

Analyzing Affective States using Acoustic and Linguistic Features

Branimir Dropuljić, Sandro Skansi, Robert Kopal

IN2Data Data Science Company,

Marohnićeva 1/1, 10000 Zagreb

{branimir.dropuljic, sandro.skansi, robert.kopal}@in2data.eu

Abstract. *This paper explores the hypothesis that sentiment in text is closely related to emotions in speech in terms of features needed for successful detection. We use a Croatian emotional speech corpus (CrES) and a Croatian social network textual sentiment corpus SentHR. We first perform emotional state estimation based on acoustic speech features using support vector machines in the first case and random forest in second. Accuracy between 60% and 70% was achieved for five discrete emotion classification task. Subsequently, we trained a positive naive Bayes classifier for textual sentiment, reporting an accuracy of around 70% (with a pronounced bias towards the complement). Finally, we used the trained sentiment classifier for two classification experiments on the transcripts of the CrES dataset for classifying anger and sadness. Across several iterations, the results showed that accuracy on the transcripts was around 50% for both sadness and anger, reporting a slightly higher (albeit consistently higher) accuracy on emotional state "anger".*

Keywords. Acoustic speech features, Affective states, Emotional state estimation, Sentiment, Textual sentiment analysis.

1 Introduction

Emotions were a central scientific topic from the earliest times in human history. They were studied within philosophy in a somewhat pejorative light (Adam, 2007), and only in modern times when biology and psychology became separate sciences, the focus shifted towards emotions as a positive human aspect. Darwin (Darwin, 1872) described emotions as a type of intergenerational memory, i.e. reflexes essential for survival. The Darwinian approach dominated until 1950's when the theory of cognitive appraisal emerged (Arnold, 1960) alongside the broader behavioural movement in psychology.

With the advent of cognitive science in the 1990's, a whole range of important aspects of daily life was tied to emotions, and areas such as emotional intelligence

and emotional judgment began to take shape (Damasio, 1994).

Affect is generally considered synonymous with emotions, and the suggestion from the W3C consortium reinforces this claim ("W3C Emotion Incubator Group Report", 2007), (Schroeder et al., 2011). However, in affective computing, there is a difference, where affect (or affective states) is a hypernym for emotions, and additionally includes mood, sentiment and personality traits. This is the approach taken in (Desmet, 2002). The goal of the present paper is to test and challenge these relationships.

The present paper explores two connected phenomena: emotions in acoustics and sentiment in text. Our initial hypothesis was that sentiment in text is closely related to emotions in speech in terms of features needed for successful detection. We used two datasets both in Croatian. The first is Croatian emotional speech corpus (CrES) and it contains audio and textual material annotated with discrete emotions, i.e. happiness, sadness, fear, anger and neutral state, and emotional dimensions: valence and arousal. The second dataset contains textual sentiment annotations (labels: "negative", "other") collected from social networks. We first perform emotional state estimation based on acoustic speech features using support vector machines in the first case and random forest in second. After that, we trained a positive naive Bayes classifier for textual sentiment. Finally, we used the trained sentiment classifier for two classification experiments on the transcripts of the CrES dataset: (i) classifying "sadness" as negative sentiment and "happiness" combined with "neutral state" as "other"; and (ii) classifying "anger" as negative sentiment and "happiness" combined with "neutral state" as "other".

2 Croatian Emotional Speech Corpus

Croatian emotional speech (CrES) corpus was collected and emotionally annotated from various prerecorded sources. The first part called "real-life

emotions” was collected from Internet, mostly from Croatian reality shows and from different documentaries. The second part called “acted emotions” was collected from Croatian movies, TV Shows and Books-Aloud programs. A detailed description of building the initial version of the corpus is presented in (Dropuljić et al., 2011), and an upgraded version, which will be used in this paper, is presented in (Dropuljić et al., 2013). This upgraded version contains total of 1140 utterances from 341 different male and female speakers with the total duration of approximately 85 minutes.

Utterances were categorized into five emotion categories: happiness, sadness, anger, fear and neutral state, based on subjective opinion of ten or more annotators per each utterance. Utterances were also annotated as a continuous levels of valence and arousal. Some utterances were removed during the filtering process in accordance with agreement and prevalence criterion, described in (Dropuljić et al., 2013), and a total of 1007 utterances remained for the analysis.

3 Emotion Analysis Based on Acoustic Speech Features

Measurable relation between emotions and speech was scientifically discovered in 1930-ies. In 1936, Cowan made the first analysis of acoustic features of a human voice recorded during public speeches (Cowan, 1936). Several years later, Fairbanks and Pronovost went a step further by analysing speeches recorded in the expression of a wider spectrum of emotions (Fairbanks & Pronovost, 1939). After revealing its potential, this interdisciplinary topic began expanding circles of interest. Psychologists and linguists were joined by neurologists and more recently, by computer experts, who contribute greatly to this field by developing computer systems for automatic emotion recognition, as well as for analysis and selection of appropriate voice features using statistical methods. Some of the most significant scientific breakthroughs were following works: (Scherer, 1986), (Banse & Scherer, 1996), (Schuller, Rigoll & Lang, 2004), (Lugger & Yang, 2008), (Eyben et al., 2010) and (Wei et al., 2016).

In this paper, speaker-independent estimation of discrete and dimensional emotional states is performed using acoustic speech features extracted from CrES corpus utterances. Support vector machines (SVM) and random forest (RF) are used for this purpose.

3.1 Acoustic Speech Features

A total of 472 acoustic features were considered for estimation of emotional states. One feature vector is calculated for each utterance in the corpus. Relevant acoustic cues from emotionally rich speech

expressions were taken from phonation and articulation speech processes. Features were extracted mostly from speech prosody information, i.e. pitch, energy and duration, and from spectral domain parameters like formants and mel-frequency cepstral coefficients (MFCC). Each parameter was estimated from 25ms speech analysis window, with the frame-rate of 100 fps.

Most parameters and features were calculated directly in MATLAB, while for some of them, integration with other specialized software was necessary. Additionally, all the features were normalized across the utterances, i.e. the value-origin was shifted to the means of the features, and the variance of all features was scaled to 1. Features were categorized in 7 groups as follows.

Raw speech features – Statistical measures like mean value, median, skewness and kurtosis, plus difference between mean and median of absolute value of a signal were extracted. Features were calculated from the whole utterance. Features calculated as mean value and median were included in (Schuller, Rigoll & Lang, 2004).

Speech rate features – Voiced, unvoiced and silence intervals in an utterance were calculated first. Therefore, Voicebox implementation of voice activity detector was used to separate silence from speech intervals (Sohn, Kim, & Sung, 1999), while a pitch estimator was used to distinguish voiced from unvoiced intervals. Features were also calculated from several statistical measures of silence, speech and voiced interval durations. Therefore, speech rate, voice rate and silence rate measures were calculated, as well as more complex measures. Such features were inspired from (Schuller, Rigoll & Lang, 2004) and (Lee & Narayanan, 2005).

Zero-crossing rate features – Statistical measures of ZCR contour were calculated. Additionally, relative positions of minimum and maximum zero-crossing rates in an utterance were included in the feature set.

Short term energy features – Statistical features, as well as few specific features like relative maximum of the short term energy and its position in an utterance and also mean value and standard deviation of distances between inflection energy points (Schuller, Rigoll & Lang, 2004) were calculated. Furthermore, features were calculated from only specific intervals of the STE, like rising and falling slopes, as well as minima and maxima plateaux. Similar features were introduced in (Ververidis & Kotropoulos, 2006).

Fundamental frequency features – Fundamental frequency of the periodic glottal excitation was estimated from the voiced parts of a speech signal using the Voicebox implementation of a robust algorithm for pitch tracking (RAPT) (Talkin, 1995). Statistical features from the voiced parts, as well as from only rising and falling slopes, together with minima/maxima plateaux, were extracted. Furthermore, measures of period-to-period fluctuations in fundamental frequency (jitter) and

period-to-period variability of the amplitude value (shimmer) were calculated. For jitter and shimmer measurements, Praat functions were used. It was shown in (Fuller, Horii & Conner, 1992) that such measures, also applied in (Li et al., 2007), could indicate several mental disorders related to stress. In order to include information about the relation between phonation and articulation speech processes, the harmonic-to-noise ratio (HNR) features were extracted. Noise in HNR is related to the non-periodic part of the voice spectrum. For calculating HNR features, VoiceSauce implementation of Krom's algorithm was used (Krom, 1993).

Spectrum features – Spectral domain features were calculated from: short-term spectrogram, long-term spectrum of the whole utterance, individually averaged short-term spectra of voiced and unvoiced parts of the utterances, and finally from MFCCs. The short-term spectrogram was used for spectral flux computation. Long-term spectrum was used for estimation of energy of several chosen frequency bands, center of gravity, spectral roll-off-point (Schuller, Reite & Rigoll, 2006), etc. Features from voiced and unvoiced short-term spectra were calculated as presented in (Banse & Scherer, 1996). Finally, features from 13 MFCCs (12 plus 0th order coefficient) were calculated in a similar way as in (Lugger & Yang, 2008).

Formant features – Features were analyzed from formant parameters that were computed using the Snack Sound Toolkit. Statistical features, inspired from (Scherer, 1986), were taken from central frequencies and bandwidths of the first four formants.

3.2 Emotional State Estimation

3.2.1 Support Vector Machines

Classification of discrete emotional states: happiness, sadness, anger, fear and neutral state was performed using LIBSVM implementation of the SVM. Following parameters were applied: 10-fold cross-validation (CV) process ($k = 10$) was selected; radial basis function (RBF) was used as a kernel function with γ set to $1/F$ (number of features, $F = 472$); the cost parameter C was set to 1; and threshold ϵ was set to 0.001. Furthermore, the sequential floating forward selection (SFFS) algorithm was used to select 50 most relevant features, with tolerance set to 2 features. Classification accuracy was used as a criterion function and the referent knowledge for the 10-fold CV was defined as described in (Dropuljić et al., 2013). Maximal obtained accuracy for 5 discrete emotions classification was 69.41%, with 40 features selected. Confusion matrix is presented in Table 1.

It can be seen that recognition rate varies across emotions and highest recall, of approximately 80%, was achieved for anger and neutral state. It can be explained with non-uniform distribution of emotions in the Croatian emotional speech corpus.

Table 1. Confusion matrix for discrete emotions classification using SVM

		Predicted Emotion					Tot.
		H	S	F	A	N	
Actual Emotion	H	103	8	11	44	13	179
	S	17	100	16	34	16	183
	F	7	12	90	24	9	142
	A	18	11	12	241	21	303
	N	4	7	2	22	165	200
Tot.		149	138	131	365	224	1007

Note: H = happiness; S = sadness; F = fear; A = anger; N = neutral state.

Estimation of emotion dimensions was also performed, using LIBSVM implementation of support vector regression (SVR) method, with the same parameters as for SVM. The mean squared errors (MSE) were set as a criterion function. The reference values of valence and arousal, i.e. utterance labels, for 10-fold CV were defined as centroids μ_s of the Gaussians, described in (Dropuljić et al., 2013). Minimal MSE for estimation of valence was 2.2497, achieved with 44 selected features, while minimal MSE for arousal was 1.8147, achieved with 51 features. It should be noted that emotion labels of each utterance are continuous variables from intervals of [1:9] for valence and arousal.

3.2.2 Random Forest

A MATLAB implementation of classification and regression RF algorithms were used for discrete and dimensional emotions analysis. In addition, feature importance was calculated. For building and evaluation of all random forests, 500 trees were used.

Classification accuracy for 5 discrete emotions, based on 472 acoustic features, was 61.77%. As a further step, feature importance for discrete emotion classification was calculated and 100 dominant acoustic features were selected. Classification accuracy using only dominant feature set is 61.97% (confusion matrix is given in Table 2). The similar recall variations per emotions can be observed in the case of RF. In this case, highest recall (of approximately 75%) was also achieved for anger and neutral state.

The MSEs of 2.07 and 0.99 were achieved for estimation of valence and arousal respectively, using 472 features. Furthermore, the MSEs of valence and arousal estimations performed using only 100 dominant acoustic features, calculated for each dimension separately, were 1.96 and 0.94.

Table 2. Confusion matrix for discrete emotions classification using RF

		Predicted Emotion					Tot.
		H	S	F	A	N	

Actual Emotion	H	80	12	8	66	13	179
	S	15	94	19	41	14	183
	F	19	19	70	25	9	142
	A	28	16	12	226	21	303
	N	5	8	3	30	154	200
Tot.	147	149	112	388	211	1007	

Note: *H* = happiness; *S* = sadness; *F* = fear; *A* = anger; *N* = neutral state.

Generally we can conclude that, in the case of CrES corpus utterances, SVM outperform RF for discrete emotion classification task, while RF is better for estimation of emotional dimensions valence and arousal. It can be seen for both SVM and RF that estimation of arousal outperforms estimation of valence. These results implicate that acoustic features are more correlated with arousal rather than valence, what is also concluded in the literature: (Douglas-Cowie et al., 2005), (Eyben et. al., 2010).

4 Sentiment and emotions in text

Russell and Norvig (Russell and Norvig 2009, pp. 33-44) provide a distinction between subhuman, human and superhuman performance. In the domain of playing games, and other adversarial contexts with delimited actions and a clear victory criteria. As such, it is easy to identify superhuman performance, and benchmarking does not play a major role. Language use is not adversarial, and the criteria for "victory" are not defined. As such in language processing the role of benchmarks is crucial: it is not possible to use a language better than humans, and the human agreement is the measure of success.

For sentiment analysis, this agreement is 82% (Wilson, Wiebe and Hoffman 2005, p. 3), which means that a sentiment analysis module that reports and accuracy of e.g. 92% is not superior, but actually inferior to the one that reports and accuracy of 80%, since it is overfitted.

In this paper we will extent the methodology for textual sentiment analysis to textual emotion analysis, by extrapolating the methodology used in speech processing.

4.1 Textual sentiment analysis

All of the main approaches such as (Archak, Ghose & Ipeirotis 2007), (Abbasi, Chen & Salem 2008) or (Bethard et al. 2004) utilize machine learning and our approach to both textual sentiment analysis and textual emotion analysis does not differ.

The basic model for textual representation is a bag of words (a JSON-like object, where features are the keys and values are used to represent the word count). Prior to being passed in the bag of words, the words are stemmed and stopwords are removed. For each document a separate bag-of-words is created, and the

most common approach is to pass it to a naïve Bayes classifier. This approach is able to classify the text as "pos" or "neg" immediately.

A second approach is to dispense with a bag of words, and just join all the tokenized documents, and propagate the document label ("pos" or "neg") to all words, and then train a classifier on it. This approach delegates the hand-crafted statistical processing to the classification part: when a given text should be classified, it can be done only by individually classifying the words and calculating the average (with an optional confidence weight). The result is then not a simple "pos" or "neg" label, but a percentage value. This approach offers greater control over the previous one and enabled the introduction of hand-crafted weights for each word. For general sentiment, this is of some practical use, as it enables the factoring of rare but significant "words" such as: :-), !!!!!!!!!!!!!, **hahahahaha**, :@, **FY** (consider the edit-similarity with **FYI**), **BS**, etc. These words in the context of sentiment and a large dataset are "absorbed" in the predictor, but for emotion detection, and more importantly multinomial emotion classification they convey a large informational value which we do not want to leave to the classifier to identify.

The problem is further compounded by the convention of all caps to represent yelling, and most sentiment classifiers during stemming simply cast all the text in lower case.

In the ideal scenario, we would have a large volume of annotated data and no feature engineering but for practical purposes we will use feature engineering since this is a new area of text processing which has not been yet developed enough.

Multinomial emotion classification is both hierarchical in nature and if more than 20 categories are used, or if categories from different hierarchy levels are used, becomes a rare event detection problem. Additionally, the information encoding different emotions in text has a similar structure (consider :D, xD and :@), making it an ideal problem for deep learning. We will not employ deep learning techniques in the present paper, but we will try to structure the problem at hand and provide a solution to the structured problem. Improvements on this result such as employing deep learning techniques will be left as open research questions and it will be addressed in a future paper.

4.2 Using sentiment to assess emotions

The main goal of our paper is to present a textual emotion classifier. Using a Positive Naïve Bayes classifier with a prior probability of 0.5 trained for textual sentiment (reporting a 68% accuracy, 95% precision and 36% recall) we tried to sort out two of the most prominent emotions and identify the level of correspondence. The emotions we isolated are anger and sadness. The emotions dataset is a reduced experimental dataset compiled from transcriptions of

CrES corpus (Dropuljić et al. 2011) (Dropuljić et al. 2013), using only the labels "happiness", "sadness", "anger" and "neutral state". The sorting criteria was to use two experiments: (1) Negative sentiment would be presumed to be equivalent to "anger", whereas the complement would be defined as a shuffled set of "neutral state" and "happiness" and (2) negative sentiment would be presumed equivalent with "sadness", and the complement would be defined as in (1). The results are presented in Table 1.

Table 1. Test results

	Accuracy	Precision	Recall
Sentiment	0.6814	0.9523	0.3636
Anger	0.5134	0.5152	0.0462
Sadness	0.4982	0.5882	0.0546

These results were obtained without any optimization and no features were removed.

5. Conclusion

Emotional state estimation based on acoustic speech features is performed in this paper using support vector machines in the first case and random forest in second. Accuracy between 60 and 70 percent was achieved for five discrete emotion classification task, while mean square errors of between 1 and 2.5 were achieved for valence and arousal estimation task (valence and arousal labels are defined within an interval [1:9]). Additionally, a positive naive Bayes classifier is trained for textual sentiment, reporting the following cross-validation metrics: accuracy of 68%, with precision 95% and recall 36%. Trained sentiment classifier is then used for two classification experiments on the transcripts of the CrES dataset: (i) classifying "sadness" as negative sentiment and "happiness" combined with "neutral state" as "other"; and (ii) classifying "anger" as negative sentiment and "happiness" combined with "neutral state" as "other". Across several iterations, the results showed that accuracy on the transcripts was around 50% for both (i) and (ii), reporting a slightly higher (albeit consistently higher) accuracy on emotional state "anger".

The poor recall results show that the classifier has not been able to ascertain with enough confidence the belonging to "sad" and "angry". The problem is solvable by augmenting the datasets used, and we leave this along with optimization and different classifier and word vectorization choices as open areas for further research.

A second and more interesting result is that we have a significant trend of better classification of "angry" as compared to "sad": across multiple iterations all of them except one have shown greater accuracy on "angry" than on "sad".

It must be noted that our results are to be interpreted as preliminary results. They are not conclusive, but

they do point to a number of factors. First, the experiments should be repeated on a larger dataset with an optimized classifier (by removing low predictivity words), and a deep architecture classifier should be assembled to capture a possibly deeper structure. One possible reason to be investigated is that the statement forms of the transcripts as compared to the social network comments might be intrinsically different, and hence more similar among them (negative comments with positive comments) then with the texts sharing the same polarity across the two forms (e.g. negative social network comments with negative transcripts).

The problem with the high bias will be addressed by enlarging the datasets used for training. Our work focused on annotated corpora for speech and text, but the possibility of additionally annotating CrES for text processing (we used audio-emotional annotations, which do not necessarily reflect the sentiment in the transcript), will be explored in a future paper.

Acknowledgments

This work was supported by the IN2data Data Science Company Ltd.

References

- Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3).
- Adam, C. (2007). *Emotions: from psychological theories to logical formalization and implementation in a BDI agent*. PhD Thesis, INP, Toulouse, France.
- Archak, N., Ghose, A. & Ipeirotis, P. (2007) Show me the money!: deriving the pricing power of product features by mining consumer reviews. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2007)*.
- Arnold, M. B. (1960). *Emotion and personality*. Columbia University Press, New York.
- Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*.
- Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *J. Personality Social Psych.* 70(3), 614–636.
- Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, H., & Jurafsky, D. (2004). Automatic extraction

- of opinion propositions and their holders. in *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Cowan, M. (1936). Pitch and intensity characteristics of stage speech. *Arch. Speech*.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam. Pub. Group.
- Darwin, C. R. (1872). *The expression of emotions in man and animals*. Murray, London.
- Desmet, P. M. A. (2002). *Designing Emotions*. PhD Thesis, TU Delft, Netherlands.
- Douglas-Cowie, E. et al. (2005). Multimodal databases of everyday emotion: Facing up to complexity. In *Proc. Interspeech'05*, (pp. 813–816).
- Dropuljić, B., Thomasz Chmura, M., Kolak, A., & Petrinović, D. (2011). Emotional Speech Corpus of Croatian Language. In *IEEE International Symposium on Image and Signal Processing and Analysis (ISPA 2011)* (pp. 95–100).
- Dropuljić, B., Popović, S., Petrinović, D. & Ćosić, K. (2013). Estimation of Emotional States Enhanced by A Priori Knowledge. In *4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2013)* (pp. 481–486).
- Eyben, F. et al. (2010). On-line Emotion Recognition in a 3-D Activation-Valence-Time Continuum using Acoustic and Linguistic Cues. *Journal on Multimodal User Interfaces*.
- Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech monograph*. 6, 87–104.
- Fuller, B. F., Horii, Y., & Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Research in Nurse & Health*. 15(5), 379–389.
- Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language and Hearing Research*.
- Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Trans. Speech and Audio Processing*. 13(2), 293–303.
- Li, X. et al. (2007) Stress and emotion classification using jitter and shimmer features. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*.
- Lugger, M., & Yang, B. (2008). Psychological motivated multi-stage emotion classification exploiting voice quality features. *Speech Recognition, In-Tech*.
- Russell, S. & Norvig, P. (2009). *Artificial Intelligence: a Modern Approach*. Harlow: Pearsons Education Ltd.
- Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. 99, 143–165.
- Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C., & Zovato, E. (2011). EmotionML – An Upcoming Standard for Representing Emotions and Related States. *Lecture Notes in Computer Science: Affective Computing and Intelligent Interaction*, 6974, 316–325.
- Schuller, B., Rigoll, G., & Lang, M. (2004). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In *IEEE Proceedings of Acoustics, Speech, and Signal Processing (ICASSP 2004)* (pp. 577–580).
- Schuller, B., Reite, S., & Rigoll, G. (2006). Evolutionary Feature Generation in Speech Emotion Recognition. In *Proc. ICME '06*.
- Sohn, J., Kim, N. S., & Sung, W. (1999). A statistical model-based voice activity detection. *IEEE Signal Processing Lett.* 6(1), 1–3.
- Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). *Speech Coding & Synthesis*.
- Ververidis, D., & Kotropoulos, C. (2006). Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections. In *Proc. XIV European Signal Processing Conf.*
- W3C Emotion Incubator Group Report. (2007). Retrieved from <http://www.w3.org/2005/Incubator/emotion/XGR-emotion/>
- Wei, J., Chen, T., Liu, G., & Yang, J. (2016). Higher-order Multivariable Polynomial Regression to Estimate Human Affective States. *Nature Scientific Reports*.
- Wilson, T., Wiebe, J. & Hoffman, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2005)*.