A Novel Real-Time Emotion Detection System for Advanced Driver Assistance Systems

Fadi Al Machot, Ahmad Haj Mosa, Alireza Fasih, Christopher Schwarzlmüller, Mouhanndad Ali and Kyandoghere Kyamakya

Abstract This paper presents a real-time emotion recognition concept of voice streams. A comprehensive solution based on Bayesian Quadratic Discriminate Classifier(QDC) is developed. The developed system supports Advanced Driver Assistance Systems (ADAS) to detect the mood of the driver based on the fact that aggressive behavior on road leads to traffic accidents. We use only 12 features to classify between 5 different classes of emotions. We illustrate that the extracted emotion features are highly overlapped and how each emotion class is effecting the recognition ratio. Finally, we show that the Bayesian Quadratic Discriminate Classifier is an appropriate solution for emotion detection systems, where a real-time detection is deeply needed with a low number of features.

1 Introduction

Every minute, on average, at least one person dies in vehicle crash [9]. For this reason, different approaches to increase the safety on road have been developed. These methods are generally called Advanced Driver Assistance Systems (ADAS) and supports the driver in its driving process. Due to the fact, that every reaction on road is time crucial, ADAS must propose real-time processing. This is the major requirement for such systems.

Typical representatives of ADAS are e.g. Adaptive Cruise Control (ACC) and Lane Departure Warning Systems (LDWS) which are systems to increase safety. Both systems can intervene in the driving process, to avoid hazardous situations for the driver. Other ADAS systems monitor drivers fatigue by analyzing its facial features (eyes), the inclination of its head or the characteristic of its voice [8].

Alpen-Adria-University Klagenfurt, Institute of Smart System Technologies, Transportation Informatics Group, 9020 Klagenfurt, Universitätsstrasse 65-67, Austria, e-mail: forename.surname@uni-klu.ac.at

The system is able to determine the reaction time of the driver and can warn him. In general, an audio signal is used for warning.

We developed a ”Driver Fatigue Warning System” which is based on voice analysis with the purpose of preventing a driver from crashing. The idea is to recognize sadness, angriness and normal mood of the driver. The motivation for mood detection is based on the fact, that aggressive behavior on road leads also to traffic accidents. In that aggressive mood, driver's pitch and volume of his/her voice increase. This change is observed and the system can respond adequately. This system supports other ADAS for monitoring and hence, increases the total safety of on road.

In the recent results of speech emotion recognition systems, researchers use a lot of features. Yacoub *et al.* use 37 features of the voice streams,Wu *et al.* use 52 features and Jian *et al.* use 32 features.

In this paper we use only 12 features to classify between 5 types of emotions based on Bayesian Quadratic Discriminate Classifier and we obtain a very high performance compared to the other existing works.

2 Related Works

Most of the previous works on emotion detection by analyzing audio streams are based on supervised learning by using different emotion classes and apply the standard pattern recognition procedure to train a classification model.

The state-of-the-art of emotion detection is divided into two branches; the first branch is the detection of emotion in music and the second one is the detection of emotions in voice streams.

In the branch of emotion detection in music, several emotion detection methods have been published, for example, Thayer maps the emotion classes on four quadrants in Thayer's arousal-valance emotion plane. He suggests a two dimensional emotion model that is simple but powerful. (In organizing different emotions response: stress and energy). The dimension of stress is called valance, while the dimension of energy is called arousal [4].

Yang *et al.* (2006), apply fuzzy classifiers, which assign a fuzzy vector for a song to indicate the relative strength of each class. The proposed system can be divided into two parts: the Model Generator (MG) and the Emotion Classifier (EC). The MG generates a model according to the features of the training samples, while the EC applies the resulting model to classify the input samples [1].

Synak *et al.* (2005), developes a multi-label classifiers which is able to detect more than one emotion (class) to the same song. Principally, the ideas of [1] and [5] are to provide emotion intensity measurement for each emotion class. They focus on automatic detection of emotion by analyzing audio streams, using features on spectral contents. The data set consists of a few hundred music pieces. The emotion are grouped into 6 or 13 classes [5].

Yang *et al.* (2007), formulates Music Emotion Recognition (MER) as a regression problem and Support Vector Regression (SVR) is applied to predict the arousal and valence (AV) values. With this regression approach, the problems inherent to categorical approaches, are avoided. For instance, besides the quadrant to which the song belongs, one can further calculate, the emotion intensity the song expresses, by examining its arousal and valence (AV) values [3].

Han *et al.* (2009), their recognition system consists of three steps, the first step is the extraction of seven main features from music pieces, the second step is the mapping into eleven emotion categories on Thayers's two-dimensional emotion model. Finally, two regression functions are trained by using Support Vector Regression (SVR), followed by the predicting of arousal and valance values.

In the branch of emotion recognition of a human voice, a few research based on rough Set theory is done. Zhou *et al.*, use an approach based on rough Set theory and SVM for speech emotion recognition. The experiment results show that this approach can reduce the calculation cost while keeping high recognition rate [6].

Yacoub *et al.*, they distinguish between different classes of emotions e.g. sadness, boredom, happiness, and cold anger. They compare results from using neural networks, Support Vector Machines (SVM), K-Nearest Neighbors, and decision trees [7].

The approach of Nwe *et al.* (2003) has several classes of emotions namely, the archetypal emotions of anger, disgust, fear, joy, sadness and surprise. They create a data base of 60 emotional utterances, which are used to train and the proposed system by using Hidden Markov Models. They compare the Low-Frequency-Coefficient (LFPC) features with feature of the Linear Prediction Cepstral Coefficients (LPCC) and Mel-Frequency Cepstral Coefficients (MFCC) features [10].

In this paper, we show that the usage of Bayesian Quadratic Discriminate Classifier (QDC) enables a real-time processing with a low amount of features (12 features). We are able to reduce the calculation cost while keeping high recognition rate.

3 Overall architecture of the emotion detection system

In this section, we present the overall architecture of the emotion detection system (see Figure 1). As training data, the Berlin emotional speech database is used to classify discrete emotions. This publicly available database is one of the most popular databases used for emotion recognition, thus facilitating comparisons with other works. Ten actors $5m/5f$) each uttered 10 everyday sentences (five short and five long, typically between 1.5 and 4 s) in German; sentences that can be interpreted in all of seven emotions acted. The raw database (prior to screening) has approximately sentences and is further evaluated by a subjective perception test with 20 listeners. Utterances scoring higher than 80% emotion recognition rate and considered natural by more than 60% listeners are included in the final database.

In total, we extract 12 features from each sample (see Table 1):

- The minimum, the maximum, the mean and the median of the energy.
- The minimum, the maximum, the mean and the median of the pitch of the signal.
- The minimum, the maximum, the mean and the median of the Mel-frequency cepstrum (MFCC) of the signal.

To extract these features, we use the statistical moments (minimum, maximum, mean and median) of 3 features (MFCC, Pitch and energy) and then for the classification, we use a quadratic discriminate classifier (QDC).

Features
$ENERGY - MAX$
$ENERGY - MIN$
$ENERGY - MEAN$
$ENERGY - MEDIAN$
$PITCH - MAX$
$PITCH - MIN$
$PITCH - MEAN$
$PITCH - MEDIAN$
$MFCC - MAX$
$MFCC - MIN$
$MFCC - MEAN$
$MFCC - MEDIAN$

Table 1 The extracted Features

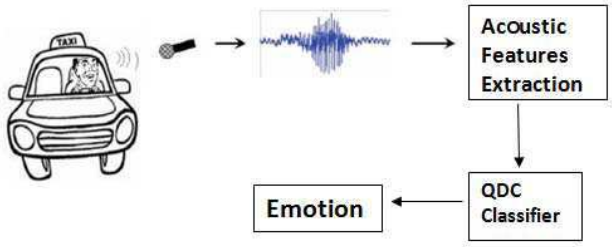


Fig. 1 The overall architecture of the emotion detection system

3.1 The mel-frequency cepstral coefficients (MFCCs)

MFCC features: the mel-frequency cepstral coefficients (MFCCs), firstintroduced in (Davis and Mermelstein, 1980) and successfully applied to automatic speech recognition, are popular short-term spectral features used for emotion recognition[13].

3.2 The Pitch of a signal

Fundamentally, this algorithm exploits the fact that a periodic signal, even if it is not a pure sine wave, will be similar from one period to the next. This is true even if the amplitude of the signal is changing in time, provided those changes do not occur too quickly. A pitch detector is basically an algorithm which determines the fundamental period of an input speech signal. Pitch detection algorithms can be divided into two groups: time-domain pitch detectors and frequency domain pitch detectors [14].

3.3 The energy of a signal

The size of a signal is very important for different applications. We define the signal energy as the is the area under the squared signal [12]:

$$E_f = \int_{-\infty}^{\infty} |f(t)^2| dt \tag{1}$$

Class	Class symbol
Afraid	1
Normal	2
Angry	3
Sad	4
Happy	5

Table 2 The classes of the decision table

4 Support vector machines SVM

In this section, most of results of the state-of-the-art are presented using Support Vector Machine (SVM). Support vector machines (SVMs) (Vapnik, 1995) are used for recognition of both discrete and continuous emotions.While support vector classification finds the separation hyperplane that maximizes the margin between two classes, support vector regression determines the regression hyperplane that approximates most data points with precision. The SVM implementation in (Chang and Lin, 2009) is adopted with the radial basis function (RBF) kernel employed. The design parameters of SVM are selected using training data via a grid search on a base logarithmic scale. In general, the RBF kernel can be a good choice as justified in because:

- It can model the non-linear relation between attributes and target values well.
- The linear kernel is a special case of RBF kernel.
- It has less hyperparameters than the polynomial kernel.
- It has less numerical difficulties compared to polynomial and sigmoid kernels [11].

5 Bayesian Quadratic Discriminate Classifier(QDC)

Bayesian Quadratic Discriminate classifier or QDC is based on the probability distribution of features vector in each class. To estimate the discriminate function between classes we have to first estimate the PDF of each class . For example, let us see the case of our three classes ω_1, ω_2 and ω_3 the first required statistical quantity is the prior probability of each class denoted by $P(\omega_1)$, $P(\omega_2)$ and $P(\omega_3)$ which can be obtained by the training data as [15]:

$$P(\omega_1) = \frac{N_1}{N} \tag{2}$$

$$P(\omega_2) = \frac{N_2}{N} \tag{3}$$

$$P(\omega_3) = \frac{N_3}{N} \tag{4}$$

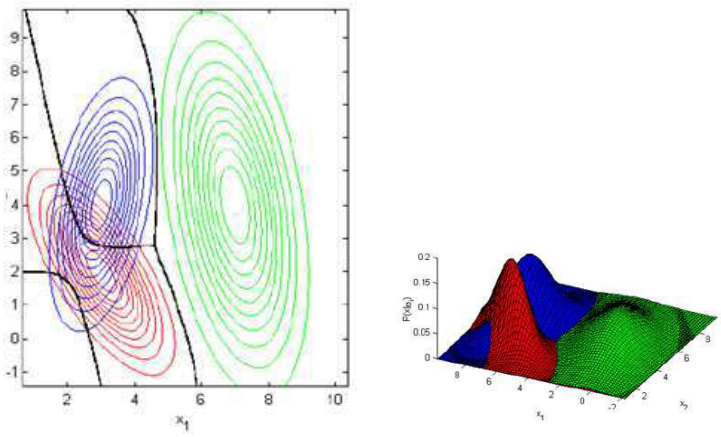
Where N is the total number of the available training data , and N_1,N_2 and N_3 are the number of the features which belong to ω_1, ω_2 and ω_3. Another required statistical quantity is the class conditional PDF $P(x|\omega_1)$, $P(x|\omega_2)$ and $P(x|\omega_3)$ or the likelihood function, commonly Gaussian PDF is used as [15]:

$$p(x|\omega_i) = \frac{1}{2\pi^{n/2}|\sum_i|^{1/2}} exp[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)] \tag{5}$$

6 Maximum likelihood Parameter Estimation

ML method is used to estimate the unknown probability distribution function, for instance suppose $P(x|\omega_1;\theta)$ is the likelihood function with unknown parameter (θ), the ML method estimates the unknown parameter so that the ML function become maximum [15]. Suppose the following function is the log-likelihood function:

$$L(\theta) = lnP(x|\omega_1;\theta) \tag{6}$$



(a) The decision boundaries 2D (b) The decision boundaries 3D

Fig. 2 The decision boundaries are hyper ellipses or hyper-paraboloids (quadratic) in 2 dimensions and three dimensions

So we take the first derivative with respect to The maximum ML of θ is related to zero value of the first derivative:

$$\frac{d(L(\theta)}{d(\theta)} = 0 \geq MaxL \tag{7}$$

In our case the mean μ_i and the covariance matrix $\sum i$ are the unknown parameters for the class conditional PDF, by using ML we estimate μ_i and $\sum i$ for each class as:

$$\mu_{iML} = \frac{1}{N}\sum_{k=1}^{N} x_{ik} \tag{8}$$

Quadratic discernment function:

$$\Sigma_{ijML} = \frac{1}{N}\sum_{k=1}^{N}(x_{ik} - \mu_i)(x_{jk} - \mu_j) \tag{9}$$

Let $g_1(x), g_2(x)$ be the cost function of classes ω_1, ω_2 so x is classified to ω_1 if :

$$g_1(x) > g_2(x) \tag{10}$$

The decision surface which separates the two regions is:

$$g_{12} \equiv g_1(x) - g_2(x) = 0 \tag{11}$$

In our case the cost function:

$$g(x) = -\frac{1}{2}(x-\mu_i)^T \sum_i^{-1}(x-\mu_i) - \frac{1}{2}log(|\sum_i|) + log(P(\omega_i)) \quad (12)$$

The decision boundaries are hyper-ellipses or hyper-paraboloids (quadratic) as shown in 2(a) and 2(b).

7 Results

In the recent results of speech emotion recognition systems, researchers in [6] use 37 features of the voice streams. They classify 6 types of emotions, their total accuracy was 74% based on a combination of support vector machine (SVM) and Rough Set theory and 77,91% based on only SVM. The recognition rates of normal emotion is 90,50%, anger emotion is 86% and sadness is 66%.

In [11], the number of features is between (30-52), using Berlin database features of the voice streams. They classify 7 types of emotions, their total accuracy was 91.6% based on a Speaker Normalization (SN) and Linear Discriminant Analysis (LDA). The recognition rates of normal emotion is 77%, anger emotion is 82% and sadness is 92%.

In [7], authors use 4 statistical moments (mean, maximum, minimum and standard deviation) of 13 features. They classify 4 types of emotions(hot anger, cold anger, neutral and sadness). Their total accuracy was 87% based on Support Vector Machine (SVM).

In our case study, we use only 12 features to classify between 5 types of emotions based on Bayesian Quadratic Discriminate Classifier. For ADAS system, we focus on the 3 classes (sad, normal and angry), because they are strongly related to the representation of the behavior of drivers.

	Normal	Sad	Angry	Total
Training set	69	52	117	238
Test set	10	10	10	30
False Positive	1	1	2	4
Detection ratio	90%	90%	80%	86,67%

Table 3 The obtained results using three emotions (sad, angry and normal using QDC

Here, we present the experimental results after using different combinations of emotion classes. We use a quadratic discriminate classifier (QDC) and Berlin data base for emotional voices . For feature extraction, the statistical moments (minimum, maximum, mean and median) of 3 features (MFCC, Pitch and energy) are extracted. Table 3 shows that we used for training (69 voice files for normal, 52 for sad and 117 for angry) and for testing (10 for normal, 10 for Sad, 10 for angry). A 90% of normal and sad signals are correct classified while 80% of angry, So a 86.67% where the total result of our classifier.

	Happy	Normal	Sad	Angry	Total
Training set	60	69	52	117	298
Test set	10	10	10	10	30
False Positive	7	1	1	3	12
Detection ratio	30	90%	90%	70%	70%

Table 4 The obtained results after adding the fear emotion using QDC

Table 5 shows that the fear emotion is added to the training set to see the influence of this class on the classification in general. We retrain the QDC classifier and then we use the following test data (9 of fear, 10 of normal, 10 of sad, 10 of angry), we obtained the following results, 6 of 9 Afraid voices where false positives which forms 33.33% of success. This result is low because of the lack of training data (60 only) but the positive side of this experiment is that the recognition ratio of (normal, sad and angry) emotions is increased to (93.33%, 91.66% and 84.32%).

	fear	Normal	Sad	Angry
Training set	60	69	52	117
Test set	9	10	10	10
False Positive	6	1	3	12
Detection ratio	33,33	93,33%	91,66%	84%

Table 5 The obtained results after adding the fear emotion using QDC

Table 4 shows that the happiness emotion is added to the training set to see the influence of this class on the classification in general. We retrain the QDC classifier and then we use the following test data (10 of happy, 10 of normal, 10 of sad, 10 of angry). We obtained the following results, 7 of 10 happy voices are false positives which form 30% of success, this low detection ration is because of the low number of the training sets. We also realize that the recognition ration is decreased (70%) for angry class, while no change is occurred to sadness and normal classes.

8 Conclusion

Speech emotion recognition system will be useful to understand the state and emotion of a driver. In this work, we came to know that the acoustic information could help to increase the performance of ADAS systems.

However, we show that the usage of Bayesian Quadratic Discriminate Classifier (QDC) enables a real-time processing with a low amount of features (12 features). We are able to reduce the calculation cost while keeping high recognition rate.

In our future work, we will perform the evaluation over different databases to check the robustness of the algorithms and to see the scalability of the algorithms. Further this work can be extended in the direction of reducing the acoustic noise

generated by the vehicle and the vehicle entertainment systems to improve the quality of the ADAS system.

References

1. Yang, Y.H. and Liu, C.C. and Chen, H.H., *Music emotion classification: a fuzzy approach*, 3rd ed. Proceedings of the 14th annual ACM international conference on Multimedia, ACM, pp 81-84, 2006.
2. Trohidis, K. and Tsoumakas, G. and Kalliris, G. and Vlahavas, I., *Multilabel classification of music into emotions*, 3rd ed. Proc. 9th International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia, PA, USA, 2008.
3. Yang, Y.H. and Lin, Y.C. and Su, Y.F. and Chen, H.H., *Music emotion classification: A regression approach*, 3rd ed. Multimedia and Expo, 2007 IEEE International Conference, pp 208-211, 2007.
4. Thayer, Robert E., *The biopsychology of mood and arousal*, 3rd ed. Book (ISBN 0195051629), New York, 1989.
5. Synak, P. and Wieczorkowska, A., *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 3rd ed. Journal pp 314-322, Springer, 2005.
6. Jian Zhou, Sch. of Inf. Sci.& Technol., Southwest Jiaotong Univ., Chengdu, *Speech Emotion Recognition Based on Rough Set and SVM*, 3rd ed. 5th IEEE International Conference on Cognitive Informatics, Beijing, 2006.
7. Yacoub, S. and Simske, S. and Lin, X. and Burns, J., *Recognition of emotions in interactive voice response systems*, 3rd ed. Eighth European conference on speech communication and technology, 2003.
8. J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 3, pp. 218 – 224, 2009. [Online]. Available: `http://www.sciencedirect.com/science/article/B6VN8-4VFBYG1-1/2/07087f8c3b6f88f9e9ed6996388d01ed`
9. W. Jones, "Keeping cars from crashing," *Spectrum, IEEE*, vol. 38, no. 9, pp. 40–45, Sep. 2001. [Online]. Available: `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=946636\&tag=1`
10. Nwe, T.L. and Foo, S.W. and De Silva, L.C.s, *Speech emotion recognition using hidden Markov models*, 3rd ed. Journal of Speech communication, volume 41, pp 603–623, Elsevier, 2003.
11. Wu, S. and Falk, T.H. and Chan, W.Y., "Automatic speech emotion recognition using modulation spectral features," *Speech Communication*, issn 0167-6393, Elsevier, 2010.
12. Maragos, P. and Kaiser, J.F. and Quatieri, T.F., "Energy separation in signal modulations with application to speech analysis," *Signal Processing, IEEE Transactions on*, volume 41, pp 3024–3051, 1993.
13. Molau, S. and Pitz, M. and Schluter, R. and Ney, H., "Computing Mel-frequency cepstral coefficients on the power spectrum," *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 1, pp 73–76, 2001.
14. Leslie Jr, G.F. and MacKay, K.W., "Method and apparatus for pitch controlled voice signal processing," *Google Patents*,US Patent 4,700,391, 1987.
15. Srivastava, S. and Gupta, M.R. and Frigyik, B.A., "Bayesian quadratic discriminant analysis," *Journal of Machine Learning Research*, volume 8, pp 1287–1314, 2007.