SINAI at Affect Task in MediaEval 2010 *

José M. Perea-Ortega, Arturo Montejo-Ráez, M. Teresa Martín-Valdivia and L. Alfonso Ureña-López
SINAI Research Group, Computer Science Department
University of Jaén
23071 - Jaén, Spain
{jmperea,amontejo,maite,laurena}@ujaen.es

ABSTRACT

In this paper, a basic approach to establish the *boredom score* of several videos using their speech transcriptions is presented. We carried out an analysis of the development data provided trying to establish a correlation between the terms of the transcriptions and their *boredom scores* predicted for each video. For the development data, we analyzed different weighting schemes such as TF, TF·IDF, binary or correlation weights, due to the small size of the texts involved. We also studied how the use of stopper and stemmer influences the system. Finally, we concluded that apply stopper and stemmer is interesting in this task and TF and correlation weights obtained the best results. Therefore, two experiments were run, using TF and correlation weights as weighting schemes, always preprocessing the transcriptions of the videos. The experiment that used correlation weights obtained the best result for the test videos provided using the *Kendall tau distance* as evaluation metric.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

Keywords

Affective computing, Sentiment Analysis, Opinion Mining, Speech transcriptions

1. MOTIVATION AND RELATED WORK

Year 2001 marked the beginning of widespread of the research problems and opportunities that sentiment analysis and opinion mining raise. Both of them denote the same field of study, which itself can be considered a sub-area of subjectivity analysis [7]. Sentiment Analysis (SA) is a discipline that deals with the quantitative and qualitative analysis of text for determining opinion properties [3]. The term sentiment analysis stands for a broad area of natural language processing, computational linguistics and text mining. It aims to extract attributes and components of the object that have been commented on a document [6]. For example, with rapid expansion of the Web and online merchants, more people buy products on the Web. In order to enhance customer satisfaction, it becomes common for customers to submit and express opinions on the products that they buy. Some products get hundreds of reviews which makes difficult to read them in order to decide which product to choose. From this point of view, an automatic mining opinion system which is able to capture the general perspective and summarize customer viewpoints become a valuable tool. Sentiment analysis classification has several characteristics [1], including various tasks, features, techniques, and application domains. One of them is treated in the Affect task of MediaEval 2010, whose aim is to predict the degree of boredom of several videos, using the speech transcriptions of them.

*This work has been partially supported by a grant from the Spanish Government, project TEXT-COOL 2.0 (TIN2009-13391-C04-02), a grant from the Andalusian Government, project GeOasis (P08-TIC-41999), and two grants from the University of Jaen, project RFC/PP2008/UJA-08-16-14 and project UJA2009/12/14.

MediaEval 2010 Workshop, October 24, 2010, Pisa, Italy

2. DESCRIPTION OF THE TASK

The Affect task in MediaEval 2010 involves automatically predicting the level of user boredom for a video [8]. The aim of this task is the *boredom detection*, i.e. to distinguish when the video content causes the viewer to feel bored and when the video content makes user feel entertained. For the Affect task 2010, the participants can make use of the speech transcriptions, the visual features and the audio content of videos as well as the metadata provided. With respect to the data provided, the video set consisted of short videos from a documentary called "*My Name is Bill*". This documentary is made through the *Bill Bowel's travel* project[1], where each episode tells a story about a place visited during his travel around the world. The videos are about two to five minutes long and chosen to vary along a broad spectrum with respect to their potential to be either boring or entertaining. The videos, the spoken content of the videos transcribed by automated speech recognition, annotations and available metadata including the popularity of the episodes are provided for development purposes. On the one hand, the development data set consisted of 42 episodes that were selected to provide good coverage from videos that viewers report to be the most boring to the most entertaining. On the other hand, the test data set consisted of 82 episodes. Finally, the participants can take two main approaches to establish the *boredom score* for each video: to predict directly the *boredom score* (which is an integer value between 1 and 9, being 1 the most boring score and 9 the least boring

[1] http://www.mynameisbill.com/travelproject/

Run name	*Kendall tau distance*	*Kendall tau correlation*	*Spearman correlation (Rho)*	*Spearman footrule*
SINAI-run1	1683	-0.087	-0.126	29.488
SINAI-run2	1385	-0.018	-0.027	27.098

Table 1: Experiments and results obtained by SINAI in the Affect task

score) or predict the rank position of the video from most boring to least boring, receiving a rank of "1" the most boring video, a rank of "2" the second most boring video, etc. For the experiments carried out in this paper, we directly predicted the *boredom score* for each video.

3. EVALUATION OF RESULTS

Several experiments were carried out in order to analyze the development corpus provided. The aim of this study was to establish the *suitable parameters* for a good correlation between the terms of the transcriptions and the *boredom score* assigned to each development video. Then, these *suitable parameters* were used to predict the *boredom score* for each test video. The devel corpus consisted of 42 documents with 240 words per document on average, and a total number of 1,179 tokens, once stopper and stemmer were applied. The RapidMiner[2] tool was selected to carry out the experiments, using a linear regression model as predictive approach. In addition, the *leave-one-out cross-validation* technique [5] was applied for this model in order to study how accurately will predict the test data. *Leave-one-out cross-validation* is usually very expensive from a computational point of view because of the large number of times the training process is repeated. For this reason, the dimensionality of the preprocessed devel corpus was reduced using PCA[3] with 0.95 as variance threshold (reducing from 1,179 to 32 attributes).

The first experiment was to study different weighting schemes such us TF, TF·IDF, binary or correlation weights. For the correlation weights, a weight is assigned to each term based on the *boredom score* determined for the document of that term. The analysis of the results for this experiment determined that TF·IDF and binary were not interesting due to short length of the transcriptions, obtaining TF and correlation weights the best results. Secondly, we analyzed the performance of applying stopper and stemmer, concluding that the use of both techniques was interesting.

Analyzing the previous results, finally we proposed two main experiments: ***SINAI-run1*** (using stopper and stemmer as preprocessing and TF as weighting scheme) and ***SINAI-run2*** (using stopper and stemmer as preprocessing and correlation weights as weighting scheme). The official results obtained in Affect task of MediaEval 2010 are shown in Table 1. For the evaluation of the results were used four ranking distance metrics: *Kendall tau distance* [4], *Kendall tau correlation*, *Spearman Correlation* (Rho) [2] and *Spearman footrule* distance or mean absolute ranking distance. The main evaluation metric is the *Kendall tau* distance, which counts the number of pairwise disagreements between two lists. The larger the distance, more dissimilar are the two lists. Analyzing the results obtained, we can see that using correlation weights as weighting scheme instead of TF reaches a lower Kendall tau distance (-298 points), so it is more interesting for our system to use correlation as weighting scheme, so each term is assigned a sort of "*dullness*" level.

[2]This toolkit provides several techniques along with other interesting features to process textual documents. It is available from `http://rapid-i.com/`

[3]Principal Component Analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.

4. CONCLUSIONS

We have applied a classical approach to the problem of determining a *boredom score* for video transcriptions. Our main finding is that, due to the relative small size of the transcripted videos (240 words), approaches like TF·IDF does not contribute to relevant weights for the terms. Thus, a correlation approach was devised as alternative, resulting in better results.

5. REFERENCES

[1] Abbasi, A., Chen, H., and Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst. 26*, 3 (2008).

[2] Diaconis, P. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[3] Esuli, A., and Sebastiani, F. Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of the Association for Computational Linguistics (ACL)* (2007).

[4] Kendall, M. *Rank correlation methods*. Griffin, London, 1948.

[5] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (1995), pp. 1137–1145.

[6] Liu, B. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL, 2010.

[7] Pang, B., and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*, 1-2 (2008), 1–135.

[8] Soleymani, M., and Larson, M. Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. In *SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)* (2010).