Is deception emotional? An emotion-driven predictive approach

Shahin Amiriparian[1,3], *Jouni Pohjalainen* [1], *Erik Marchi*[1],
Sergey Pugachevskiy[1], *Björn Schuller*[1,2]

[1]Chair of Complex and Intelligent Systems, University of Passau, Germany
[2]Machine Learning Group, Imperial College London, U.K.
[3]Machine Intelligence & Signal Processing Group, Technische Universität München, Germany
shahin.amiriparian@uni-passau.de

Abstract

In this paper, we propose a method for automatically detecting deceptive speech by relying on predicted scores derived from emotion dimensions such as arousal, valence, regulation, and emotion categories. The scores are derived from task-dependent models trained on the GEMEP emotional speech database. Inputs from the INTERSPEECH 2016 Computational Paralinguistics Deception sub-challenge are processed to obtain predictions of emotion attributes and associated scores that are then used as features in detecting deception. We show that using the new emotion-related features, it is possible to improve upon the challenge baseline.

Index Terms: computational paralinguistics, emotion, deception

1. Introduction

Deception is generally defined as "to cause to believe what is false" or "a deliberate attempt to mislead others" [1]. For centuries, practitioners and laypersons [2] have been interested in the question, do people behave in discernibly different ways when they are lying compared with when they are telling the truth? Assuming this to be the case leads to a practical challenge – detecting deception – which is a well-known task for its difficulty also for humans to perform reliably and consistently.

Detecting deception has long been important in the domains of psychology, law enforcement and other government agencies, international business, national security and research. Most scientific works and experimental studies focus on behavioural and visual cues to deception, such as facial expressions [3] or on traditional biometric cues used in polygraphy [4], [5], [6] or on body gestures [7], [8]. An improvement in detecting deception can be made by analysing non-verbal cues like voice, verbal style or facial expressions: During talking, acting and especially while telling a lie, micro-expressions occur involuntary and express concealed emotions [9].

Recent studies focused on computational linguistics by developing intelligent systems in the scope of distinguishing between deceptive and non-deceptive speech using machine learning techniques [10, 11, 12]. However, as in every machine learning approach, the accuracy of these systems relies highly on the quality and quantity of the available data.

In addition to these factors, several studies investigated how emotions are influencing facial and vocal expression in a plethora of domains such as human-human interaction [13], human-robot interaction [14], and human-computer interaction [15, 16]. In particular, Zuckerman et al. proposed that truth tellers show less undifferentiated arousal than liars. Lying is often indicated by increased blinking, greater pupil dilation and/or higher tone of voice [4]. These studies corroborates our assumption that emotions – and in particular arousal – are playing an important rule in a deceptive expression. It has been observed that the performance of emotional lie detection (based on micro-expression training tools or subtle-expression training tools) is higher than that of unemotional lie detection [17]. It is also theoretically and experimentally more accurate and defensible to interpret arousal-related scores as indicative of deceptive speech [18, 19, 20, 21].

This paper describes our contribution to the Deception sub-challenge as part of the INTERSPEECH 2016 Computational Paralinguistics Challenge (ComParE) [22]. Our approach relies on predicted emotion-related attributes, such as arousal, valence, regulation, and emotion categories. These high-level features are then used as new derived attribute vector for detecting deception. If our assumption is correct, that emotional cues are highly correlated to deception, then a system relying of purely emotion-derived features can be implemented to detect deception. Furthermore, it has to be noted that labelled emotional speech databases are comparably widely available than databases containing deceptive speech.

The following paper is structured as follows. Firstly, the proposed system is introduced in Section 2. Then, Section 3 demonstrates the experimental set-up followed by extensive evaluations. Finally, conclusions and future work are drawn in Section 4.

2. Proposed System

An overview of the proposed system is depicted in Fig. 1. It consists of two main functional components: 1) feature generation which involves classifiers for the emotional attributes arousal, valence, regulation and emotion classes, producing emotion-derived features and 2) the main binary classifier which uses the emotion-related features and is trained to discriminate deceptive vs. non-deceptive speech.

In the present work, feature generation is implemented by training a set of k-Nearest-Neighbour (kNN) classifiers each operating on a specific task, namely arousal, valence, regulation and emotion classes. Each of the four predictions is associated with two relevance values. Feature vectors comprising (a subset of) these 12 features are used in the main classification stage both for the training and detection phase. For the main classification system, we apply both kNN and support vector machines (SVMs). In the following sections, these components are described in more detail.

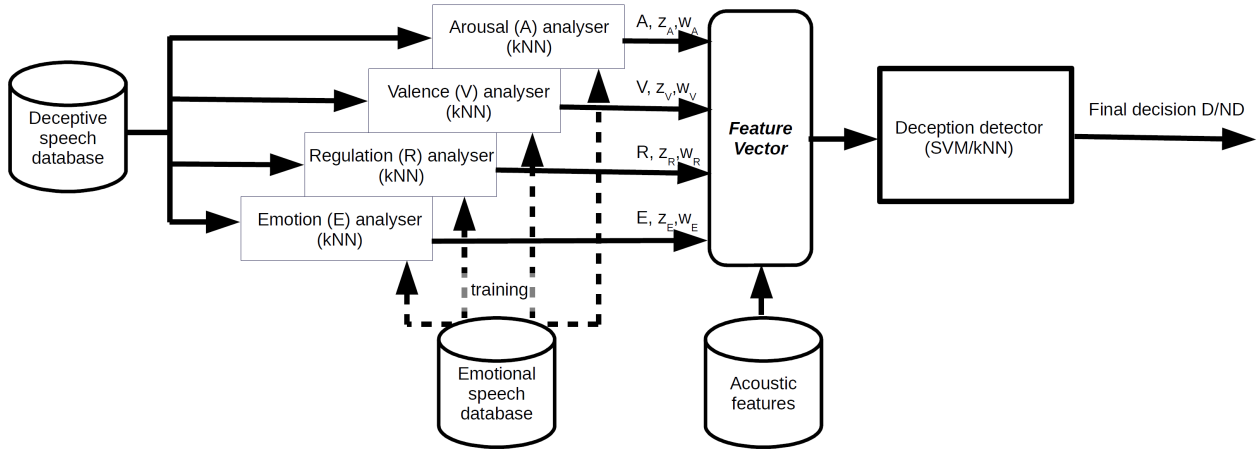

Figure 1: Flowchart of the deception recognising system: using acoustic and emotion-related features for detecting deceptive speech.

2.1. Emotion-driven features

As depicted in Fig. 1, each of the four emotion tasks – arousal (A), valence (V), regulation (R) and emotion (E) – is predicted using a k-nearest-neighbours (kNN) classifier trained on the Geneva Multimodal Emotion Portrayals (GEMEP) emotional speech database. These classifiers are built using the INTERSPEECH 2013 ComParE features set [23], which is the same set of features provided for the Deception sub-challenge by the organisers. However, we also apply feature selection based on mutual information in order to find an optimised and reduced feature subset for each task. In addition to the predicted discrete-valued labels, each emotion attribute analyser outputs two continuous-valued scores: z_X, $X \in \{A, V, R, E\}$, the average distance of each data point to the predicted (majority) class among its k nearest neighbours, and w_X, $X \in \{A, V, R, E\}$, the average distance to all the k nearest neighbours (in the task-specific feature space).

The feature selection method chosen for each of the four emotional attributes is to rank the features based on their mutual information (MI) with the attribute label [24]. This is a simple approach which does not consider feature interdependencies and is thus not designed to obtain compact feature sets, unlike for example the popular, similarly mutual-information-based, minimum-redundance-maximal-relevance (MRMR) method [25]. Nevertheless, the MI feature scoring method, when combined with a reasonable way of deciding upon the number of features to select, has shown good performance in similar paralinguistic tasks, outperforming typical approaches such as MRMR and sequential forward selection [24]. The features are ranked in descending order of the mutual information

$$MI = \sum_{y \in Y} \sum_{z \in Z} p(y, z) \log \left(\frac{p(y, z)}{p(y)p(z)} \right), \qquad (1)$$

where Y is the set of discrete values of a quantised feature and Z is the set of class labels. In previous work, the feature-specific quantisation scale to produce Y has been adaptively determined in such a manner that each quantisation bin contains roughly the same number of samples over the data set under study [24]. In the MRMR method, a three-level quantisation scale with limits at one standard deviation on either side of the mean value is used [25]. In the present work, we have experimented with the aforementioned sample-count-equalising quantisation approach [24], which has a fixed number of quantisation levels, but have obtained better results with a one-dimensional clustering approach that automatically chooses the number of clusters/quantisation levels. It increases the number of clusters one by one and stops at the first point where the rate of decrease of total squared quantisation error, obtained while increasing the number of quantisation levels, has started to diminish noticeably. The goal here is to locate the first obvious "turning point" after which adding more clusters does not improve the modeling of the data as much any more [26]. This is illustrated in Fig. 2.

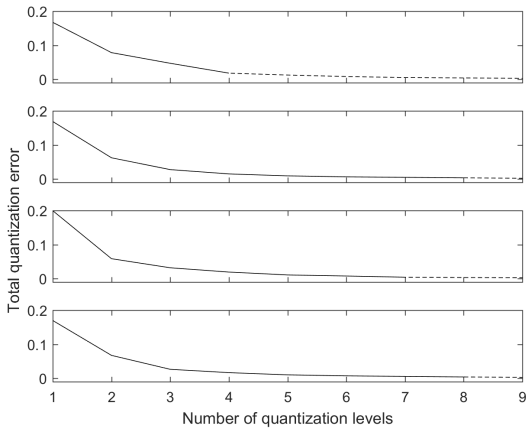

Figure 2: Behaviour of the total quantisation error as a function of quantisation levels for four features when constructing the feature-specific quantisation scale using k-means clustering. The end point of the solid line indicates the number of quantisation levels chosen by the present method, which looks for a point where the relative decrease rate of the quantisation error has stabilised.

The underlying motivation is that the method should explicitly favour such features, whose observed samples form distinct clusters that are in agreement with the labeling – assuming that such features do exist in the base feature set. This add an unsu-

pervised learning aspect to the feature selection approach. The maximum number of quantisation levels $|Y|$ considered in this process has been experimentally chosen as 10 based on the Development set performance of the deception detection system.

For each task, feature ranking by mutual information with respect to the labelling is performed on the train and development sets of GEMEP. The number of ranked features to use for classification is then determined by classifying the Test set using kNN via a grid search with k ranging from 1 to 562 (the number of samples in the Train subset) and with the number of features $d \in \{50, 100, \ldots, 5000\}$. The optimal number of features and optimal k are chosen after median filtering the result grid of Test set unweighted average recall (UAR) separately in both k and d dimensions and taking the minimum of the two. To obtain the final classifier for new data, we concatenate the Train, Development and Test sets of GEMEP and scale the optimal k up in proportion to the increased training data size.

In kNN classification, the hypothesised class label for each test instance is determined as the label seen most frequently among the k labeled training instances closest to the sample in terms of the Euclidean distance [27]. Despite its simplicity, kNN is a powerful pattern classification method that, given enough training data, can model complex nonlinear decision boundaries in the feature space [26]. It also lends itself well to generating nearest-neighbour-based relevance values for each class decision, as described earlier. However, kNN is susceptible to the effects of the curse of dimensionality [26, 27] and thus requires relatively high-quality features to give good results, which is also why we have focused on improving feature selection for the emotion recognition tasks.

2.2. Deception detector

As final stage classifier for the detection of deception, we apply both kNN and linear SVM. Emotion-derived features are first normalised to zero mean and unit variance based on statistics of the training set. In deception detection, kNN is applied in the same form as in generating the emotion features. However, due to the aforementioned curse of dimensionality problem of kNN, we limit the kNN classification experiments to low-dimensional feature sets consisting only of the newly generated features.

Adopting the Weka toolkit [28], SVMs with linear kernel were trained with the Sequential Minimal Optimization (SMO) algorithm. SVMs have been chosen as classifier since they are a well known standard method for emotion recognition due to their capability to handle high and low dimensional data.

3. Experiments

3.1. Material and test setup

Since all data sets are unbalanced (i.e. one class is underrepresented in the data), the unweighted average recall (UAR) of the classes is used as the scoring measure. The SVM training has been performed at different complexity constant values $C \in \{0.004, 0.005, 30, 80, 0.2, 9, 3, 90\}$.

According to the guidelines of the INTERSPEECH 2016 ComParE Deception sub-challenge, we apply the DECEPTIVE SPEECH DATABASE (DSD) created at the University of Arizona which has been divided by the challenge organisers into a Train, Development and Test set.

Firstly, we extract the 12 emotion-based features for all of the material, based on the 6373 features used by the baseline system. Then, our approach is to rely only on these extracted features. Firstly, we train our models with emotion features extracted from the Train set and classify the Development set. Using this setup, we tune the classifier parameters (complexity for linear-kernel SVM and number of neighbours for kNN). Next, we retrain the most promising model configurations with the same emotion features from combined Train and Development set and aim to classify the Test data. As before, the performance measure under study will be the UAR.

3.2. Results

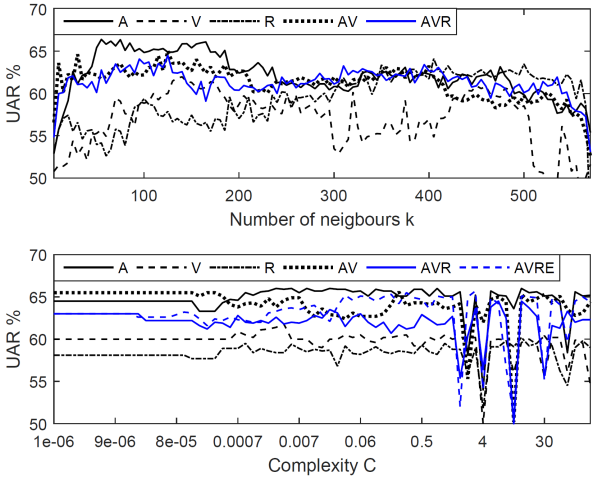

Figure 3: Unweighted average recall (UAR) for small subsets of emotional features arousal (A), valence (V), regulation (R) and emotion category (E) using two classifiers, kNN (upper panel) and SVM (lower panel). Each of the four features involves a predicted categorical label and two continuous-valued scores reflecting the prediction confidence. The classifiers are evaluated by training on the Train subset and predicting the Development subset, using a wide range of the relevant tuning parameter (number of neighbours k for KNN and complexity C for SVM). Results are shown for values of k that are multiples of 5 and for the same values of C as used in [22].

Fig. 3 shows the behavior of both classifiers for selected feature sets when the main parameter (the number of neighbours k for kNN, the complexity C for SVM) is varied. Some observations can be made. Among arousal, valence and regulation, arousal appears to be the strongest individual feature for identifying deception. With several small subsets of emotional features (that include arousal), good deception detection performance, clearly over the Development set baseline, can be achieved in a stable manner over large intervals of the tuning parameter. We can also observe that the quality of the features is not classifier-dependent.

Table 1 shows the results on the Development set with both classifiers using various subsets of the 12 emotion-related features ($A, z_A, w_A, V, z_V, w_V, R, z_R, w_R, E, z_E, w_E$) by themselves, and then the results of SVM classifier using fusion of the 6373 baseline features with the emotion features. From the former approach, it is seen that the Development set baseline can be exceeded using the emotion features only. The latter approach leads to improved performance on the Test set, even though one might expect the large number of the baseline acoustic features to dominate the decision. This finding further suggests that the proposed emotion features have a high discriminative power in the deception classification task.

Table 1: *Deception classification performance on the Development set (UAR %). The emotion feature combinations are denoted by the shorthand notation where, for example, A' indicates the pair of features (A, z_A) and A" indicates the feature triplet (A, z_A, w_A). '6373' indicates the complete baseline openSMILE (OS) feature set of 6373 features.*

Classifier and features		Chosen hyperparameter	Development maximum	Test trials
kNN	A'	k=185	66.9	N/A
	A'R'	k=273	65.5	N/A
	A"V"	k=119	65.1	N/A
	A"V"R"	k=97	64.8	N/A
SVM	A"	C=0.004	**66.0**	N/A
	V"	C=0.005	61.8	N/A
	R"	C=30	60.4	N/A
	E"	C=80	59.1	N/A
	V"E"	C=9	61.3	N/A
	A"V"	C=0.2	65.6	N/A
	A"E"	C=3	**66.5**	N/A
	A"V"E"	C=0.9	65.7	N/A
	V"R"E"	C=40	62.7	N/A
	A"V"R"	C=9	64.5	N/A
	A"V"R"E"	C=3	65.7	N/A
	A"R"E"	C=90	**67.7**	N/A
Emotion + OS				
SVM	A" + 6373	C=10^{-4}	65.2	68.8
	V" + 6373	C=10^{-4}	66.1	**68.9**
	E" + 6373	C=10^{-4}	62.1	**68.9**
Baseline				
SVM	6373	C=10^{-4}	61.9	68.3

4. Conclusions

Application of emotion-related features in detecting deceptive speech was studied. We trained classifiers for categorical emotion attributes on an emotional speech database and applied those on the deceptive speech data in order to generate emotion-related features. We then used these in the deception classification task in combination with different pattern classification methods.

We showed that the emotional features have a relatively high predictive power in the deception task even when used by themselves. Remarkably, by means of fusion of the challenge baseline feature set (6373 features) with a small number of automatically generated descriptors related to, e.g., arousal, valence or emotion, we managed to exceed the Test set baseline of the challenge. These findings imply that emotional attributes, even ones generated by machine learning systems trained on separate data, have considerable potential for detecting deceptive speech.

The approach chosen in this study for utilising emotion analysis resources – especially labeled emotion databases – for detecting deception was to try to produce high-quality features which contain information on deception, as well. Our results show that this approach holds potential for future development of deception detection systems.

5. Acknowledgements

The research leading to these results has received funding from the ECs Seventh Framework Programme through the ERC Starting Grant No. 338164 (iHEARu).

6. References

[1] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception." *Psychological bulletin*, vol. 129, no. 1, p. 74, 2003.

[2] P. V. Trovillo, "A history of lie detection," *Journal of Criminal Law and Criminology (1931-1951)*, vol. 29, no. 6, pp. 848–881, 1939.

[3] P. Ekman and W. V. Friesen, "Detecting deception from the body or face." *Journal of Personality and Social Psychology*, vol. 29, no. 3, p. 288, 1974.

[4] M. Zuckerman, B. M. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception," *Advances in experimental social psychology*, vol. 14, no. 1, p. 59, 1981.

[5] F. S. Horvath, "Verbal and nonverbal clues to truth and deception during polygraph examinations." *Journal of Police Science & Administration*, 1973.

[6] P. A. Granhag, A. Vrij, and B. Verschuere, *Detecting Deception: Current Challenges and Cognitive Approaches*. John Wiley & Sons, 2015.

[7] J. K. Burgoon, D. B. Buller, A. S. Ebesu, and P. Rockwell, "Interpersonal deception: V. accuracy in deception detection," *Communications Monographs*, vol. 61, no. 4, pp. 303–325, 1994.

[8] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *Affective Computing, IEEE Transactions on*, vol. 4, no. 1, pp. 15–33, 2013.

[9] J. P. Gaspar and M. E. Schweitzer, "The emotion deception model: a review of deception in negotiation and the role of emotion in deception," *Negotiation and Conflict Management Research*, vol. 6, no. 3, pp. 160–179, 2013.

[10] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis *et al.*, "Distinguishing deceptive from non-deceptive speech," 2005.

[11] S. Benus, F. Enos, J. B. Hirschberg, and E. Shriberg, "Pauses in deceptive speech." Proc. ISCA 3rd International Conference on Speech Prosody, 2006.

[12] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, and S. Kajarekar, "Combining prosodic lexical and cepstral systems for deceptive speech detection," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 1. IEEE, 2006, pp. I–I.

[13] V. Petrushin, "Emotion in speech: Recognition and application to call centers," in *Proc. of Artificial Neural Networks in Engineering*, vol. 710, St. Louis, MO, 1999, pp. 7–10.

[14] E. Marchi, F. Ringeval, and B. Schuller, "Voice-enabled assistive robots for handling autism spectrum conditions: An examination of the role of prosody," in *Speech and Automata in the Health Care*, A. Neustein, Ed. Walter de Gruyter GmbH & Co KG, 2014, pp. 207–236.

[15] B. Schuller, E. Marchi, S. Baron-Cohen, A. Lassalle, H. O'Reilly *et al.*, "Recent developments and results of asc-inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions," in *Proc. of IDGEI*, Atlanta, GA, 2015, no pagination.

[16] E. Marchi, B. Schuller, S. Baron-Cohen, A. Lassalle *et al.*, "Voice Emotion Games: Language and Emotion in the Voice of Children with Autism Spectrum Condition," in *Proc. of IDGEI*, ACM. Atlanta, GA: ACM, March 2015, 9 pages.

[17] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[18] G. G. Sparks and J. O. Greene, "On the validity of nonverbal indicators as measures of physiological arousal: A response to burgoon, kelley, newton, and keeley-dyreson." 1992.

[19] R. Neiss, "Reconceptualizing arousal: psychobiological states in motor performance." *Psychological bulletin*, vol. 103, no. 3, p. 345, 1988.

[20] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–1210, 1983.

[21] M. A. DeTurck and G. R. Miller, "Deception and arousal: Isolating the behavioral correlates of deception." *Human Communication Research*, 1985.

[22] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language," 2016.

[23] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism," in *Proceedings INTERSPEECH 2013*. Lyon, France: ISCA, 2013, pp. 148–152.

[24] J. Pohjalainen, O. Räsänen, and S. Kadioglu, "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits," *Computer Speech and Language*, vol. 29, no. 1, pp. 145–171, January 2015.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, August 2005.

[26] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 2nd ed. Academic Press, 2003.

[27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons Inc., 2001.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656278