Recognition of Multiple Drivers' Emotional State

Jinjun Wang, Yihong Gong
NEC Laboratories America, Inc., Cupertino
$\{$*jjwang, ygong*$\}$*@sv.nec-labs.com*

Abstract

The paper attempted the recognition of multiple drivers' emotional state from physiological signals. The major challenge of the research is due to the severe inter-driver variation such that the features of different emotional state are high correlated, and it is found that simple decorrelation method cannot normalize the features well to achieve acceptable classification accuracy. Hence, in this paper, we propose to apply a latent variable to represent the hidden attribute of individual driver and use statistical training. In addition, we applied temporal constraints for the inference process to improve the recognition accuracy. 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 [7]. 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 [1]. Examples related to driving applications include the ASV (Advanced Safety Vehicle) system [5] and the SmartCar project [2], where researchers sought to discover effective physiological and bio-behavioral measures to model the driver's vigilance [1], stress [3], fatigue or drowsiness. Since emotional state is closely related to mental state, using physiological feature to recognize human emotions has also been attempted [6, 4, 8] for in-door environment, and promising results were reported. In this paper, we aim to use physiological features for emotion recognize under in-car environment. Five emotional states are to be recognized, specifically "Happy", "Sad", "Angry", "Fatigue" and "Neutral".

Comparing to emotion recognition under in-door condition [6, 4, 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 the features used in [2, 6, 3], we derived an R^{25} feature vector in a 60sec sliding window

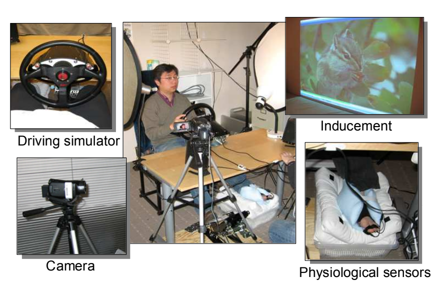

Figure 1. Data collection protocol

Sensor	Feature
RESP	$mean$, std, $sp_{0,0.1}$, $sp_{0.1,0.2}$, $sp_{0.2,0.3}$, $sp_{0.3,0.4}$
SC	$mean$, $mean_d$, $sp_{0,0.1}$, $sp_{0.1,0.2}$, $sp_{0.2,0.3}$, $sp_{0.3,0.4}$
TEMP	$mean$, $mean_d$, $sp_{0,0.1}$, $sp_{0.1,0.2}$, $sp_{0.2,0.3}$, $sp_{0.3,0.4}$
BVP	$mean$, $mean_d$, $sp_{0,0.1}$, $sp_{0.1,0.2}$, $sp_{0.2,0.3}$, $sp_{0.3,0.4}$, $sp_{0,0.08}/sp_{0.15,0.5}$

Table 1. Feature Derivation

$mean_d$ is the mean of the 1st order forward difference; $sp_{a,b}$ is the spectrum power between $a \sim b$Hz

with 10sec step-size (Table 1).

Recognizing the driver's emotional state is a classification problem. For single subject [6, 8] or subjects from a very limited age group [4], Support Vector Machine (SVM) or even simple linear classifiers (FDA) 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, FDA/SVM can achieve 90%/96% accuracy. However, if we concatenate multiple drivers' features together for training, the obtained classifier perform poorly (As listed in Table 2 in section 3). Analyzing the feature distribution reveals that, although the emotional states are separable within a single driver, features of different emotional state from different drivers overlap with each other due to high inter-driver variation, which makes the recognition of multiple drivers' emotion difficult. To cope with the problem, in this paper, we propose to use a latent variable to model the attribute from the individual driver to assist the final recognition process. In addition, under the driving scenario, certain prior knowledge can also be modeled to further improve accuracy.

The remainder of the paper is organized as follows: Section 2 presents our proposed algorithms for multiple drivers' emotion recognition; Section 3 lists the experimental results; And section 4 concludes the work and discusses some future work.

2 Dynamic Model with Latent Variable

As explained in Section 1, the features of different emotional state overlap with each other due to the existence of a bias attribute for different drivers' feature. For example, certain peoples' heart rate is always higher than the others. One possibility to solve the problem is to normalize the features from different drivers. However, simple decorrelation method does not improve the overall recognition accuracy in our experiments. Hence we instead apply a latent variable z to represent the attribute of the driver, and use statistical method to discover such attribute. In addition, from the same driver, his attribute, as well as his emotional state, should possess temporal consistencies, which can also be utilized to assist emotion recognition. In brief, we can first guess which driver most likely generates the observed feature vector, and then use that driver's model, as well as temporal information, to recognize his emotional state. Both the latent variable, as well as the temporally preceding estimation, can be intergraded out for robustness. The following subsections elaborate the proposed algorithm in more detail.

2.1 Latent Representation

Let x be the derived physiological feature vector, and y be the emotional state. Instead of directly computing MAP $P(y|x)$ (Figure.2.a), we want to apply a latent variable z to encode the attribute of a particular driver, and seek to maximize $P(y|x,z)$ to obtain more robust estimation (Figure.2.b). Let $y \in \{1,2,3,4,5\} = \{$Angry, Happy, Sad, Fatigue, Neutral$\}$, and $z \in \{1,2,...,M\}$ where M is the number of drivers. To simplify the notation, denote $P(z=i)$ as $P(z_i)$ and $P(y=j)$ as $P(y_j)$. We can have

$$P(y|x) = \sum_{i=1}^{M} \left(P(y|z_i,x)P(z_i|x) \right), \qquad (1)$$

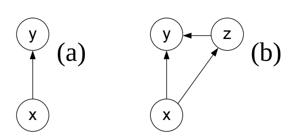

Figure 2. Latent Variable Model

2.2 Temporal Consistence of Latent Variable

The next strategy to improve the recognition accuracy is to use temporal information. In real driving scenario, the feature vectors come from the same driver, hence they possess consistent attribute encoded by z (Figure.3.a), which can be utilized for inference. To elaborate, let suffix t denote the time instance, e.g., x^t means the feature vector at time t. Also, let X^t be a matrix containing all the feature vectors from time instance 1 to t. To apply the temporal consistence constraint for z^t, in contrast to Eq (1), we are to estimate $P(y^t|X^t)$ by

$$P(y^t|X^t) = \sum_{i=1}^{M} \left(P(y^t|z_i^t,x^t)P(z_i^t|X^t) \right), \qquad (2)$$

where

$$P(z^t|X^t) = \sum_k \big(P(z^t|z_k^{t-1}, x^t)P(z_k^{t-1}|X^{t-1})\big) \quad (3)$$

$$= \sum_k \Big(\frac{P(x^t|z^t)P(z^t|z_k^{t-1})}{\sum_p Pr\big((x^t|z_p^t)P(z_p^t|z_k^{t-1})\big)}P(z_k^{t-1}|X^{t-1})\Big),$$

which is recursive with $P(z^0|X^0) = P(z|\theta_l)$. And

$$P(y^t|z^t, x^t) = \frac{P(x^t|z^t, y^t)P(y^t|z^t)}{\sum_q \big(P(x^t|z^t, y_q^t)P(y_q^t|z^t)\big)} . \quad (4)$$

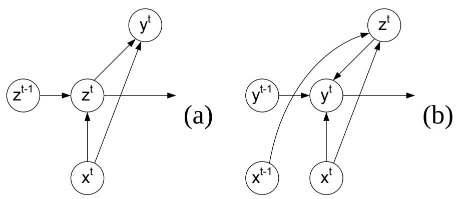

Figure 3. Temporal Transition Model

2.3 Temporal Evolving of Emotional State

An intuitive extension to subsection 2.2 is to also apply temporal constraints on the emotional state y^t, because a driver's emotion will not change too fast. Hence a more elegant dynamic model is to evolve y^t from y^{t-1}. On the other side, now we don't need to evolve z because otherwise we will have two dynamic variables. Instead, we can simply regard the probability of $P(z|X^{t-1})$ as the prior of z for current estimation, and regard the obtained posterior $P(z|X^t)$ as the prior for the next time instance (Figure.3.b). Hence, in contrast to Eq (2), we are to estimate $P(y^t|X^t)$ by

$$P(y^t|X^t) = \sum_{i=1}^{M} \big(P(y^t|z_i, X^t)P(z_i|X^t)\big), \quad (5)$$

where

$$P(y^t|z_i, X^t) = \sum_q \big(P(y^t|y_q^{t-1}, z_i, x^t)P(y_q^{t-1}|z_i, X^{t-1})\big)$$

$$= \sum_q \Big(\frac{P(x^t|y^t, z_i)P(y^t|y_q^{t-1}, z_i)}{\sum_p P(x^t|y_p^t, z_i, y_q^{t-1})P(y_p^t|y_q^{t-1}, z_i)}$$

$$P(y_q^{t-1}|z_i, X^{t-1})\Big), \quad (6)$$

which is recursive with $P(y^0|z_i, X^0) = P(y|z_i)$. And

$$P(z_i|X^t) = \frac{P(x^t|z_i)}{P(x^t|\theta_l)}P(z_i|X^{t-1}), \quad (7)$$

which is also recursive, with $P(z_i|X^0) = P(z_i|\theta_l)$.

2.4 Training

Training the system requires learning the parameters for the following models: First, the two generative models, specifically $P(x|z)$ in Eq (3) (7), and $P(x|y, z)$ in Eq (4) (6). They are simply modeled using Gaussian Mixture model (GMM). And second, the two transition probabilities, $P(z_i|z_k)$ in Eq (3), and $P(y_j|y_q, z)$ in Eq (6). This is difficult because we don't have such transitions in our training data. To cope with the difficulty, we assume that, similar generative models should have high transition probability between each other. Hence we can use the obtained $P(x|z)$ and $P(x|y, z)$ to learn the transition probability of $P(z_i|z_k)$ and $P(y_j|y_q, z)$ respectively. To depict, let's use $P(z_i|z_k)$ for example, assuming $Pr(x|z_i) \sim \{[w_1, \mu_1, \Sigma_1]^i,...,[w_{L_i}, \mu_{L_i}, \Sigma_{L_i}]^i\}$ and $Pr(x|z_k) \sim \{[w_1, \mu_1, \Sigma_1]^k,...,[w_{L_k}, \mu_{L_k}, \Sigma_{L_k}]^k\}$, then

$$P(z_i|z_k) = exp\big(-Dist(z_i, z_k)/500\big),$$

where the distance measure $Dist(z_i, z_k)$ is computed based on KL divergence, i.e.,

$$Dist(z_i, z_k) = \min_{j=1,\dots,L_i} \sum_{p=1}^{L_k} w_p^k KL\big([\mu_j^i, \Sigma_j^i], [\mu_p^k, \Sigma_p^k]\big),$$

where

$$KL\big([\mu_j^i, \sigma_j^i], [\mu_p^k, \sigma_p^k]\big)$$
$$= \frac{1}{2}\mathrm{Tr}(\Sigma_p^{k\,-1}\Sigma_j^i) - \frac{1}{2}\mathrm{In}\big(\mathrm{Det}(\Sigma_p^{k\,-1}\Sigma_j^i)\big) - \frac{D}{2}$$
$$+ \frac{1}{2}\mathrm{Tr}\big(\Sigma_p^{k\,-1}(\mu_j^i - \mu_p^k)(\mu_j^i - \mu_p^k)^\top\big).$$

where D is the dimension of feature vector x. $P(z_i|z_k)$ is then normalized according to $\sum_{i=1}^{M} P(z_i|z_k) = 1$.

3 Experimental Results

In the experiment, we used the beginning half of every driver's data together for training, and use the second half of each driver's data for testing. Table 2 lists the recognition accuracy of the second half in comparison amongst FDA, GMM, SVM, and the methods proposed in subsection 2.2 (Model 1) and 2.3 (Model 2) respectively. As can be seen from Table 2, both of our proposed methods outperform the benchmark algorithms.

Table 2. Recognition Accuracy

FDA	GMM	SVM	Model 1	Model 2
38%	43%	69%	78%	80%

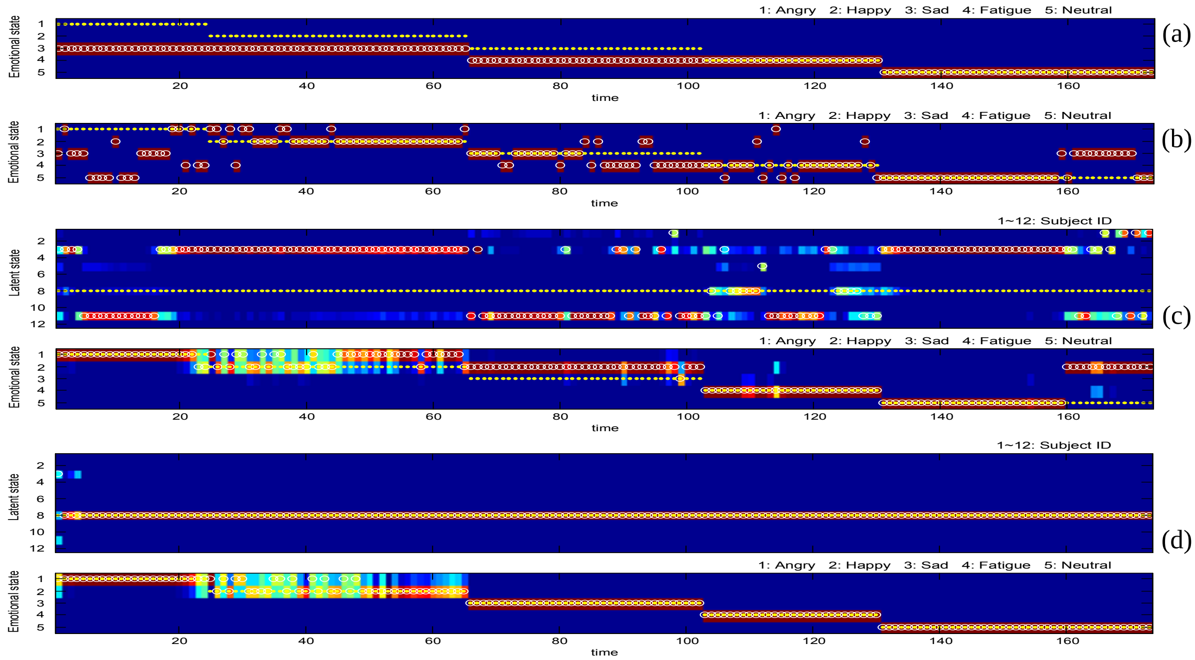

Figure 4. Dynamic Inference Process for Subject 8

(a) FDA; (b) SVM; (c) Model 1; (d) Model 2. The yellow dot represents the ground-truth; the white circle represents the predicted result; the background color represents the probability.

Next, we concatenated the five driving sessions of the same driver together to examine the inference process. As can be seen from Figure.4, both of our proposed methods achieve higher accuracy than FDA and SVM by first estimating the latent variable (the upper row of Figure.4.c and .d) to assist emotion recognition. Beside, they can also capture the emotion change quickly. The reason to select the 8^{th} driver's data for illustration is because, in his example, the estimation of the latent variable by model 1 (Figure.4.c) is poorer than model 2 (Figure.4.d), hence the final emotion recognition accuracy is also poorer. This proves the importance of using the latent variable. In most cases, model 2 seems more practical for the data, hence its performance is slightly better than model 1 (Table 2).

4 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. In addition, we apply temporal constraints for the inference process to improve the recognition accuracy. 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.

References

[1] L. Bergasa, et al. Real-time system for monitoring driver vigilance. *IEEE Trans. on Intelligent Transportation System*, (1):63–77, 2006.

[2] J. Healey and R. Picard. Smatcar: Detecting driver stress. *Proc. of ICPR'00*, pages 218–221, 2000.

[3] J. Healey and R. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. on Intelligent Transportation System*, 2005.

[4] K. Kim, S.Bang, and S. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Med. Biol. Eng. Comput.*, pages 419–427, 2004.

[5] A. Kircher, M. Uddman, and J. Sandin. Vehicle control and drowsiness. *Swedish National Road and Transport Research Institute*, 2002.

[6] R. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE PAMI*, (10):1175–1191, 2001.

[7] N. Sebe, et al. Multimodal approaches for emotion recognition: A survey. *Proc. of the SPIE Internet Imaging*, pages 56–67, 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 ICME'05*, pages 940–943, 2005.