Sensor feature selection and combination for stress identification using combinatorial fusion

Yong DENG  
*Peking University*

Zhonghai WU  
*Peking University*

Chao-Hsien CHU  
*Pennsylvania State University*, chchu@smu.edu.sg

Qixun ZHANG

D. Frank HSU

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis_research](https://ink.library.smu.edu.sg/sis_research)

Part of the Artificial Intelligence and Robotics Commons

Citation

DENG, Yong; WU, Zhonghai; CHU, Chao-Hsien; ZHANG, Qixun; and HSU, D. Frank. Sensor feature selection and combination for stress identification using combinatorial fusion. (2013). *International Journal of Advanced Robotic Systems*. 10, 306-313. Research Collection School Of Information Systems.  
Available at: [https://ink.library.smu.edu.sg/sis_research/2236](https://ink.library.smu.edu.sg/sis_research/2236)

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.
Sensor Feature Selection and Combination for Stress Identification Using Combinatorial Fusion

Yong Deng¹, Zhonghai Wu²⁺, Chao-Hsien Chu³, Qixun Zhang², and D. Frank Hsu⁴

1 School of Electronic Engineering and Computer Science, Peking University, Beijing, P.R. China
2 School of Software and Microelectronics, Peking University, Beijing, P.R. China
3 College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA
4 Department of Computer and Information Science, Fordham University, New York, NY, USA

* Corresponding author E-mail: wuzh@pku.edu.cn

Received 18 May 2012; Accepted 8 Mar 2013

© 2013 Deng et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract The identification of stressfulness under certain driving condition is an important issue for safety, security and health. Sensors and systems have been placed or implemented as wearable devices for drivers. Features are extracted from the data collected and combined to predict symptoms. The challenge is to select the feature set most relevant for stress. In this paper, we propose a feature selection method based on the performance and the diversity between two features. The feature sets selected are then combined using a combinatorial fusion. We also compare our results with other combination methods such as naïve Bayes, support vector machine, C4.5, linear discriminant function (LDF), and k-nearest neighbour (kNN). Our experimental results demonstrate that combinatorial fusion is an efficient approach for feature selection and feature combination. It can also improve the stress recognition rate.

Keywords Combinatorial Fusion, Feature Selection, Feature Fusion, Stress Identification, Sensor Fusion

1. Introduction

In a recent survey by the American Psychological Association [2], more than half of the Americans surveyed indicated that stress is a major cause of personal health problems. In addition, more than 94% of these adults believe that stress is an essential factor in the development of illnesses such as depression, heart disease and obesity. Stress can also trigger heart attacks, arrhythmias and sudden death. Therefore, it is important to track and understand an individual’s stress patterns by constantly and efficiently detecting his/her stress levels. By doing so, physicians are provided with much more reliable data and information with which to perform interventions and stress reduction if necessary.

In recent years, identifying the stress of human beings using multiple psychological sensors has received a lot of attention as a research topic. Existing studies ([7], [8], [12], [3], [19], [21], [4]) have shown that psychosocial stress can be recognized from the physiological information of a
human being. The physiological information can be acquired by biological or physiological sensors, which usually include: ECG (electrocardiogram), GSR (galvanic skin response), EMG (electromyogram) and RESP (respiration).

In the process of analysing and interpreting data, features were extracted from the raw physiological sensor data using the feature extraction methods. Then the most sensitive and relevant features are selected by using certain feature selection heuristics. Next, based on the selected features, a feature fusion procedure is applied to identify the stress level. Many feature fusion heuristics may be used in this procedure and different decision results can be acquired accordingly. Finally, we obtain the predicted stress level based on the feature fusion results (see Figure 1).

![Figure 1. Procedure of multiple sensor fusion](image1)

In this paper, we continue our previous work [5, 6] to focus on three main issues: (1) how to select the most relevant features based on individual performance and pair diversity, (2) how to combine features in order to accurately detect the stress level, and (3) how does the combinatorial fusion method compare to other conventional feature combination methods.

The organization of this paper is as follows. In section 2 we briefly review related work. Feature selection using performance and cognitive diversity is introduced in Section 3. The use of CFA to fuse the features is presented in Section 4. Our experimental results are summarized and described in Section 5. Finally, we give a conclusion and suggest future work that could result from the paper in Section 6.

2. Related work

In the experiment conducted by Healey and Picard [8], five kinds of wearable sensors were used: an EEG sensor, a hand and a foot GSR sensor, and a RESP sensor. People with these five sensors drove in and around downtown Boston on a pre-determined path. They went through three kinds of driving conditions classified as “rest”, “highway” and “city”. The signals of these sensors were recorded and analysed. A total of 22 features were extracted and feature combination was performed, using Linear Discriminant Function (LDF), to predict the driver’s stress level [7].

The data acquired by Healey and Picard’s experiment have been partially published on the website PHYSIONET [14]. It enables other researchers to explore and study stress detection. Akbas [1] presented an evaluation based on the driver dataset of PHYSIONET [14]. In his work, only 10 groups of those total 16 recordings were used. The remaining six groups are not evaluated because of the incompleteness of the sensor information. Zhang et al [24] gave a systematic approach using a Bayesian Network to combine sensor features.

![Figure 2. Placement of sensors](image2)

Combinatorial Fusion ([9], [10], [11]) provides a useful information fusion method and metric in analysing the combination and fusion of multiple scoring systems (MSS). It has been used in many application domains such as video target tracking [14], the virtual screening of molecular compound libraries [23], protein structure prediction [13] and on-line learning algorithms [15].

Let $D=\{d_1, d_2, ..., d_n\}$ be a set of candidates, such as sensor data, genes, documents, images, locations or classes. Let $A$ be a scoring system with a score function from the set $D$ to the set of real number $R$. In the framework of combinatorial fusion, each scoring system $A$ consists of two functions: a score function $s_A$ and a rank function $r_A$ derived from $s_A$ by sorting the function values of $s_A$ in descending order. A Rank-Score Characteristic (RSC) function is defined as: $f_A: N \rightarrow R[11]$.

![Figure 3. Rank-Score Characteristic (RSC) function](image3)

For a set of $p$ scoring systems $A_1, A_2, ..., A_p$ on the set $D$, at least two different approaches can be used to combine them: Score Combination (SC) and Rank Combination (RC). The equations are as follows:

$$s_{sc}(d_i) = \sum_{j=1}^{p} s_{A_j}(d_i) / p \quad (1)$$

$$s_{rc}(d_i) = \sum_{j=1}^{p} r_{A_j}(d_i) / p \quad (2)$$

where $d_i$ is in $D$, and $s_A$ and $r_A$ are the score function and rank function from $D$ to $R$ and $N$ respectively.
For a pair of two scoring systems A and B, the diversity between A and B, \( d(A,B) \), can be defined as \( d(s_A,s_B) \), \( d(r_A,r_B) \) and \( d(f_A,f_B) \) using score functions, rank functions, and rank-score characteristics (RSC) functions respectively. We use the last one in our paper and call this cognitive diversity.

3. Sensor feature selection using combinatorial fusion

3.1 Feature extraction

We use the same feature extraction results as Healey [7,8] and extract the 22 features from the driver stress dataset. Table 1 presents a detailed description. Upon further investigation into the 10 available driver datasets of the 17 driver datasets in PHYSIONET [16], we found that of these ten groups, seven drivers’ data sets (drivers 6, 7, 8, 10, 11, 12 and 15) are complete. These not only include all the sensor information but also have a clear mark identification. Three drivers’ sets (drivers 5, 9, and 16) are partially complete but can be used in the experiment. Driver05’s first highway period lacks heart rate information. Driver09’s second city period is less than five minutes and last rest period lacks a clear mark. Driver16’s second city period and last rest period are both less than five minutes. The remaining seven drivers’ data sets (drivers 1, 2, 3, 4, 13, 14 and 17) do not contain all the sensor information and the mark of the different driving period is not clear. Based on the complete portion of the sensor information, we acquired 65 segments with 22 features for each segment.

| Category                | Number | Symbol | Name                      | Feature Description                             |
|-------------------------|--------|--------|---------------------------|-------------------------------------------------|
| EMG Feature             | 1      | A      | EMG_mean                  | The normalized mean of the EMG data              |
|                         |        | B      | FGSR_mean                 | The normalized mean of the foot GSR data         |
|                         |        | C      | FGSR_std                  | The standard deviation of foot GSR data          |
|                         |        | D      | FGSR_freq                 | The total number of orienting responses of a segment for foot GSR |
|                         |        | E      | FGSR_mag                  | The summary of the startle magnitudes of orienting responses of a segment for foot GSR |
|                         |        | F      | FGSR_dur                  | The summary of the duration of orienting responses of a segment for foot GSR |
| Skin Conductivity Feature | 12    | G      | HGSR_area                 | The summary of the area of orienting responses of a segment for foot GSR |
|                         |        | H      | HGSR_mean                 | The normalized mean of the hand GSR data         |
|                         |        | I      | HGSR_std                  | The standard deviation of hand GSR data          |
|                         |        | J      | HGSR_freq                 | The total number of orienting responses of a segment for hand GSR |
|                         |        | K      | HGSR_mag                  | The summary of the startle magnitudes of orienting responses of a segment for hand GSR |
|                         |        | L      | HGSR_dur                  | The summary of the duration of orienting responses of a segment for hand GSR |
|                         |        | M      | HGSR_area                 | The summary of the area of orienting responses of a segment for hand GSR |
| Respiration Feature     | 6      | N      | RESP_mean                 | The normalized mean of the Respiration data      |
|                         |        | O      | RESP_std                  | The standard deviation of Respiration data       |
|                         |        | P      | RESP0-0.1                 | The summary of respiration energy in the bands 0-0.1Hz |
|                         |        | Q      | RESP0.1-0.2               | The summary of respiration energy in the bands 0.1-0.2Hz |
|                         |        | R      | RESP0.2-0.3               | The summary of respiration energy in the bands 0.2-0.3Hz |
|                         |        | S      | RESP0.3-0.4               | The summary of respiration energy in the bands 0.3-0.4Hz |
| Heart Rate Feature      | 3      | T      | HR_mean                   | The normalized mean of the heart rate data       |
|                         |        | U      | HR_std                    | The standard deviation of heart rate data        |
|                         |        | V      | HR_Ir                     | The total energy of Heart Rate in the low frequency band (0-0.08 Hz) |

Table 1. Description of the 22 extracted features

3.2 Feature selection using performance and cognitive diversity

Every feature extraction method is actually a score assignment metric. So every feature generation system can be regarded as a scoring system. We can use the performance as well as the diversity of multiple scoring systems to select the most important features, which can result in better performance when combined. The general principles are: (a) relatively good performance features can often result in better combination performance, and (b) those features with higher diversity can often result in better combination performance. Therefore, the aim of our feature selection is to find features with relatively good performance as well as relatively high diversity.

3.2.1 Performance sorted for single feature

We assume that the performance of a feature is in accordance or discordance with the increment of the stress level. Since stress has three different levels: low, middle and high, the values for a feature can be divided into three groups: low, middle and high. For the increase-increase case, the group with low values corresponds with low-level stress, the group with middle values
corresponds with middle-level stress and the group with high values corresponds with high-level stress. For the decrease-increase case, the result is simply inversely.

In our experiment, there are 22 features in total and 65 possible values for each feature. For each feature we sort their values in ascending order as well as descending order. The first 18 values are regarded as low level stress, the second 19 values are predicted to correspond with middle level stress and the last 28 values are assumed to belong to the high level stress. We compare the prediction results with the standard answers and calculate the correct rate with which to evaluate the performance. The feature with a higher correct rate between increase-decrease case and decrease-increase case means that the feature value increases or decreases with the increase of the stress level. We then select the higher correct rate as the final performance result. Our experimental results show that, features E, F, G, J and V are decrease-increase case and the other features are cases of increase-increase.

Figure 4 shows the sorted final performance of each feature in decreasing order. We can see that feature F has the highest correct rate with a value of 76.92% and feature U has the lowest correct rate with a value of 33.85%.

3.2.2 Cognitive diversity between features

The diversity of a feature pair \(\{F_i, F_j\}\) can be calculated as in Equation 3. \(f_i\) and \(f_j\) are the rank-score characteristic function for \(F_i\) and \(F_j\) respectively. Both functions \(f_i\) and \(f_j\) have total n score values and a rank sequence with range from 1 to n.

\[
d(F_i, F_j) = \frac{\sum_{i=1}^{n} |f_i(i) - f_j(i)|}{n} \quad (3)
\]

The diversity of a feature set \(S=\{F_1, F_2, ..., F_n\}\) is calculated as in Equation 4. \(C(n,2)\) is the combination of n features taking two at a time in feature set S. \(|C(n,2)|\) represents the total number of combining of two features in S.

\[
d(S) = \sum_{i<j} d(F_i, F_j) / |C(n,2)| \quad (4)
\]

The diversity between two feature sets \(S_1\) and \(S_2\) is calculated as in Equation 5. \(|S_1\) and \(|S_2\) are the cardinal number of \(S_1\) and \(S_2\).

\[
d(S_1, S_2) = \sum_{i<j} d(F_i, F_j) / |S_1 \cap S_2| \quad (5)
\]

Figure 5 shows the sorted pair diversity for the 231 feature pairs of 22 features in our experiment in decreasing order.

3.2.3 Feature selection algorithm (FSA)

Our algorithm for feature selection based on combinational fusion is as follows:

1. Performance analysis;
   a) calculate the performance of each feature,
   b) sort feature performance in decreasing order;

2. Diversity analysis;
   a) divide the total features into different groups based on the different sensor types, 
   b) calculate the average performance and average diversity for each feature group,
   c) calculate the inter diversity between each pair of two features from different groups;

3. Selection based on performance and diversity;
   a) select a feature set with \(m = 5, 7, 9, 11\) in our experiment) features with both high performance and high diversity;
   b) repeat step 3(a) and generate \(p (p = 4\) for each set in our experiment) different feature groups to carry out further feature combination and evaluation;

Figure 6 presents the modalities, performance and diversities of features. We divided the 22 features into four modalities according to the corresponding sensor types. These four modalities are: 1. EMG Features; 2. RESP Features; 3. GSR Features; 4. HR features. The features in each of the four modalities are in decreasing order according to performance. Bold font denotes the top 13 features. The number on the link between the two nodes is the diversity of these two feature sets. The number on the curve within a node is the average diversity of a feature set. It is assumed that, when the diversity between two modalities is bigger, it is more likely to generate a better combination performance. What’s more, the features with better individual performance are more likely to result in a better combination performance.
Figure 6. Feature classification using diversity and performance in four modalities

3.3 Feature selection results

Based on Figure 6, we selected features from the top-13-performance features and the diversity of the selected feature set is as high as possible. Using this feature selection method, we selected a total of four 5-feature sets (A), four 7-feature sets (C), four 9-feature sets (D) and four 11-feature sets (E). In order to perform comparisons, we also randomly select four 5-feature sets (B). The selected results are shown in Table 2.

| Feature Group | Feature Symbols | Average Performance | Average Diversity |
|---------------|-----------------|---------------------|-------------------|
| (A) 5-feature set | A(5,1) A,E,F,O,T | 0.66 | 0.255 |
| | A(5,2) D,E,O,P,T | 0.63 | 0.179 |
| | A(5,5) D,E,L,Q,T | 0.64 | 0.217 |
| | A(5,4) A,E,L,P,T | 0.62 | 0.248 |
| (B) 5-feature set (randomly selected) | B(5,1) A,F,H,T,O | 0.60 | 0.295 |
| | B(5,2) E,K,O,P,U | 0.56 | 0.258 |
| | B(5,3) D,E,N,S,T | 0.60 | 0.272 |
| | B(5,4) B,I,N,R,V | 0.43 | 0.400 |
| (C) 7-feature set | C(7,1) A,D,E,F,O,P,T | 0.64 | 0.229 |
| | C(7,2) A,D,E,L,Q,O,T | 0.63 | 0.221 |
| | C(7,3) A,E,F,L,O,P,T | 0.64 | 0.257 |
| | C(7,4) D,E,F,G,L,M,T | 0.68 | 0.207 |
| (D) 9-feature set | D(9,1) A,D,E,F,G,L,O,P,T | 0.64 | 0.230 |
| | D(9,2) A,D,E,F,L,O,P,O,T | 0.63 | 0.235 |
| | D(9,3) A,D,E,F,L,O,P,T | 0.63 | 0.232 |
| | D(9,4) D,E,F,G,K,L,M,O,T | 0.67 | 0.209 |
| (E) 11-feature set | E(11,1) A,D,E,F,G,L,M,O,P,Q,T | 0.63 | 0.227 |
| | E(11,2) A,D,E,F,T,K,L,M,O,P,Q | 0.63 | 0.227 |
| | E(11,3) A,D,E,F,G,J,K,L,O,P,T | 0.63 | 0.217 |
| | E(11,4) A,D,E,F,G,J,K,L,M,O,T | 0.65 | 0.212 |

Table 2. Feature selection results

4. Feature combination using combinatorial fusion

4.1 Leave-one-out based on combinatorial fusion

We use a leave-one-out metric based on combinational fusion to evaluate feature selection results. There are 65 groups of data in total and each group contains five feature values. In each round of evaluation, one group of data is selected as a test case and the other 64 groups of data are used as training cases. The test case can then be predicted based on the trained data. The above steps are repeated 65 times and each time the test case is unique. In this way, we can obtain 65 predicted results. These predicted results are compared with the standard answers and the correct rate is calculated. First, every feature is regarded as a group with 65 scores. For features N, C, P, S and V, we negate their scores. Normalizing the scores for each feature would ensure that the score value is between “0” and “1”. Sorting the scores in decreasing order implies that every feature has both 65 score values and 65 rank values. Next, the features from the feature set are selected and score combination and rank combination are performed individually according to Equation 1 and Equation 2. Finally, the score combination is sorted in decreasing order and the rank combination in increasing order respectively.

In the testing procedure, score combination and rank combination are also calculated. Then we compare the testing combination results with the training case combination results. For either score combination or rank combination, if the testing case is within the top 28 sequences, then the stress level is high. If testing case is between the 29th and 47th sequences, the stress level is medium. Finally if testing case is within the last 18 sequences, then the stress level is low.
4.2 Feature fusion result

4.2.1 Fusion results of 5-feature sets

The feature fusion correct rate of both score combination and rank combination for the 5-feature sets in both (A) and (B) in Table 2 is presented in Figure 7 and Figure 8.

Figure 7. Combinatorial fusion of 5-feature sets in (A)

(a). A(5,1)
(b). A(5,2)
(c). A(5,3)
(d). A(5,4)

Figure 8. Combinatorial fusion of 5-feature sets in (B)

(a). B(5,1)
(b). B(5,2)
(c). B(5,3)
(d). B(5,4)

In Figure 7(a), the maximum correct rate is 83.08% resulting from the score combination of features F and T. In Figure 7(b), the maximum correct rate is 86.15% resulting from the rank combination of features E, D, O and P. In Figure 7(c), the maximum correct rate is 87.69% resulting from the rank combination of features Q, D, L, T...
and E. In Figure 7(d), both the rank combination and core combination of features E, T, L, A, and P can result in the highest correct rate 86.15%. In Figure 8(a), the maximum correct rate is 83.08% resulting from the rank combination of features F and T. In Figure 8(b), the maximum correct rate comes from the score combination of features T and D. In Figure 8(d), the maximum correct rate is 58.46% resulting from the rank combination of features R and I. Figure 7 (a) – Figure 7(d) and Figure 8(a) – Figure 8(d) belongs to 5-feature sets in (A) and in (B) respectively (See Table 2).

The overall performance of the 5-feature sets in (A) is much better than that of the 5-feature sets in (B).

4.2.2 Fusion results of t-feature sets, t=7, 9, 11

Tables 3(a) to 3(c) show the combination results of the 7-feature sets, 9-feature sets and 11-feature sets respectively. Since the total combination is a large number, we only list the results of the highest combination performance of each feature group and its corresponding features.

| Feature Set | C(7,1) | C(7,2) | C(7,3) | C(7,4) |
|-------------|--------|--------|--------|--------|
| Highest Performance | 86.15 | 87.69 | 86.15 | 83.08 |
| Combined Features | rank combination: E,D,O,P | rank combination: E,T,D,L,Q | rank combination: E,T,L,A,P | score combination: E,T |

(a). Combination result of 7-feature sets

| Feature Set | D(9,1) | D(9,2) | D(9,3) | D(9,4) |
|-------------|--------|--------|--------|--------|
| Highest Performance | 86.15 | 87.69 | 86.15 | 84.62 |
| Combined Features | score combination: E,D,O,P; E,T,L,A,P; T,D,L,G,A,P | rank combination: E,T,D,L,Q | score combination: E,T,D,L,A,P; E,T,D,L,A,P; E,D,O,P; E,T,L,A,P | rank combination: E,T,D,L,O,K |

(b). Combination result of 9-feature sets

| Feature Set | E(11,1) | E(11,2) | E(11,3) | E(11,4) |
|-------------|---------|---------|---------|---------|
| Highest Performance | 87.69 | 87.69 | 87.69 | 86.15 |
| Combined Features | rank combination: E,T,D,L,Q; F,E,T,D,M,A,Q; F,E,T,D,M,A,P; F,E,T,D,L,M,A,Q,P | rank combination: E,T,D,L,Q; E,T,D,K,A,Q; F,E,T,D,M,A,Q; F,E,T,D,M,A,P; F,E,T,D,K,A,Q; F,E,T,D,K,A,P; F,E,T,D,L,M,A,Q,P | rank combination: E,T,D,K,A,P | score combination: F,E,T,D,K,A,P; F,E,T,D,G,A,K; E,T,D,L,O,K,A; F,E,T,D,L,G,M,O,J,A |

(c). Combination result of 11-feature sets

Table 3. Combination result of t-feature sets (t=7, 9, 11)

5. Fusion results comparison

5.1 Other feature fusion methods

In order to evaluate our feature fusion methods, we use another five methods as the feature fusion algorithms: LDF (Linear Discriminant Function), Decision Tree C4.5, SVM (Support Vector Machine), NB (Naïve Bayes) and KNN (K-Nearest Neighbours).

- In Healey’s work [7], a linear discriminant function (Equation 6) was used to classify the stress levels.

\[
 g_c(\hat{y}) = 2m_c^{-1}K^{-1}\hat{y} - m_c^{-1}K^{-1}m_c + 2\ln(Pr[W_c])
\]  

(6)

The stress class is assumed to have a Gaussian distribution, with \(m_c\) as the mean. The covariance \(K\) is the pooled covariance. A linear classifier is implemented by assigning each test sample to the class \(c\) for which the value of the function is the maximum. \(Pr[W_c]\) is the priori probability of belonging to class \(c\). \(Pr[W_i]=1/n_c\) and \(n_c\) is the numbers in class \(c\).
• C4.5 algorithm, is one of the most popular and practical methods for inductive inference [17,18], which uses information entropy as the metric to evaluate performance and uses the information acquired to select the nodes of the tree.

• Pioneered by Vapnik [22], Support Vector Machine (SVM) is a statistical learning algorithm, the basic idea of which is to find an optimal hyper-plane that can maximize the margin between two groups of samples. The vectors nearest to the optimal hyper-plane are called support vectors.

• A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions [25].

• KNN is an instance-based learning, where the function is only approximated locally and all computation is deferred until after classification [20].

5.2 Comparison of results

Table 4 shows the comparison of predictions of correct rates for different feature sets selected by using different feature selection metrics as well as different feature fusion methods. We use the maximum correct rate value of all the combination results for that feature set as the correct rate of combinational fusion used in this table.

From Table 4, we can see that, for feature sets selected by using our method, the combinatorial fusion can result in better performance than the other five fusion algorithms. CFA generates a higher value of both the maximum correct rate and the average correct rate for the 5-feature set (A), 7-feature set (C), 9-feature set (D) and 11-feature set (E). Over all, CFA can result in a better performance than the other five fusion methods.

From Table 4 we can also see that when using combinational fusion, the maximum correct rate is 87.69% for the 5-feature set, 87.67% for 7-feature set, 87.69% for 9-feature set and 87.69% for 11-feature set. The correct rate has not increased with the increase of the feature number. The 5-feature set selected by our feature selection method can result in the same highest performance as the 11-feature set.

| Feature Set | Fusion Method | C4.5 | NB | LDF | SVM | KNN | CFA |
|-------------|---------------|------|-----|-----|-----|-----|-----|
| **5-feature set** | | | | | | | |
| (A) | | | | | | | |
| A(5,1) | 80 | 81.54 | 73.85 | 69.23 | 76.92 | 83.08 | |
| A(5,2) | 81.54 | 81.54 | 69.23 | 64.62 | 73.85 | 86.15 | |
| A(5,3) | 63.08 | 78.46 | 61.54 | 67.69 | 61.54 | 87.69 | |
| A(5,4) | 56.92 | 70.76 | 64.62 | 60 | 66.15 | 86.15 | |
| Average | 70.39 | **78.08** | 67.31 | 65.39 | 69.62 | 85.77 | |
| (B) | | | | | | | |
| B(5,1) | 81.54 | 83.08 | 73.85 | 73.85 | 72.31 | 83.08 | |
| B(5,2) | 86.15 | 80 | 75.38 | 72.31 | 72.31 | 76.92 | |
| B(5,3) | 53.85 | 69.23 | 66.15 | 61.54 | 69.23 | 80 | |
| B(5,4) | 46.15 | 41.54 | 43.08 | 41.54 | 40 | 58.46 | |
| Average | 66.92 | **68.46** | 64.62 | 62.31 | 63.46 | 74.62 | |
| (C) | | | | | | | |
| C(7,1) | 75.38 | 83.08 | 75.38 | 70.77 | 73.85 | 86.15 | |
| C(7,2) | 81.54 | 76.92 | 63.08 | 70.77 | 70.77 | **87.69** | |
| C(7,3) | 76.92 | 80 | 70.77 | 70.77 | 73.85 | 86.15 | |
| C(7,4) | 67.69 | 72.31 | 67.69 | 80 | 72.31 | 83.08 | |
| Average | 75.38 | **78.08** | 69.23 | 73.08 | 72.70 | 85.77 | |
| (D) | | | | | | | |
| D(9,1) | 72.31 | 80 | 67.69 | 72.31 | 70.77 | 86.15 | |
| D(9,2) | 73.85 | 80 | 60 | 72.31 | 72.31 | 87.69 | |
| D(9,3) | 73.85 | 78.46 | 61.54 | 73.85 | 73.85 | 86.15 | |
| D(9,4) | 76.92 | 73.85 | 66.15 | 80 | 78.46 | 84.62 | |
| Average | 74.23 | **78.08** | 63.95 | 74.62 | 73.85 | 86.15 | |
| (E) | | | | | | | |
| E(11,1) | 72.31 | 78.46 | 67.69 | 75.38 | 73.85 | **87.69** | |
| E(11,2) | 75.38 | 78.46 | 64.62 | 76.92 | 73.85 | **87.69** | |
| E(11,3) | 76.92 | 76.92 | 69.23 | 69.23 | 75.38 | **87.69** | |
| E(11,4) | 76.92 | 75.38 | 69.23 | 75.38 | 75.38 | 86.15 | |
| Average | 75.38 | **77.31** | 67.69 | 74.23 | 74.62 | **87.31** | |

Table 4. Comparison of prediction performance (%)
For the randomly selected 5-feature set (B), the best performance is 86.15%, which is generated using the C4.5 method on feature set B(5,2). The highest performance using combinatorial fusion is 83.08% on feature set B(3,1). However, the average performance of combinatorial fusion method is 74.62% which is lower than 86.15% but better than 68.46%, the highest (B) of the five fusion methods. So, overall, combinatorial fusion is better than the other five fusion methods for the four randomly selected 5-feature set (B).

6. Conclusion and future work

In this paper, we demonstrated how to use combinatorial fusion to select features and fuse physiological sensor information to determine drivers’ stress levels. Our results showed that combinatorial fusion provides a good method with which to fuse sensor information. The correct rate can even achieve a much higher point when we use the features selected by our feature selection metric based on both performance and diversity.

The main contributions of this paper are: (1) we proposed an individual feature selection method based on both performance and diversity, (2) we used the combinatorial fusion method to fuse physiological sensor information to detect drivers’ stress, and (3) we performed a comparison of the combinatorial fusion method with five machine learning methods. Our work showed that the combinatorial fusion can result in better correct rates in several cases.

In the future, we will study decision level fusion based on combinatorial fusion. In addition, other kinds of feature selection and feature fusion methods will be investigated.

7. Acknowledgments

The work of Yong Deng was performed during his visiting year 2011-2012 at School of Information Science and Technology, Pennsylvania State University. This research is supported by the State Key Program of National Natural Science of China (grant no. 61232005) and the National Key Technology R&D Program (no: 2012BAH06B01).

8. References

[1] Akbas, A. (2011). Evaluation of the physiological data indicating the dynamic stress level of drivers. Scientific Research and Essays, 6(2), pp. 430-439.
[2] APA (American Psychological Association) (2012). Stress in America: Our Health at Risk. URL: http://www.apa.org/news/press/releases/stress/index.aspx
[3] Angus, F., and Zhai, J. (2005). Front-end analog preprocessing for real time psychophysiological stress measurements. Proceedings of the 9th World Multi-Conference on Systematics, Cybernetics and Informatics(WMSCI05), pp. 218-221.
[4] Bakker, J., Pechenizkiy, M., and Sidorava, N. (2011). What’s your current stress level? Detection of stress patterns from GSR sensor data. Proceedings of the 11th IEEE International Conference on Data Mining Workshops, pp. 573-580.
[5] Deng, Y., Hsu, D. F., Wu, Z., Chu, C. (2012). Combining Multiple Sensor Features for Stress Detection using Combinatorial Fusion. Journal of Interconnection Networks, Vol 13, Issue 3n04, DOI: 10.1142/S0219265912500089.
[6] Deng, Y., Chu, C., Wu, Z., Si, H., Zhang, Q. (2012). An Investigation of Decision Analytic Methodologies for Stress Identification. The International Journal on Smart Sensing and Intelligent Systems (ISSN: 1178 -5608).
[7] Healey, J. A. (2000). Wearable and automotive systems for affect recognition from physiology. Doctoral dissertation, Massachusetts Institute of Technology, MA.
[8] Healey, J. A., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transaction on Intelligent Transportation System, 6(2), pp. 156-166.
[9] Hsu, D. F., and Taka, J. (2005). Comparing rank and score combination methods for data fusion in information retrieval. Information Retrieval, 8(3), pp. 449-480.
[10] Hsu, D. F., Chung, Y. S., and Kristal, B. S. (2006). Combinatorial fusion analysis: methods and practice of combining multiple scoring systems. Advanced Data Mining Technologies in Bioinformatics, Idea Group Inc., pp. 32-62.
[11] Hsu, D.F., Kristal, B.S., and Schweikert, C. (2010). Rank-score characteristic (RSC) function and cognitive diversity. Brain Informatics, LNAI 6334, Springer, pp. 42-54.
[12] Jovanov, E., O’Donnell Lords, A., Raskovic, D., Cox, P. G., Adhami, R., and Andrasik, F. (2003). Stress monitoring using a distributed wireless intelligent sensor system. IEEE Engineering in Medicine and Biology Magazine, 22, pp. 49-55.
[13] Lin, C. Y., Lin, K. L., Huang, C. D., Chang, H. M., Yang, C. Y., Lin, C. T., Tang, C. Y., and Hsu, D. F. (2007). Feature selection and combination criteria for improving accuracy in protein structure prediction. IEEE Trans. Nanobiosci. 6(2), pp. 186-196.
[14] Lyons, D. M., and Hsu, D. F. (2009). Combining multiple scoring systems for target tracking using rank-score characteristics. Information Fusion, Vol. 10, No. 2, 2009, pp. 124–136.
[15] Meterharm, C., and Hsu, D. F. (2008). Combinatorial fusion with on-line learning algorithms. The 11th International Conference on Information Fusion, pp. 1117-1124.
[16] PHYSIONET (2010). Stress Recognition in Automobile Drivers (drivedb). URL: http://physionet.org/cgi-bin/atm/ATM/.

www.intechopen.com Yong Deng, Zhonghai Wu, Chao-Hsien Chu, Qixun Zhang and D. Frank Hsu: Sensor Feature Selection and Combination for Stress Identification Using Combinatorial Fusion
[17] Polat, K., and Güneş, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications, 36 (2009) pp. 1587-1592.

[18] Ruggieri, S. (2002). Efficient C4.5, IEEE Transactions on Knowledge and Data Engineering, Vol. 14, No. 2, March/April 2002.

[19] Salahuddin, L., and Kim, D. (2006). Detection of acute stress by heart rate variability using a prototype mobile ECG Sensor," International Conference on Hybrid Information Technology, Proceeding in IEEE CS, vol. 2, pp. 453-459.

[20] Duda, R., Hart, P., and Stork, D. (2001). Pattern Classification, (2nd Ed.). Wiley Inter-science.

[21] Sun F.T., Kuo C., Cheng, H.T., Buthpitiya S., Collins P., and Griss, M. (2010). Activity-aware Mental Stress Detection Using Physiological Sensors. Silicon Valley Campus, pp. 23.

[22] Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY, USA. ISBN: 0-387-94559-8.

[23] Yang, J. M., Chen, Y. F., Shen, T. W., Kristal, B. S., and Hsu, D. F. (2005). Consensus scoring criteria for improving enrichment in virtual screen. Journal of Chemical Information and Modeling, Vol. 45, no.4, pp. 1134-1146.

[24] Zhang, L., Tamminedi, T., Ganguli, A., Yosiphon, G., and Yadegar, J. (2010). Hierarchical multiple sensor fusion using structurally learned Bayesian network. Proceedings of Wireless Health, pp. 174-183.

[25] Rish, I. (2001). An empirical study of the naive bayes classifier. In Proceedings of IJCAI-01 Workshop on Empirical Methods in AI, Sicily, Italy, pp. 41-46.