforementioned is found in the text, the score of the corresponding "evaluation" label is increased by one unit. That way, a global score for each "evaluation" label is obtained as an output for each processed text.

In order to include the extracted affective information into the statistical method, the obtained "evaluation" scores are added to the classification module. This is achieved by adding 4 new locations in the term vector build in the feature extraction module: one for each defined "evaluation" labels. In doing so, the semantic affective information is unified with the statistical method and a hybrid statistical-semantic system is achieved (see Figure 1). The results obtained after integrating the statistical and semantic modules are presented and discussed in section 5.

4. CORPUS

The influence of emotional word information in affect sensing classification texts is analyzed by means of a Spanish movie review corpus [10]. The movie review corpus contains 3.878 films reviews from "www.muchocine.com", rated in Spanish by non-specialized users and mapped onto classes from one to five stars .

We must emphasize the difficulty of sentiment classification due to the fact that movies' corpus is inconsistent, sometimes grammatically incorrect, contains repetitions and inexact wordings. The movie review database has been written and rated by different people. This raises a lot of questions, e.g. if a critic has the same or at least a comparable impression of the same movie.

To decrease the complexity of this task, the corpus labeling was reduced to two classes: "positive" and "negative" affect. Furthermore, texts without a clear sentiment where removed.

5. RESULTS

Three different sets of experiments have been carried out with the Spanish corpus. The first experiment only takes into account the pure statistical algorithm. The second experiment manages only semantic information. Finally, the last experiment combines both techniques.

Table 1 represents the results obtained for the different groups of experiments using a 10 fold cross-validation. In each experiment, two well known classification algorithms in the text mining field, Support Vector Machine (SVM) [6] and Multilayer Perceptron (Neural Network – NN) [5] were adjusted to obtaining the best results. Finally, the results are compared with the best results yet published using this data [10].

Table 1. Affect sensing results for each experiment.

Experiment	Algorithm	Accuracy on positive texts	Accuracy on negative texts	Average Results
Statistic	SVM	80,15%	73,94%	77,05%
	NN	76,00%	78,26%	77,13%
Semantic	SVM	68,52%	66,09%	67,31%
	NN	69,41%	65,86%	67,64%
Hybrid	SVM	81,41%	80,30%	80,86%
	NN	80,44%	78,18%	79,31%
Fermín L. Cruz [4][1]	Multiple seeds	73%	83%	77,50%

Table 2 reflects the substantial increase of accuracy when the hybrid approach is compared with the semantic approach and exhibits an increase around 3% when compared with statistical method.

Table 2. Accuracy improvement comparison.

Hybrid vs static	**SVM**	3,81%
	NN	2,18%
Hybrid vs semantic	**SVM**	13,55%
	NN	11,68%

In conclusion, "evaluation" dimension word measurement from Whissell's dictionary allows increasing the sentimental classification capabilities. Moreover it is expected that when the number of texts decrease, the hybrid algorithm should have a substantial improvement. It is also interesting to point out that the

[1] Results were obtained with a lower number of texts and without 10 fold cross-validation.

current implementation of the hybrid system is executed in a distributed platform with a time performance in execution below 250 ms for each text.

6. CONCLUSIONS AND FUTURE WORK

The obtained results confirm that the emotional semantic features provide significant additional information for affect sensing. It has been showed that the hybrid model, based of the Dictionary of Affect in Language, has improved the performance of the pure semantic and statistical approaches, and beat the published results with this corpus. The system has been proved to work in real-time, with performances lower than 250ms per text.

Even though the movie corpus is a first approach for affect sensing analysis, our future objective is to analyze affect in spoken language. The main difference between a movie review and a dialogue is that a dialogue can be seen as a continuous stream of information between utterances that influences. Nevertheless, a large corpus allows us to validate the influence of affective dictionaries for sentimental analysis. We plan to further improve our model by studying an extensive dialog corpus. Finally, the implemented system is going to be integrated in a real-time multimodal virtual agents platform.

7. ACKNOWLEDGMENTS

This work has been partly financed by the CETVI project (PAV-100000-2007-307), 3SME project (TSI-020302-2008-69) funded by the Spanish Ministry of Industry and the Grupo de Ingeniería Avanzada (GIA) of the Instituto Tecnológico de Aragón.

8. REFERENCES

[1] Binali, H., Potdar, V., and Wu, C. 2009. A state of the art opinion mining and its application domains. 2009. IEEE International Conference on Industrial Technology, 1-6.

[2] Osherenko, A., and André, E. 2007. Lexical Affect Sensing: Are Affect Dictionaries Necessary to Analyze Affect? Proceedings of Affective Computing and Intelligent Interaction, Springer.

[3] Shaikh, M. A. M., Prendinger, H. and Ishizuka, M. 2009. A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text. Affective Information Processing, Springer 2009, 45-73.

[4] Hammarling, S. J. 1985. The singular value decomposition in multivariate statistics, ACM Signum Letter, Vol. 20, 2-25.

[5] Haykin, S. 1999. Neural Networks: A Comprehensive Foundation. Second ed, Prentice-Hall, Englewood Cliffs, NJ.

[6] Vapnik, V. 1998. Statistical Learning Theory. Wiley, Chichester, GB.

[7] Whissell, C. M. 1989. The Dictionary of Affect in Language. Emotion: Theory, Research and Experience: Vol. 4, The Measurement of Emotions, R. Plutchik and H. Kellerman Eds. New York: Academic.

[8] Plutchik, R. 1980. Emotion: a psychoevolutionary synthesis. New York: Harper & Row.

[9] Russell, J.A. 1980. A circumplex model of affect. Journal of personality and social psychology, Vol. 39, 1161–1178.

[10] Cruz, F. L., Troyano, J. A., Enríquez, F. and Ortega, J. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. Procesamiento del Lenguaje Natural, Vol. 41, 73-80.