Affect-based Features for Humour Recognition

Antonio Reyes, Paolo Rosso and **Davide Buscaldi**
Departamento de Sistemas Informáticos y Computación
Natural Language Engineering Lab - ELiRF
Universidad Politécnica de Valencia
{areyes,prosso,dbuscaldi}@dsic.upv.es

Abstract

The actual trends in NLP are focusing on analysing knowledge beyond the language: moods, sentiments, attitudes, etc. In this paper we focused on studying the importance of affectiveness information for humour recognition. Several experiments were performed over 7,500 blogs using some features reported in the literature, besides a set of new ones. A classification task was executed in order to verify the features relevance. The results indicate an interesting behaviour regarding to affective information.

1 Introduction

The actual trends in NLP are focusing on the analysis of knowledge beyond the language. Through the analysis of textual information, knowledge related to emotions, sentiments, opinions, moods or humour, has been mined with success. For instance, Opinion Mining (Ghose et al., 2007), Sentiment Analysis (Pang et al., 2002) or Computational Humour (Mihalcea and Strapparava, 2006), have shown how to take advantage of the implicit knowledge in texts for their own purposes.

In this framework, this paper is focused on studying the importance of affectiveness information for humour recognition. In particular, we concentrate on analysing a corpus of 7,500 blogs retrieved from LiveJournal and linked to humour and moods through users tags. This means we aim at considering humour beyond typical one-liners (Mihalcea and Strapparava, 2006) applying some features reported in the literature, besides a set of new ones. A selection of features were assessed through a classification task.

The paper outline is organised as follows. Section 2 underlines the initial assumptions and the objective. Section 3 describes the experiments. In Section 4 the evaluation is presented. Finally, Section 5 concludes with some final remarks and addresses the future work.

2 Affectiveness in Humour

When speaking of humour, we must be taken into account the multiple variables that produce it. For instance, the presence of antonyms, sexual information or adult slang has been stressed as a recurrent humour property (Mihalcea and Strapparava, 2006), as well as a trend to negative orientation (Mihalcea and Pulman, 2007), or the employment of semantic ambiguity as triggers of humorous effects (Reyes et al., 2009). However, other kinds of factors exist that influence the perception of humour. Emotions, sentiments or moods impact on the manner in which humour is expressed as well as on the joke effectiveness. That is why we aim at investigating what is the relevance of analysing information related to affective knowledge for humour recognition purposes. The underlying assumption is that humour is expressed in several ways profiling some particular features: jokes, punning riddles or one-liners are just a manner to verbalise humour[1]. However, there are other kinds of features that must be considered as triggers of humour. In this case, we focus on affective information. Taking into account that humour profiles a broad spectrum of information linked to human behaviour (Ruch, 2001), it is coherent to think that there are triggers of affective stimuli which may be identified and learned in order to

[1] Hereafter it must be understood that, when speaking of humour, we refer only to verbal humour, that is, that one expressed by means of linguistic strategies (Attardo, 2001).

provide more elements for characterising humour.

In this framework, the main objective is to study humour beyond only one-liners, focusing on the analysis of a corpus of blogs related to humour in order to study how the bloggers express emotions, sentiments or feelings by means of the information they profile in their posts. This objective implies the following tasks: a) to collect a corpus related to humour; b) to evaluate this corpus; c) to identify and to learn features; d) to assess the relevance of every feature. The first task was accomplished by means of retrieving a corpus from LiveJournal. These data were evaluated twice: firstly, applying the measures proposed in (Pinto et al., 2009) for studying corpora features; the second evaluation was done utilising some of the humour features reported in the literature, especially, we focused on orientation and semantic ambiguity. The third task was performed taking advantage of WordNet-Affect (Strapparava and Valitutti, 2004). Finally, the last task was achieved employing two classifiers implemented in Weka (Witten and Frank, 2005): Naïve Bayes and Support Vector Machine.

3 Experiments

3.1 Data Sets

The corpus was automatically collected from LiveJournal simulating the process described in (Balog et al., 2006), in which the authors took advantage of the predefined tags for analysing irregularities in mood patterns. We enhanced the scope up to considering as well users tags. With respect to the predefined mood tags provided by LiveJournal, there are 132 items organised in 15 categories. We just selected two categories: angry and happy. With respect to the users tags, we just considered the blogs labelled with the humour and joke tags. The retrieval process consisted in requesting to Google and Yahoo search engines, on one hand, all the blogs labelled with the angry and happy mood tags, if and only if, they contained keywords such as punch line, humour, funny, and so on. On the other one, in requesting all the blogs labelled with the users tags: humour and joke.

A set of 7,500 blogs with these parameters were retrieved[2]. They were divided in 3 sets: angry, happy and humour; each one integrated by 2,500 blogs. Besides these sets, we collected one more set from $Wikipedia$ whose main topic was technology. This set also contains 2,500 documents and was used as counterexample.

[2]Available at: http://users.dsic.upv.es/grupos/nle/downloads.html.

Feature	Angry	Happy	Humour	Wikipedia
Terms	1,314.557	1,114.415	1,577.166	1,934.072
CVS	132.831	161.330	219.254	162.305
DL	604.394	542.558	720.496	937.959
VL	411.095	382.987	503.267	516.176
VDR	0.939821	0.944683	0.945468	0.912729
UVB	6.90692	9.27257	9.29029	6.90254
SEM	0.412067	0.404354	0.404779	0.371608

Table 1: Assessment per data set. Measures: corpus vocabulary size (CVS); document and vocabulary length (DL and VL, respectively); vocabulary and document length ratio (VDR); unsupervised vocabulary based measure (UVB); stylometric evaluation measure (SEM).

3.1.1 Corpus Evaluation

In order to provide elements to automatically justify the corpus validity, the data sets were evaluated by means of the criteria described in (Pinto et al., 2009) for the assessment of corpora features. The characteristics analysed[3] were:

i. $shortness$, whose objective is to evaluate the length of a collection considering aspects such as document length, vocabulary length, and document length ratio;

ii. $broadness$, whose objective is to evaluate the domain broadness of a collection on the basis of supervised or unsupervised[4] language modeling based measures;

iii. $stylometry$, whose objective is to give hints about the linguistic style employed for writing a document.

The results obtained are shown in Table 1. According to the values presented in this table, the inferences about the data sets indicate:

i. with respect to the shortness measures, it can be noticed that all the data sets are integrated by large documents and large vocabularies. This impacts on the complexity of every one. The VDR measure indicates that, in terms of frequency, all the sets imply high complexity;

[3]All the measures are implemented in the Watermarking Corpora On-line System (WaCOS), available at: http://users.dsic.upv.es/grupos/nle/demos.html.

[4]Due to the lack of a humour gold standard to compare the data sets with, we always selected the unsupervised version to assess the corpus.

ii. with respect to the broadness, the UVB measure points out that, broadly, all the sets tend to restrict their topics to specific contents, being the happy and humour sets the most limited to particular subjects. That is, they represent two narrow domain collections.

iii. with respect to the stylometry, the SEM measure indicates that, despite the blogs and the documents from Wikipedia are written by several persons, they share a common expression style. This can be perceived by the similarity among the angry, happy and humour sets. According to their SEM values, they show a trend to have specific language style. Considering this information, we think that, at least these 3 sets, have a kind of *identity* tag that supposes a particular pattern.

3.2 Orientation

According to the results depicted in (Mihalcea and Pulman, 2007), humour tends towards a negative orientation. That is, from a sentiment analysis viewpoint, there are more words and/or sentences related to negative connotations in humorous examples than in non humorous ones. In their experiments with one-liners and humorous news articles, the negative polarity has been an important discriminating feature. Therefore, we decided to verify whether or not this feature has the same behaviour over our data sets.

The experiment contemplated two manners of obtaining the orientation. The first way was by means of using a public tool for Sentiment Analysis: Java Associative Nervous Engine (Jane16)[5]. This tool creates a model of positive and negative words and sentences which are crawled in Internet. Depending on their occurrence, they are ranked. The labelling phase matches the information provided by the users with that one in the Jane16 database. For the second one, we employed SentiWordNet (Esuli and Sebastiani, 2006). This resource contains a set of graduated tags to cover the positive and negative polarity for the following categories: nouns, verbs, adjectives and adverbs. We only focused on nouns and adjectives, if and only if, they passed a empirically founded threshold ≥ 375 in the positive or negative scores registered in SentiWordNet.

Considering both resources, we created a dictionary including the positive and negative nouns and adjectives, which was compared against every one of the blogs and documents in the four data sets. The labelling stage computes the amount of positive and negative items for determining the final orientation. The results obtained with both resources are shown in Tables 2 and 3.

Set	Positive	Negative	Neutral
Angry	1,574	548	378
Happy	1,593	363	544
Humour	1,785	336	379
Wikipedia	1,861	147	492

Table 2: Jane16 results

Set	Positive	Negative	Neutral
Angry	2,329	115	56
Happy	2,307	133	60
Humour	2,379	80	41
Wikipedia	2,309	145	46

Table 3: SentiWordNet results

Except the Wikipedia set, the results are contrary to our expectations. The polarity profiled by all the sets trends towards a positive orientation and the difference is significant, as can be noted from the correlated results. This behaviour questions the relation among the global content in the data sets (at least in the angry, happy and humour sets) and humour. Considering that the seeds for retrieving the blogs were selected taking into account keywords related to humour, we would have expected another kind of results. The explanation we could argue to justify this outcome is to point out that the results exposed in (Mihalcea and Pulman, 2007) apply to another kind of data. Moreover, we need to take into account that, although we tried to guide the topics towards humour, the blogs are heterogeneous sites where the humour is not always expressed through a lists of jokes, one-liners, etc., but also by means of images, videos, comments and so on.

3.3 Semantic Ambiguity

In several works related to computational humour it has stressed the importance of ambiguity for generating humorous effects (Mihalcea and Strapparava, 2006; Sjöbergh and Araki, 2007; Reyes et al., 2009). In our case, we aim at analysing the semantic ambiguity applying the techniques ex-

[5]Tool available at http://opusintelligence.com/download.jsp.

Set	$\sum_W$ N	$\overline{X}$	σ
Angry	395,329	10.47	3.12
Happy	380,361	9.92	3.05
Humour	520,520	10.43	3.11
Wikipedia	632,509	9.73	2.07

Table 4: Semantic ambiguity results

posed in (Reyes et al., 2009) for measuring the dispersion degree among the senses of a given noun. The sense dispersion measure intends to quantify the differences among the senses of a word considering the hypernym distance among the WordNet synsets (Miller, 1995). The underlying concept behind the measure relies on the hypothesis about a noun with senses that differ significantly is more likely to be used to trigger humorous effects than a word with senses that differ slightly.

The experiment consisted in retrieving all the nouns from the four sets and applying the formula in (1) for getting the global hypernym distance:

$$\delta(w_s) = \frac{1}{P(|S|, 2)} \sum_{s_i, s_j \in S} d(s_i, s_j) \qquad (1)$$

where S is the set of synsets $(s_1, ..., s_n)$ for the word w; P(n,k) is the number of permutations of n objects in k slots; and $d(s_i, s_j)$ is the length of the hypernym path between synsets (s_i, s_j). The total dispersion per noun was calculated as: $\delta_{TOT} = \sum_{w_s \in W} \delta(w_s)$, where W is the set of nouns in the collection N. All the single values were summed in order to get the sense dispersion in each one of the blogs and set. The results are depicted in Table 4.

The results obtained through the sense dispersion measure suggest that the angry, happy and humour sets are the most ambiguous ones. Taking into account the amount of nouns per set and their standard deviation with respect to their variance, it can be noted how the dispersion average in the Wikipedia set is much smaller than in the other three. Consider also that, with respect to the angry and happy sets, the difference in quantity of nouns is about 30% more items. According to (Reyes et al., 2009), this is a hint about a deeper ambiguity profiled in those sets. Following their hypothesis, this underlying ambiguity can be used for creating humorous situations through words that function as humour triggers.

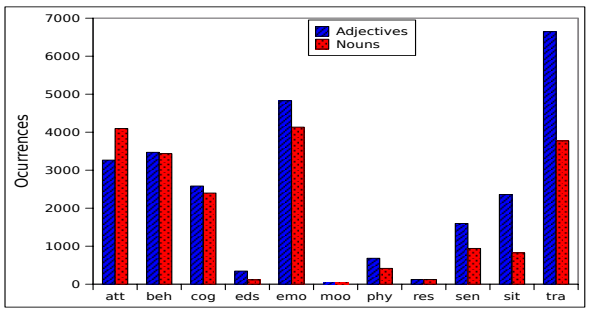

Figure 1: WordNet-Affect categories distribution

3.4 Affectiveness

We always denote affective information through the words we employ in our every day communication. This characteristic is acquiring greater importance in scenarios such as sentiment analysis, computer assisted creativity or verbal expressivity in human computer interaction (Strapparava and Mihalcea, 2008). An example is the SemEval-2007 workshop where, one of the tasks was devoted to analyse the affectiveness in text (Strapparava and Mihalcea, 2007). From a (computational) humour perspective, this task becomes more difficult because humour not only relies on the "funny" utterances produced by a speaker but also on how the hearer codifies that information (Curcó, 1995). Nonetheless the difficulty, we performed an experiment for computing, for each blog and document, the amount of affective nouns and adjectives according to the WordNet-Affect categories. These are: attitude (att), behaviour (beh), cognitive state (cog), edonic signal (eds), emotion (emo), mood (moo), physical state (phy), emotional response (res), sensation (sen), emotion-eliciting situation (sit) and trait (tra)[6]. Figure 1 shows the distribution of every category in terms of occurrences within the sets.

As can be appreciated in the figure, the affectiveness in the data sets is more representative by the adjectives and by the tra, emo, att, beh and cog categories. This implies that the bloggers express their affectiveness by means of qualifying attributes. The next step consisted in verifying what is the most representative category per set considering both morphosyntactic categories. This information is given in Figure 2.

According to the results depicted in Figure 2, it is interesting to note how affective information

[6]In (Strapparava and Valitutti, 2004) it can be found all the information about the concepts represented by these categories.

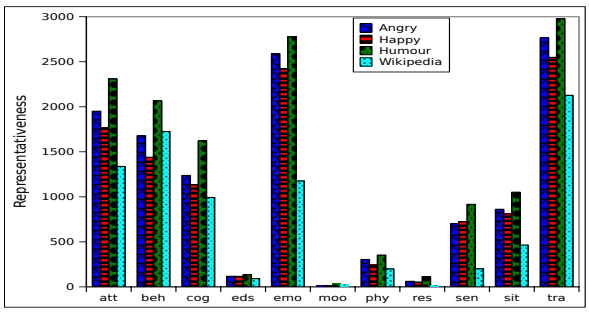

Figure 2: WordNet-Affect representativeness per set considering nouns and adjectives as one

seems to play an important role on the manner of expressing humour (through words that denote emotions, feelings, moods, etc.) by the bloggers. In accordance with this graphic, the humour set profiles a greater trend to express its content using affective features. This could be correlated to our assumption that humour takes advantage of multiple resources and techniques (superiority, incongruity, etc.) to get its effect. Moreover, the behaviour observed by the rest of sets is the expected one. Both angry and happy sets are also sufficiently representative by the affective categories to be distinguishable, at least in the same classes as the humour set, from the Wikipedia one.

4 Evaluation

The classification task described in this section was carried out in order to assess the relevance of the features previously investigated. The idea was to know how much they can help for representing the bloggers expression manner and, especially, to be considered for identifying humour in sources such as blogs.

Six classifications experiments were performed. Every one of the 7,500 blogs and 2,500 documents was represented though a feature vector. The following schema summarises the features and the order in which they were assessed:

i. semantic ambiguity (amb), considering the sense dispersion value organised according to three scales (1 - 10; 11 - 20; 21 - above), being the last value the most ambiguous;

ii. orientation (orien), considering the positive, negative and neutral polarity obtained with Jane16 and SentiWordNet;

iii. ambiguity and orientation (amb+orien), considering both sense dispersion and polarity;

iv. affectiveness (affect), considering the WordNet-Affect categories according to five scales (1 - 100; 101- 200; 201 - 300; 301 - 400; 401 - above), being the last value the one with most affective items;

v. total features (all), considering all the previous attributes together;

vi. informativeness features (infoGain), considering only the subset with most informativeness ratio. This subset was obtained by means of the information gain measure implemented in Weka (Witten and Frank, 2005).

With respect to the classifiers, the task was performed using Naïve Bayes and SVM. The classes considered to evaluate the performance were: angry, happy, humour and Wikipedia. Finally, the method used for evaluation was ten-fold cross validation. The results are depicted in Figure 3.

Despite the classification accuracy is not good, as can be noted in the graphic, the most important conclusion we can draw is to confirm the relevance of affective information for discriminating the data sets according to the emotions, pleasures, displeasures, attitudes, feelings and so on, expressed by the bloggers. This is corroborated by both classifiers. The accuracy reached with Bayes and SVM, considering both semantic ambiguity and orientation, does not achieve 30%, while considering affective information the accuracy increases almost 10%. Likewise, taking into account the features studied and their role in the classification, it is evident that, in order to recognise humour in these sources, it is not enough to consider features such as ambiguity or polarity. It must also be considered information more related to emotional and affective aspects in order to enhance the quantity and quality of variables that impact on humour. This can be observed from the graphic: the accuracy achieved through a selection of the most informative features (among them Jane16 orientation, semantic ambiguity, and six affective categories) produces a better performance, achieving a better accuracy with SVM.

5 Conclusions and Future Work

In this paper we have studied whether or not affective data could be used for humour recognition tasks. The experiments have focused on analysing a corpus of blogs related to humour and moods.

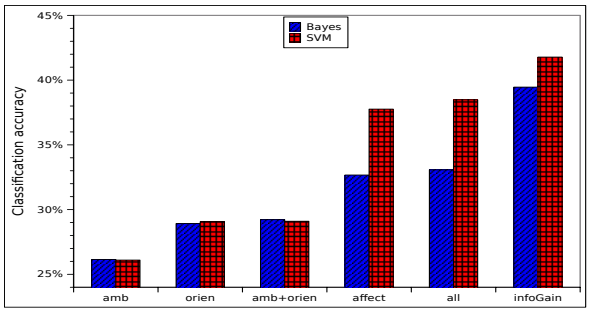

Figure 3: Classification accuracy

Two underlying evaluations were performed: one with respect to the corpus and other with respect to the features relevance. Regarding to the experiments for determining the corpus validity, it is obvious that, although the evaluations show hints about the presence of humour in the data, not all the information is related to humour. Thus, the results must be understood under this perspective. Regarding to the features relevance, the classification task showed that, although affective information did not help so much for classifying the data sets according to their content, it could be useful for characterising humour. Finally, as future work, we will verify the results with more data and contemplating other kind of sources, besides analysing aspects such as irony or sarcasm.

Acknowledgements

The TEXT-ENTERPRISE 2.0 (TIN2009-13391-C04-03) project has partially funded this work.

References

S. Attardo. 2001. *Humorous Texts: A semantic and pragmatic analysis*. Mouton de Gruyter.

K. Balog, G. Mishne, and M. Rijke. 2006. Why are they excited? identifying and explaining spikes in blog mood levels. In *European Chapter of the Association of Computational Linguistics (EACL 2006)*.

C. Curcó. 1995. Some observations on the pragmatics of humorous interpretations: a relevance theoretic approach. In *UCL Working Papers in Linguistics*, number 7 in Working Papers in Linguistics, pages 27–47. UCL.

A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 417–422.

A. Ghose, P. Ipeirotis, and A. Sundararajan. 2007. Opinion mining using econometrics: A case study on reputation systems. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 416–423. Association for Computational Linguistics.

R. Mihalcea and S. Pulman. 2007. Characterizing humour: An exploration of features in humorous texts. In *8th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2007*, volume 4394, pages 337–347.

R. Mihalcea and C. Strapparava. 2006. Learning to Laugh (Automatically): Computational Models for Humor Recognition. *Journal of Computational Intelligence*, 22(2):126–142.

G. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

D. Pinto, P. Rosso, and H. Jimnez. 2009. On the assessment of text corpora. In *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems*.

A. Reyes, D. Buscaldi, and P. Rosso. 2009. The impact of semantic and morphosyntactic ambiguity on automatic humour recognition. In *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems*.

W. Ruch. 2001. The perception of humor. In World Scientific, editor, *Emotions, Qualia, and Consciousness. Proceedings of the International School of Biocybernetics*, pages 410–425.

J. Sjöbergh and K. Araki. 2007. Recognizing humor without recognizing meaning. In *3rd Workshop on Cross Language Information Processing, CLIP-2007, Int. Conf. WILF-2007*, volume 4578, pages 469–476.

C. Strapparava and R. Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007)*.

C. Strapparava and R. Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied Computing*, pages 1556–1560.

C. Strapparava and A. Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 4, pages 1083–1086.

I. Witten and E. Frank. 2005. *Data Mining. Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers. Elsevier.