

# NORMALIZING MULTI-SUBJECT VARIATION FOR DRIVERS' EMOTION RECOGNITION

Jinjun Wang, Yihong Gong

NEC Laboratories America, Inc. Cupertino, CA 95014, USA  
{jjwang,ygong}@sv.nec-labs.com

## ABSTRACT

The paper attempts the recognition of multiple drivers' emotional state from physiological signals. The major challenge of the research is the severe inter-subject variation such that it is extreme difficult to build a general model for multiple drivers. In this paper, we focus on discovering an optimal feature mapping by utilizing the additional attribute from the drivers. Two models are reported, specifically an auxiliary dimension model and a factorization model. Experimental results show that the proposed method outperform existing algorithms used for emotional state recognition.

## 1. INTRODUCTION

The availability of on-board electronics and in-vehicle information systems has demanded the development of more intelligent vehicles. One such important intelligence is the possibility to recognize the driver's emotional state to prevent potential driving risks or to develop more user friendly driver-car-interactions. Due to wide application scenarios and great commercial potentials, understanding the driver's emotional state has been listed as one of the key area for improving intelligent transportation systems by many leading global car industries and manufacturers.

To recognize human emotion, many researchers have focused on facial expression and/or speech analysis [1]. However, under the in-car driving environment, the heavy noise significantly decreases the performance of these techniques. On the other side, the physiological features have been proven to be very effective for monitoring human mental state [2]. Examples related to driving applications include the ASV (Advanced Safety Vehicle) system [3] and the SmartCar project [4], where researchers sought to discover effective physiological and bio-behavioral measures to model the driver's vigilance [2], stress [5], fatigue or drowsiness. Since emotional state is closely related to mental state, using physiological feature to recognize human emotions has also been attempted [6, 7, 8] for in-door environment, and promising results were reported.

Comparing to emotion recognition under in-door condition [6, 7, 8], collecting the physiological data during driving is more difficult. To remove motion artifacts, we attached

all the sensors to the driver's left foot which is relatively more stable; To induce and sustain the driver into required emotional state, our psychologist carefully designed different guidance voices and driving courses. After the driving session, the subject is required to finish a questionnaire from which our psychologist can judge whether the emotional state is correct and reliable. We used a physiological sensing system called "FlexComp Infiniti" to connect four sensors to the driver, specifically the *Respiration (RESP)*, *skin conductance (SC)*, *temperature (TEMP)* and *blood volume pulse (BVP)* sensors. Figure 1 shows the data collection setup. 13 subjects participated in the data collection. Each subject drove five sessions under the five emotional state respectively. Each session is 15 to 20 minutes long. Based on these sensor signals, we derived an  $R^{25}$  statistical feature vector in a 60sec sliding window with 10sec step-size [9]. In this paper, we aim to detect five emotional states, specifically "Happy", "Sad", "Angry", "Fatigue" and "Neutral".

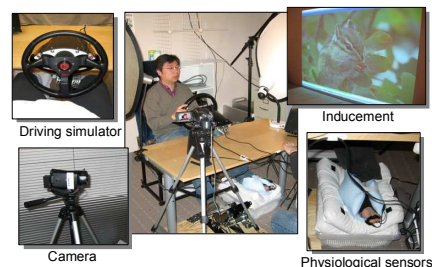


Fig. 1. Data collection protocol

The remainder of the paper is organized as follows: Section 2 discussed the classification difficulty caused inter-subject variation; Section 3 presents our proposed method to normalize such variation by discovering optimal subject-dependent representations and subject-independent classification model; Section 4 lists the experimental results; And section 5 concludes the work and discusses some future work.

## 2. INTER-SUBJECT VARIATION

Recognizing the driver's emotional state is a classification problem. For single subject [6, 8] or subjects from a very limited age group [7], Support Vector Machine (SVM) or even

simple linear classifiers (LDA) can be used to recognize their emotional states satisfactorily. This conforms to our experimental results that, if we train/test using any single driver's data, LDA/SVM can achieve 90%/96% accuracy. However, if we concatenate multiple drivers' features together for training, the obtained classifier perform poorly. As can be seen from figure 2, although the emotional states are separable within a single driver (figure 2.a and b), due to high inter-subject variation, features of different emotional state from different drivers overlap with each other (figure 2.c), which makes the recognition of multiple drivers' emotion extremely difficult. To cope with the problem, in a previous work [9], we proposed a subject representation model to recognize the driver ID for model selection, and applied additional constraint, such as temporal smoothness, to improve the recognition accuracy 9% better than simple SVM classifier. The method in [9] works fine when each subject has contributed some samples for training. However, the leave-one-out cross-validation result was poor because with only 13 subjects (i.e., 12 can be used for training), the learned latent model has very low generalization ability to predict any unseen subject's data.

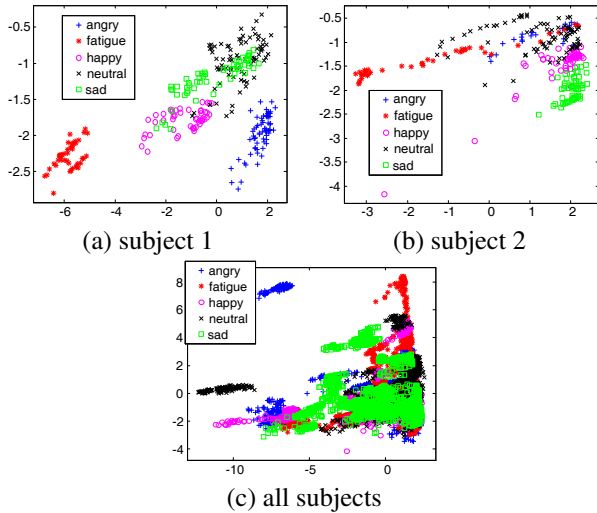


Fig. 2. Latent Variable Model

### 3. LATENT SUBJECT REPRESENTATION

To cope with the aforementioned difficulty, we want to build a model with higher generalization ability. Our idea is to apply a latent variable to represent the attribute from the subject, and apply this latent representation together with the subject's physiological data for emotion recognition. In contrast to [9] where the auxiliary variable represents the discrete subject ID for model selection, the latent variable in this paper is continuous. It can be understood as an optimal mapping for a particular subject such that his transformed features can be classified using a generic classification model.

To elaborate, denote each input feature vector from subject  $i$  as  $\mathbf{x}_i$  which is an  $D_x \times 1$  column vector. We encode the fact that  $\mathbf{x}_i$  belongs to emotion class  $j \in \{1, 2, \dots, M\}$  by a  $M$ -dimensional 0/1 valued row vector  $\mathbf{y}_i = [y_1, \dots, y_M]^\top$ , where  $y_j = 1$  and all other coordinates are 0. Now we want to build a function  $f_{\theta_j}(\mathbf{x}) = Pr(y_j = 1 | \mathbf{x}, \theta_j)$  that predicts the probability that  $\mathbf{x}_i$  belongs to emotional state  $j$ . As we explained above, the  $\mathbf{x}_i$  are biased from the characteristic of subject  $i$ . Hence to normalize this inter-subject variation, we further applied an  $D_w \times 1$  parameter  $\mathbf{w}_i$  with different values for different subject to transform the  $\mathbf{x}_i$ . Now the training process strive to find a general parameter  $\theta_j$  for all subject, and the subject-dependent parameter  $\mathbf{w}_i$  for the  $i^{th}$  subject. Finally, another function  $\mathbf{w}^* = g_\phi(\mathbf{x})$  is trained to infer  $\mathbf{w}$  from  $\mathbf{x}$ , such that for feature vectors  $\mathbf{x}$  from an unseen subject, we can first predict his subject-dependent mapping  $\mathbf{w}$ , then use both  $\mathbf{w}$  and  $\mathbf{x}$  to predict the emotional state.

Now we have both the samples  $\mathbf{x}_i$  and its label  $\mathbf{y}_i$ , as well as the subject ID  $i$  from which  $\mathbf{x}_i$  is extracted, we can use the following strategy to train  $\mathbf{w}_i$  and  $\theta$ .

#### 3.1. Linear regression

If we assume that  $Pr(\mathbf{y}_i | \mathbf{x}_i, \mathbf{w}_i, \theta)$  follows Gaussian distribution, i.e.,

$$Pr(\mathbf{y}_i | \mathbf{x}_i, \mathbf{w}_i, \theta) = \alpha \exp \left[ \frac{-(f_\theta(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{y}_i)^\top (f_\theta(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{y}_i)}{2\Sigma} \right], \quad (1)$$

where  $\Sigma$  is an unknown but fixed covariance, and

$$f_\theta(\mathbf{x}_i, \mathbf{w}_i) = [f_{\theta_1}(\mathbf{x}_i, \mathbf{w}_i), f_{\theta_2}(\mathbf{x}_i, \mathbf{w}_i), \dots, f_{\theta_M}(\mathbf{x}_i, \mathbf{w}_i)]. \quad (2)$$

It can be easily proven that the Maximum-likelihood (ML) estimation of  $\mathbf{w}_i$  and  $\theta$  equals to minimizing the following cost function,

$$l(\mathbf{w}_i, \theta) = (f_\theta(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{y}_i)^\top (f_\theta(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{y}_i). \quad (3)$$

For simplicity, the cost function with respect to all the subjects' training data can be written as

$$l(\mathbf{W}, \theta) = \sum_{i=1}^P \left[ tr(f_\theta(\mathbf{X}_i, \mathbf{W}_i) - \mathbf{Y}_i)^\top (f_\theta(\mathbf{X}_i, \mathbf{W}_i) - \mathbf{Y}_i) \right], \quad (4)$$

where  $\mathbf{X}_i$  is an  $D_x \times D_i$  matrix representing all the training samples  $\{\mathbf{x}_i\}$  from the  $i^{th}$  subject,  $\mathbf{Y}_i$  is the  $M \times D_i$  matrix representing all their labels, and  $\mathbf{W}_i$  is an  $D_w \times D_i$  matrix obtained by repeating  $\mathbf{w}_i$   $D_i$  times, and  $D_i$  is the number of samples in the  $i^{th}$  subject's data.

Usually the ML estimation is not robust. Hence we applied the Tikhonov regularization to obtain the Maximum-a-posteriori (MAP) estimation. The cost function for MAP estimation can be written as

$$l(\mathbf{W}, \theta) = \sum_{i=1}^P \left\{ \text{tr} \left[ (f_{\theta}(\mathbf{X}_i, \mathbf{W}_i) - \mathbf{Y}_i)^{\top} (f_{\theta}(\mathbf{X}_i, \mathbf{W}_i) - \mathbf{Y}_i) \right] + \lambda \mathbf{w}_i^{\top} \mathbf{w}_i \right\} + \alpha \text{tr}(\theta^{\top} \theta), \quad (5)$$

where  $\lambda$  and  $\alpha$  are given parameters defining the weights of the regularization term.

### 3.2. Multinomial Logistic Regression

In the previous section, as can be seen from Eq. (2), in the linear regression model, training one class is independent of any other classes. Hence the learned  $\theta$  does not impose a competition between different classes. For this reason, the multinomial logistic regression is more suitable for the multi-class classification problem. In multinomial logistic regression, the probability of latent variable model is

$$\text{Pr}(\mathbf{y}_i | \mathbf{x}_i, \mathbf{w}_i, \theta) = \frac{\exp(f_{\theta}^{\top}(\mathbf{x}_i, \mathbf{w}_i) \mathbf{y}_i)}{\exp(f_{\theta}^{\top}(\mathbf{x}_i, \mathbf{w}_i) \mathbf{I}_M)} \quad (6)$$

where  $\mathbf{I}_M$  is the  $M \times 1$  unity vector.

ML estimation of  $\theta$  and  $\mathbf{w}_i$ ,  $i = 1, \dots, P$  equals to minimizing the negative log-likelihood cost function

$$l(\mathbf{w}_i, \theta) = \ln \left[ \exp(f_{\theta}^{\top}(\mathbf{x}_i, \mathbf{w}_i) \mathbf{I}_M) \right] - f_{\theta}^{\top}(\mathbf{x}_i, \mathbf{w}_i) \mathbf{y}_i \quad (7)$$

Similar to Eq. (4), the cost function with respect to all the training data can be written as

$$l(\mathbf{W}, \theta) = \sum_{i=1}^P \left\{ \ln \left[ \exp(f_{\theta}^{\top}(\mathbf{X}_i, \mathbf{W}_i) \mathbf{I}_M) \right] \mathbf{I}_{D_i} - \text{tr}(f_{\theta}^{\top}(\mathbf{X}_i, \mathbf{W}_i) \mathbf{Y}_i) \right\}, \quad (8)$$

where  $\mathbf{I}_{D_i}$  is an  $D_i \times 1$  unity vector.

To improve robustness, the MAP estimation can be obtained by minimizing the following cost function

$$l(\mathbf{W}, \theta) = \sum_{i=1}^P \left\{ \ln \left[ \exp(f_{\theta}^{\top}(\mathbf{X}_i, \mathbf{W}_i) \mathbf{I}_M) \right] \mathbf{I}_{D_i} - \text{tr}(f_{\theta}^{\top}(\mathbf{X}_i, \mathbf{W}_i) \mathbf{Y}_i) + \lambda \mathbf{w}_i^{\top} \mathbf{w}_i \right\} + \alpha \text{tr}(\theta^{\top} \theta), \quad (9)$$

### 3.3. Inferring the optimal $\mathbf{w}$ and Training

As mentioned above, we need a further function  $g_{\phi}(\mathbf{x})$  to infer the optimal  $\mathbf{w}^*$ . With no prior knowledge of  $\phi$ , it is reasonable to assume an Gaussian distribution for  $g_{\phi}(\mathbf{x})$ . Hence  $\phi$  can be estimated using regression method to minimize the following cost,

$$l(\phi) = \text{tr} \left[ (g_{\phi}(\mathbf{X}) - \mathbf{W})^{\top} (g_{\phi}(\mathbf{X}) - \mathbf{W}) \right] + \beta \text{tr}(\phi^{\top} \phi), \quad (10)$$

Now training the whole system equals to finding the optimal  $\theta$  and  $\phi$ , using  $\mathbf{W}$  as an interim variable. For efficient training, both  $f_{\theta_j}(\mathbf{x}, \mathbf{w})$  and  $g_{\phi}(\mathbf{x})$  are given smooth parametric forms such that the optimal  $\theta$  and  $\phi$  can be solved analytically. In our implementation, we used the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) method [10] for optimization.

## 4. EXPERIMENTAL RESULTS

### 4.1. Auxiliary dimension model

The forms of the  $f_{\theta_j}(\mathbf{x}, \mathbf{w})$  function evaluated in our experiments can be grouped into two categories. The first category is called the *Auxiliary dimension model* (ADM) where  $\mathbf{w}$  is used by concatenating itself to  $\mathbf{x}$  to make a new vector. Then both a linear function and a multi-RBF function were tested, where for the linear function,

$$f(\mathbf{x}, \mathbf{w}, \theta_j) = \mathbf{a}_j \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \end{bmatrix} + \mathbf{b}_j, \quad (11)$$

and for the multi-RBF function

$$f(\mathbf{x}, \mathbf{w}, \theta_j) = \sum_{t=1}^T (\alpha_{tj} \exp(-\sigma \| \begin{bmatrix} \mathbf{x} \\ \mathbf{w}_t \end{bmatrix} - \boldsymbol{\mu}_{tj} \|^2)), \quad (12)$$

where  $T$  is the number of RBFs.

The  $g_{\phi}(\mathbf{x})$  also can take either the linear or the multi-RBF function. The following figure 3.(a) lists the classification accuracy with respect to different  $D_w$  value (dimension of  $\mathbf{w}$ ), in comparison to LDA, GMM and SVM classifiers, as well as the method proposed in [9].

### 4.2. Factorization model

Alternatively, we can apply the  $\mathbf{w}$  by factorizing  $\theta$  into  $K$  components, i.e.,  $\mathbf{w} = [w_1, w_2, \dots, w_K]^{\top}$ . This is called the *Factorization model* (FM) Similar to subsection 4.1, we can have the linear function as

$$f_{\theta_j}(\mathbf{x}, \mathbf{w}) = \sum_{k=1}^K w_k (\mathbf{a}_{kj} + \mathbf{b}_{kj}), \quad (13)$$

and the multi-RBF function

$$f_{\theta_j}(\mathbf{x}, \mathbf{w}) = \sum_{k=1}^K (w_k \exp(-\sigma \|\mathbf{x} - \boldsymbol{\mu}_{kj}\|^2)), \quad (14)$$

The same  $g_{\phi}(\mathbf{x})$  functions as in subsection 4.1 were also tested. can take either the linear or the multi-RBF function. The following figure 3.(a) lists the classification accuracy with respect to different  $D_w$  value (dimension of  $\mathbf{w}$ ), in comparison to LDA, GMM and SVM classifiers, as well as the method proposed in [9].

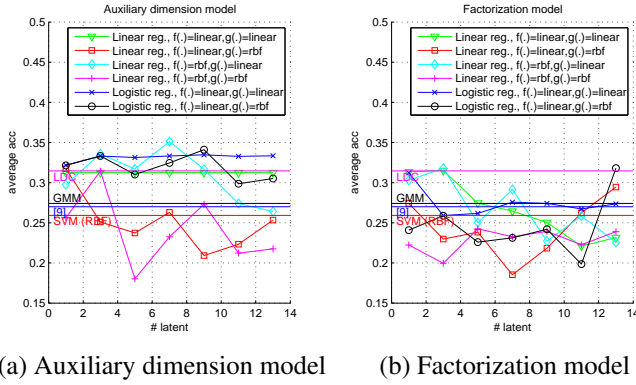


Fig. 3. Classification accuracy

As can be seen from figure 3, the FM model gives comparable performance to those benchmark methods, and the ADM works better than these methods. This results demonstrate the effectiveness of using a latent subject representation  $\mathbf{w}$  to normalize the inter-subject variation.

To further explain why ADM works better, according to Eq. (11) and Eq. (12), although different dimensions were tested for  $\mathbf{w}$ , the variable that really take effect is always a single scaler, e.g., in Eq. (11),  $\mathbf{a}_j \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \end{bmatrix} + \mathbf{b}_j = \mathbf{a}_j^x \mathbf{x} + \mathbf{a}_j^w \mathbf{w} + \mathbf{b}_j$ , hence no matter what dimension  $\mathbf{w}$  is, the latent variable that takes effect is  $\mathbf{a}_j^w \mathbf{w}$  which is a one dimension scaler. This means that the number of free variables to be trained in the ADM is less than in the FM. Since theoretically the ADM can be regarded as a special case of the FM, the later should give no worse accuracy than the former. Hence the results indicates that number of training samples is not enough for the FM model. On the other side, when the dimension of  $\mathbf{w}$  takes the smallest possible value one in the FM, the classification accuracy is most comparable to the ADM.

In addition, comparing the two regression strategies used in the ADM, the logistic regression achieved a very constant accuracy under different  $\mathbf{w}$  dimension, while the linear regression leads to significant variation. Given the above discussion where the training sample size may not be large enough, we tend to value the model obtained by logistic regression.

## 5. CONCLUSION AND FUTURE WORK

In this paper we propose a method to recognize multiple drivers' emotional state. To deal with inter-driver variation, we apply a latent variable to represent the attribute of individual driver and use statistical method for training. The applied latent representation works as an optimal feature mapping that normalize the bias from each individual subject, such that different subjects' data can be classified using a same model. Experimental results show that the proposed method achieves higher accuracy than existing algorithms for multiple drivers' emotional state recognition problem. In the future, we'll collect addition data for validation.

## 6. REFERENCES

- [1] Nicu Sebe, Ira Cohen, Theo Gevers, and T-S. Huang, "Multimodal approaches for emotion recognition: A survey," *Proc. of the SPIE Internet Imaging*, pp. 56–67, 2004.
- [2] L. Bergasa, J. Nuevo, M. Steloand, R. Barea, and M. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Trans. on ITS*, , no. 1, pp. 63–77, 2006.
- [3] Albert Kircher, Marcus Uddman, and Jesper Sandin, "Vehicle control and drowsiness," *Swedish National Road and Transport Research Institute*, 2002.
- [4] J. Healey and R. Picard, "Smatcar: Detecting driver stress," *Proc. of IEEE ICPR'00*, pp. 218–221, 2000.
- [5] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. on ITS*, 2005.
- [6] Rosalind W Picard, Elias Vyzas, and Jennifer Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. on PAMI*, , no. 10, pp. 1175–1191, 2001.
- [7] K. Kim, S.Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, pp. 419–427, 2004.
- [8] J. Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *Proc. of IEEE ICME'05*, pp. 940–943, 2005.
- [9] Jinjun Wang and Yihong Gong, "Recognition of multiple drivers' emotional state," *Proc. of IEEE ICPR'08*, 2008.
- [10] C. Zhu, R. H. Byrd, and J. Nocedal, "L-bfgs-b: Algorithm 778: L-bfgs-b, fortran routines for large scale bound constrained optimization," *ACM Transactions on Mathematical Software*, pp. 550–560, 1997.