

Quality Assessment of Subjectively Labelled Training Data for Improving the Reliability of Status Perception

Ying Dai

Iwate Prefectural University, Japan

Abstract

In order to improve the reliability of the status perception, this paper defines a metric QoSTD to measure the quality of the subjectively labelled training data used with K-means clustering. On the basis of QoSTD, we propose a method that utilizes a support vector support (SVM) model to predict the specified states of an individual's status. We also present a way to determine a threshold for QoSTD so that it rejects states that cannot be predicted with sufficient certainty. A high positive correlation between the QoSTD and the quality of the perception of the status is verified by experimental results with predictions based on traditional Chinese medicine (TCM) Zheng's. In addition, we verified that we can use QoSTD to determine whether a change of training data for a feature's modality affects the quality of the status perception.

1. Introduction

We now have the technology to collect human data through sensors, and there is an increasing amount of personal information available through social networks. There is thus increasing interest in research in the areas of unstructured and uncertain data computing, such as affective computing, personality computing, and Kansei engineering [1] [2] [3]. Reference [1] presents the intellectual framework for affective computing. The design and construction of affective computing are discussed; it also includes the requirements for emotional computers and applications for affective computing. Reference [2] discusses Kansei engineering, which focuses on qualitative and quantitative evaluations, such as the measurement and analysis of the sense perceptions of groups and individuals, as well as the development of procedures for the design of products that use this information. Reference [3] discusses personality computing domains concerned with personality and considers three main problems: automatic personality recognition, perception, and synthesis.

All of these approaches are aimed at determining the status of individuals, including their emotional

state, their Kansei, and their personality; this is done by using a training data set that is unconstructive and is interpreted or labelled subjectively by a third person. Furthermore, such data sets may be incomplete. Thus, the collected training data are very person dependent. In order to assure the reliability of the results, it is important to assess the quality of the training data used by the model to determine the status of an individual. However, this issue was not explicitly addressed in the papers discussed above.

Focusing on the above issue, we defined a metric QoSTD that is intended to measure the quality of the subjectively labelled training data that are used to predict the specified status of an individual. QoSTD reflects the ability of the clustering and partitioning of the training data set to predict an individual's status. For the training data, the features include the eigenfeatures extracted from the multimodal sensor data of the target and the subjective scores of various items in a first person questionnaire. The labelled states data are the scores given by third persons. By using this metric, we can analyze the influence of multimodal sensor data on the quality of the training data, and we can estimate the sufficiency of the training data for predicting the status. When QoSTD for a particular state is less than a predetermined threshold, no prediction is made.

We also trained a well-known support vector machine (SVM) classifier to predict an individual's status. The average of SVM score and the classification loss were utilized to evaluate the performance of the prediction. To examine the effectiveness of the proposed method, we used TCM Zheng training data set that was used in [4] and [5]. The experimental results show that there is a definite threshold for the QoSTD, beyond which the performance of the prediction is dramatically improved. We thus verified that QoSTD can be used to assure the reliability of the state prediction, and determine whether the given sensors are sufficient and effective for doing so. Also, this shows that we can estimate the performance of the classification by analyzing the values of QoSTD even though the classifier isn't trained by the subjectively labelled training data

We now have the technology to collect human data through sensors, and there is an increasing amount of personal information available through social networks. There is thus increasing interest in research in the areas of unstructured and uncertain data computing, such as affective computing, personality computing, and Kansei engineering [1] [2] [3]. Reference [1] presents the intellectual framework for affective computing. The design and construction of affective computing are discussed; it also includes the requirements for emotional computers and applications for affective computing. Reference [2] discusses Kansei engineering, which focuses on qualitative and quantitative evaluations, such as the measurement and analysis of the sense perceptions of groups and individuals, as well as the development of procedures for the design of products that use this information. Reference [3] discusses personality computing domains concerned with personality and considers three main problems: automatic personality recognition, perception, and synthesis.

All of these approaches are aimed at determining the status of individuals, including their emotional state, their Kansei, and their personality; this is done by using a training data set that is unconstructive and is interpreted or labelled subjectively by a third person. Furthermore, such data sets may be incomplete. Thus, the collected training data are very person dependent. In order to assure the reliability of the results, it is important to assess the quality of the training data used by the model to determine the status of an individual. However, this issue was not explicitly addressed in the papers discussed above.

Focusing on the above issue, we defined a metric QoSTD that is intended to measure the quality of the subjectively labelled training data that are used to predict the specified status of an individual. QoSTD reflects the ability of the clustering and partitioning of the training data set to predict an individual's status. For the training data, the features include the eigenfeatures extracted from the multimodal sensor data of the target and the subjective scores of various items in a first person questionnaire. The labelled states data are the scores given by third persons. By using this metric, we can analyze the influence of multimodal sensor data on the quality of the training data, and we can estimate the sufficiency of the training data for predicting the status. When QoSTD for a particular state is less than a predetermined threshold, no prediction is made.

We also trained a well-known support vector machine (SVM) classifier to predict an individual's status. The average of SVM score and the classification loss were utilized to evaluate the performance of the prediction. To examine the effectiveness of the proposed method, we used TCM Zheng training data set that was used in [4] and [5]. The experimental results show that there is a definite

threshold for the QoSTD, beyond which the performance of the prediction is dramatically improved. We thus verified that QoSTD can be used to assure the reliability of the state prediction, and determine whether the given sensors are sufficient and effective for doing so. Also, this shows that we can estimate the performance of the classification by analyzing the values of QoSTD even though the classifier isn't trained by the subjectively labelled training data.

2. Related works

The literature contains many studies on data integration and subjective inference. Reference [9] introduces a framework called smart sensor integration that can be used for multimodal recognition of emotions. Reference [10] presents a distributed filter that allows the nodes of a sensor network to track the average sensor measurements when solving a data fusion problem. Reference [11] discusses the advantages of using image processing to integrate multiple and varied sensors, and it identifies a number of associated problems. Reference [12] analyzes a number of pieces of evidence supporting a single subjective hypothesis within a Bayesian framework. Reference [13] introduces an emotion-processing system that is based on fuzzy inference and subjective observation. However, these papers do not consider ways to measure the quality of the training data acquired from multimodal sensors, and this is necessary in order to assure the reliability and effectiveness of the subjective inferences.

Recently, due to its ambiguous nature, labelling the values of subjective properties for learning a prediction is challenging. In [14], to make the annotation more reliable, the proposed method integrates local pairwise comparison labels together to minimize a cost that corresponds to global inconsistency of ranking order. In [15], the authors construct subjective classification systems to predict sensation of reality from multimedia experiences based on EEG and peripheral physiological signals such as heart rate and respiration. In [16], the authors propose a machine learning based data fusion algorithm that can provide real time per frame training and decision based cooperative spectrum sensing. For the labelled data imbalance, the authors in [17] propose a framework based on the correlations generated between concepts. The general idea is to identify negative data instances which have certain positive correlations with data instances in the target concept to facilitate the classification task. In [18], robust principal component analysis and linear discriminant analysis are used to identify the features, and support vector machine (SVM) is applied to classify the tumor samples of gene expression data based on the identified features.

On the other hand, various methods have been proposed that use TCM to infer the health status of an individual as a means of auto-diagnosing. References [4] and [5] propose methods that use TCM Zheng to infer the health status of individuals by using images of their face and eyes, data on their emotional and physical state, and Zheng scores assigned by different TCM doctors (TCMDs). Reference [7] analyzes the effect of multimodal sensor data on the Zheng classification.

However, all of these papers don't consider to improve the reliability of the prediction by assessing the quality of the subjectively labelled training data.

3. Measuring the quality of subjectively labelled training data

Because the target to be determined is the status of a person, the collected data representing that person's properties are generally diverse. However, the data from the first-person questionnaires, the data measured by the sensor, and those measured by the equipment are all handled in the same way as the features of different modes. Although the modes are heterogeneous, all of the features belonging to the different modes are considered to be homogeneous. For example, the histogram, shape, and the texture of an individual image are the features of the image mode, and the blood pressure measured by a bio-sensor is the feature of bio-sensor mode. All of these features are considered to be homogeneous. They are denoted as $a_o^{m,n}$, which are normalized for each data set. Here, o , m , and n indicate indices of the sample, the mode, and the feature. The combined features yield a matrix A .

On the other hand, the states of each individual's status are labelled by third persons and are based on first-person questionnaires, on data measured by sensors, or by direct observation of the target. The labelled state scores for each individual's status are considered to be the subjectively labelled data set. These scores are denoted as z_s^{ij} , and the values range from 0 to 10. Here, s , i , and j indicate indices of the sample, the observer, and the state, respectively.

A matrix of eigenfeature vectors of the samples is obtained by calculating the eigenvalues and eigenvectors of $A^T * A$; this is based on the method of principle component analysis (PCA) [6]. It is denoted as

$$EA = \{ea_{s,p} : s \in [1, S], p \in [1, P]\} \quad (1)$$

where s and p are indices indicating sample and eigenfeature, respectively, and S and P are the numbers of samples and eigenfeatures, respectively. Thus, the size of EA is $S * P$.

The eigenfeature vectors are then used to represent the samples. The samples belonging to a given state and those not belonging to that state are considered to overlap due to the subjectivity of the

labelling. Accordingly, a matrix called OoSTD is defined to reflect how well the training data set can be divided into binary classes. This allows us to explore the influence of the features and the subjectively labelled data on the state that is perceived. OoSTD is based on two components (the partitioning and the clustering) and these determine the performance of using a binary classifier for the training data set. Let the score of State j for Sample s labelled by Observer i be denoted as z_s^{ij} , and let the threshold of that state's score be denoted as t_z (a variable). In the training data set, those that have scores larger than t_z for state j compose the data set $PZ^{ij} = \{z_s^{ij} : z_s^{ij} \geq t_z, s \leq S\}$, and those that have a score of 0 for State j compose the data set $NZ^{ij} = \{z_s^{ij} : z_s^{ij} = 0, s \leq S\}$. We used K-means clustering to divide the data set into two groups. If we assume that one of these groups is the positive cluster of State j , denoted as $PC^{ij} = \{pc_s^{ij}, s \leq S\}$, the other cluster is denoted as $NC^{ij} = \{nc_s^{ij}, s \leq S\}$. Accordingly, the performance of the separation of the data set for State j labelled by Observer i is defined as

$$par1^{ij} = \frac{\#(PZ^{ij} \cap PC^{ij})}{\#PC^{ij}}, \quad (2)$$

and the performance of the clustering of the data set for State j labelled by Observer i is defined as

$$clu1^{ij} = \frac{\#(PZ^{ij} \cap PC^{ij})}{\#PZ^{ij}}, \quad (3)$$

where $\#$ indicates the number of data points.

In this instance, the metric $QoSTD1^{ij}$ is defined by the following expression, which is an aggregation of $par1^{ij}$ and $clu1^{ij}$:

$$QoSTD1^{ij} = w_1 par1^{ij} + w_2 clu1^{ij}. \quad (4)$$

Here, w_1 and w_2 are both set to 0.5, because it is considered that $par1^{ij}$ and $clu1^{ij}$ have the equivalent weights to $QoSTD1^{ij}$.

Next, let the other group be the positive cluster. Accordingly, the performance of the separation of the data set for State j labelled by Observer i is defined as

$$par2^{ij} = \frac{\#(PZ^{ij} \cap NC^{ij})}{\#NC^{ij}}, \quad (5)$$

and the performance of the clustering of the data set for State j labelled by Observer i is defined as

$$clu2^{ij} = \frac{\#(PZ^{ij} \cap NC^{ij})}{\#NZ^{ij}}, \quad (6)$$

where $\#$ indicates the number of data points.

In this instance, the metric $QoSTD2^{ij}$ is defined by the following expression, which is an aggregation of $par2^{ij}$ and $clu2^{ij}$:

$$QoSTD2^{ij} = w_1 par2^{ij} + w_2 clu2^{ij}. \quad (7)$$

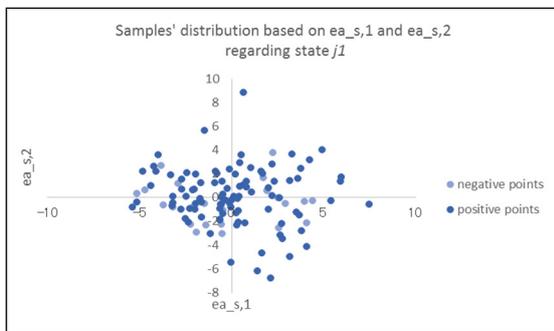
Here, w_1 and w_2 are both set to 0.5, because it is considered that $par2^{ij}$ and $clu2^{ij}$ have the equivalent weights to $QoSTD2^{ij}$.

Comparing the values of $QoSTD1^{ij}$ and $QoSTD2^{ij}$, the larger value is considered as the quality of the training data set used for classifying State j labelled by Observer i .

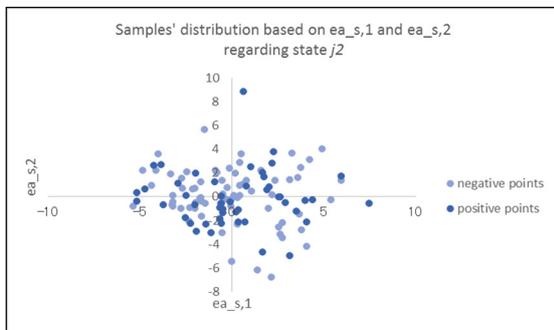
$$QoSTD^{i,j} = \max\{QoSTD1^{i,j}, QoSTD2^{i,j}\} \quad (8)$$

The value of $QoSTD^{ij}$ averaged over all of the states of a person's specified status is denoted as ave_QoSTD^i ; this will vary by the observer labelling the data set.

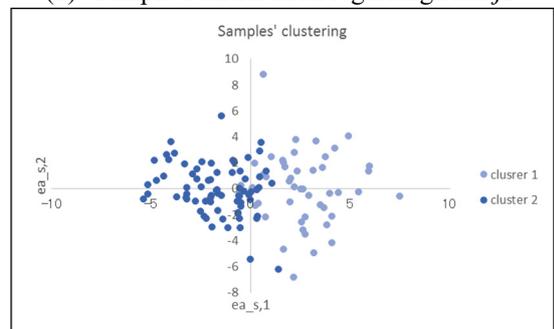
Fig. 1 shows the example of 120 data points of $ea_{s,1}$ and $ea_{s,2}$. Fig 1 (a) is the samples' distribution regarding state $j1$, and Fig.1 (b) is the samples' distribution regarding state $j2$. The dark blue points indicate the corresponding samples assigned to that state, and the light blue points indicate the samples not assigned to that. Fig. 1 (c) is the clustering of the data points by K-means. The light blue points are clustered into cluster 1, and the dark blue points are clustered into cluster 2.



(a) Samples' distribution regarding state $j1$



(b) Sample's distribution regarding state $j2$



(c) Samples' Clustering

Figure 1. Samples' distribution and clustering

From Fig. 1, we can see that the samples' distribution is overlapped although the degree of the overlapping is different for state $j1$ and state $j2$. Considering that the performance of the training data set regarding the separation and the clustering influences the ability of the classification, the quality of the instance of (a) is better than that in the instance of (b) when those data are used to train the classifier. In fact, the value of $QoSTD^{ij}$ regarding the instance (a) is 0.78, and the value of that regarding the instance (b) is 0.43. So, we think that the larger the values of $QoSTD^{ij}$ and ave_QoSTD^i are, the better the quality of the training data set labelled by Observer i for classifying State j is. When $QoSTD^{ij} = 1$, this indicates that the data set labelled by Observer i can be divided appropriately into two classes with a positive or negative State j .

4. Using QoSTD to predict a person's status

As mentioned above, the training data used to predict a person's status are ambiguous. It is necessary to determine whether a given training data set can be used to reliably perceive that person's status. Here, the metric $QoSTD^{ij}$ is used to judge the training data set. For State j of the person's status, the prediction is rejected if the values of $QoSTD^{ij}$ are less than a given threshold T_{rej} . The procedure for predicting the states by using the subjectively labelled training data set is as follows.

- Calculating $QoSTD^{ij}$

The data modes used as the training set are determined based on the context in which the data were collected and the capacity for computations. Next, the features of matrix A are extracted from the multimodal data set. The eigenfeatures matrix EA is obtained based on the eigenvalues and eigenvectors of $A' * A$. Then, the value of $QoSTD^{ij}$ for State j labelled by Observer i is calculated using equations (2), (3), and (4).

- Determining the states that should be rejected

If the value of $QoSTD^{ij}$ is less than T_{rej} , the prediction of State j using the training data set labelled by Observer i should be rejected. The value of T_{rej} is chosen so that the prediction can achieve an expected level of reliability.

Each time new data modes are chosen to extract the features, the value of $QoSTD^{ij}$ is recalculated. Thus, the context in which State j is rejected may vary for different values of $QoSTD^{ij}$.

- Building the classification model to predict the states

Generally, existing supervised machine learning classifiers, for example, k-nearest neighbor (KNN), support vector machine (SVM), or decision tree (DT), could be utilized to predict the states. We use the

SVM model to complete the tasks in this paper. The SVM model of State j is trained by the eigenfeatures-based training data set using the MATLAB (Mathworks, Natick, MA, USA) function *fitcsvm*. The kernel function here is a polynomial of order three.

The SVM score for a state j labelled by observer i is obtained by using the MATLAB function *predict*, and it is denoted as $score^{i,j}$. This for classifying sample x is the signed distance from x to the decision boundary ranging from $-\infty$ to $+\infty$. A positive score for the state j indicates that x is predicted to be that class. The larger the value of $score^{i,j}$ is, the more confidentially the x is assigned to the state j . The value of $score^{i,j}$ averaged over all of the samples with the positive scores for the state j is denoted as $ave_score^{i,j}$.

On the other hand, with k-fold cross-validation, the classification loss of state j labelled by observer i for samples not used for training is obtained by using the MATLAB function *crossval* and *kfoldLoss*, and it is denoted as $loss^{i,j}$. The less the value of $loss^{i,j}$ is, the more reliable the prediction results of the samples belonging to the state j is.

The measurements of $ave_score^{i,j}$ and $loss^{i,j}$ are used to show how the metric $QoSTD^{i,j}$ is utilized to improve the reliability of the status prediction in section 6.

5. Training data set used in the experiment

In this study, we used the same training data set that was used in previous studies [4] [5] for predicting the individual's health status represented by the states of TCM's thirteen Zhengs (寒, 熱, 氣虛, 氣滯, 血虛, 血瘀, 津虛, 痰飲, 心, 肺, 脾, 肝, 腎), because we think that it is sufficient to examine the proposed method. The other reason that the data set is not be updated is that it is not easy to let TCMD label Zheng scores again and again. This data set contains multimodal sensor data about the health status of various individuals. These data include scores of measured physical states and reports of subjective information obtained by first-person questionnaires; in addition, features are extracted from images of the individual's tongue, face, and eyes. The corresponding labelled data set comprises the scores of thirteen Zhengs given by four TCM doctors (TCMDs) who inspected and diagnosed the provided samples. The labelled Zheng scores range from 0 to 10. However, most of these data have values less than 5 because the subject volunteers were students at the university, and thus they were generally healthy. The data from the first-person questionnaires contains nine types of feelings and thirteen physical states related to health status, as proposed by the World Health Organization (WHO). The scores of the corresponding items range from 0 to 5. The features that were extracted from the images of the faces, eyes,

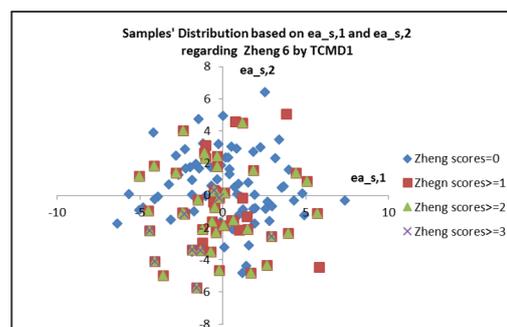
and tongues are explained in reference [4]. The extracted features were combined with the above feelings and physical states to create the features used to represent the health status of each person. Each of these items are the modes used to create the features. The training data set includes five modes: Feelings, Physical States, Eye, Tongue, and Face. The modes and the number of features for each mode are shown in Table 1. The total number of features is 71. There are 150 samples from 32 individuals in the data set, each of which includes 71 features and the corresponding thirteen Zheng scores labelled by the four TCMDs.

Table 1. Modes and the number of features

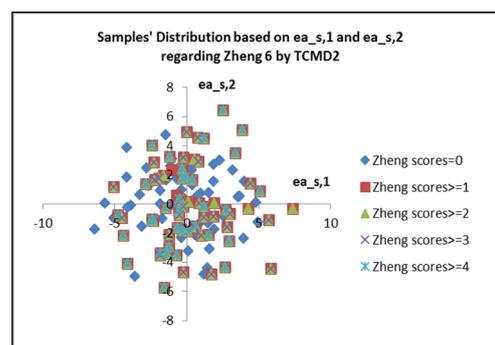
Symbol	M_1	M_2	M_3	M_4	M_5
Mode	Feeling	Physical States	Eye	Tongue	Face
Number of features	9	25	9	20	18

A matrix of eigenfeature vectors of the samples is obtained by calculating the eigenvalues and eigenvectors of above feature data sets; this is based on the method of principle component analysis (PCA) [6]. It is denoted as the expression (1).

The distribution of the samples, based on the first and second eigenfeatures for Zheng 6 (血瘀), are shown in Fig. 2. In Fig. 2 (a), (b), (c), and (d), the Zheng scores of the samples were from TCMD1, TCMD2, TCMD3, and TCMD4, respectively.



(a) Zheng scores from TCMD1



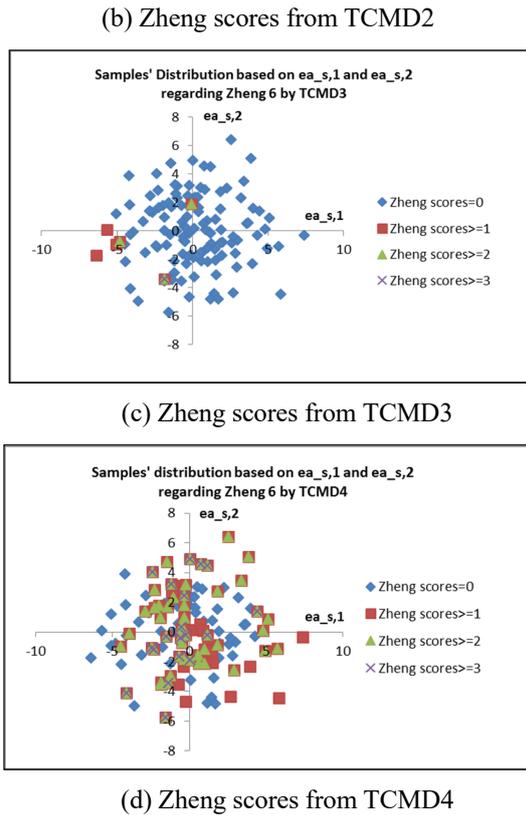


Fig. 2 Distribution of samples with Zheng scores of Zheng 6

In Fig. 2, the blue diamonds indicate samples with a Zheng score of zero; this means that the corresponding samples do not belong to Zheng 6. The remaining points indicate samples that belong to Zheng 6. The red squares indicate samples with Zheng scores larger than zero; the green triangles indicate samples with Zheng scores greater or equal to 2; and the gray crosses indicate those having the Zheng scores greater or equal to 3. In Fig. 2, it can be seen that the samples diagnosed as belonging to Zheng 6 overlap with those that do not belong, and there is no obvious boundary between the two classes. The degree of overlap varies by TCMD.

It is clear that the subjective score labelling of the unconstructive data is often imprecise.

6. Experimental results and analysis

- $QoSTD^{i,j}$

The results of $QoSTD^{i,6}$ for the training data set are shown in Fig. 3. From this, we can evaluate the quality of the training data set labelled by four different TCMDs for the binary classification of Zheng 6; this is shown for three different values of t_z (1, 2, and 3).

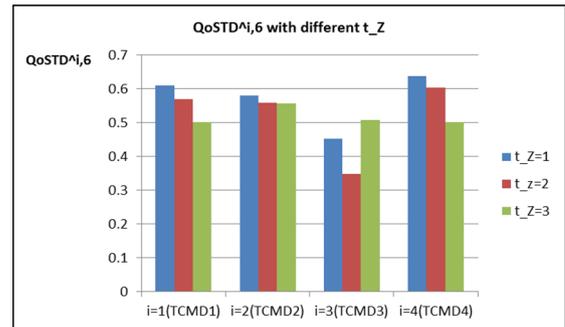


Figure 3. Quality of training data set for Zheng 6

From Fig. 3, it can be seen that the values of $QoSTD^{1,6}$, $QoSTD^{2,6}$, and $QoSTD^{4,6}$, which correspond to the quality of the training data sets labelled by TCMD1, TCMD2, and TCMD4, respectively, are relatively high. In particular, when $t_z = 1$, the values for TCMD1 and TCMD4 are larger than 0.6. Furthermore, in most cases, the value of $QoSTD^{i,6}$ decreases when t_z increases. That is, when the labelled state scores have low values, increasing the threshold does not improve the quality of the training data set for the classification. This is in accordance with the observations we made about the distribution of the training data set for Zheng 6, as shown in Fig. 2. The other $QoSTD^{ij}$ values show a similar behavior. Thus, $QoSTD^{ij}$ actually reflects the quality of the subjective training data set labelled by TCMD i for classifying State j .

Fig. 4 shows the values of $QoSTD^{ij}$ for all thirteen Zhengs, as labelled by four TCMDs, for which $t_z = 1$.

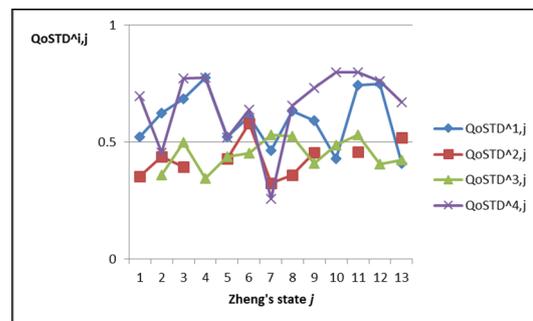


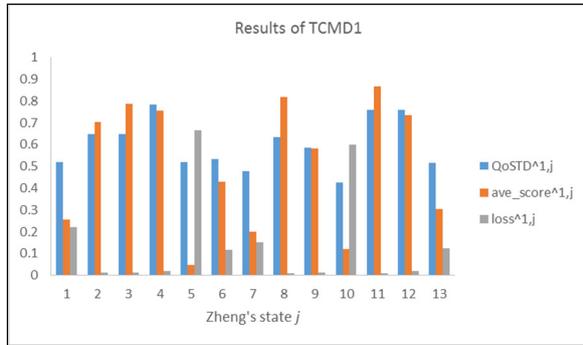
Figure 4. Quality of the training data set for all thirteen Zhengs

From Fig. 4, we can see that the quality of the training data sets labelled by TCMD1 and TCMD4 have relatively high values for most of the Zhengs. For TCMD1, eight of the Zhengs have a $QoSTD^{ij}$ value larger than 0.6; for TCMD4, ten of them have a $QoSTD^{ij}$ value larger than 0.6. However, for the data sets labelled by TCMD2 and TCMD3, the values of $QoSTD^{ij}$ are relatively low. Most of Zhengs have a $QoSTD^{ij}$ value less than 0.6. It is observed that the quality of Zheng scores labelled by TCMD2 and TCMD3 are not as good as those labelled by TCMD1 and TCMD4. We thus think that

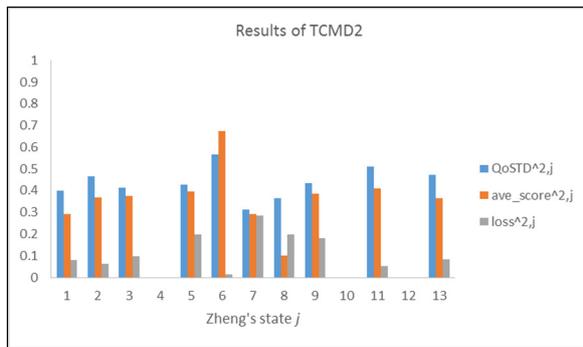
the $QoSTD^{i,j}$ can be used as a criterion for judging the quality of the subjectively labelled training data sets. If the $QoSTD^{i,j}$ is less than a threshold, it is necessary to reject the prediction of state j using the training data labelled by TCMD i , so as to ensure the reliability of the state prediction. Although a larger $QoSTD^{i,j}$ value indicates that the training data set is of better quality, a higher threshold will result in a higher rate of rejecting the states to be predicted. Thus, the value of T_{rej} for $QoSTD^{i,j}$ is trade-off. It could be set to 0.6 in the instance of Fig. 4. Of course, in practice, this threshold can be adjusted automatically.

- Exploring the correlation among $QoSTD^{i,j}$, $ave_score^{i,j}$ and $loss^{i,j}$

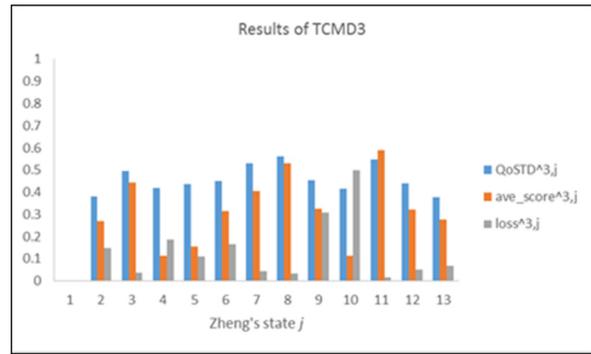
The values of $QoSTD^{i,j}$, $ave_score^{i,j}$ and $loss^{i,j}$ for the thirteen Zhengs, based on the labelled scores of four TCMDs, are shown in Fig. 5; $t_z = 1$. All features of the five modes were used for extracting the eigenfeatures. In this case, $i \in [1,4]$, and $j \in [1,13]$.



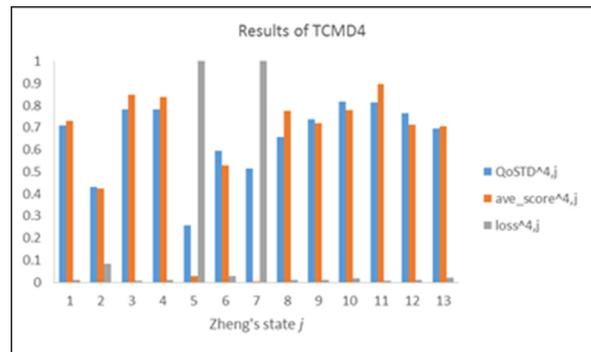
(a) Labelled by TCMD1



(b) Labelled by TCMD2



(c) Labelled by TCMD3



(d) Labelled by TCMD4

Figure 5. Values of $QoSTD^{i,j}$, $ave_score^{i,j}$ and $loss^{i,j}$

In Fig. 5, the blue bars indicates the values of $QoSTD^{i,j}$, the orange bars indicates the values of $ave_score^{i,j}$ normalized into the range of $[0, 1]$, and gray bars indicate the values of $loss^{i,j}$. From Figs. 5, it is obvious that, in almost all cases, the values of $ave_score^{i,j}$ are large, and the values of $loss^{i,j}$ is small when the corresponding values of $QoSTD^{i,j}$ are large. The correlation coefficients between $QoSTD^{i,j}$ and $loss^{i,j}$ for thirteen Zhengs labelled by four TCMDs are 0.90, 0.91, 0.63 and 0.89, respectively, and those between $QoSTD^{i,j}$ and $loss^{i,j}$ are -0.66, -0.75, -0.42 and -0.75, respectively. This indicates that $ave_score^{i,j}$ has a high positive correlation with $QoSTD^{i,j}$, and $loss^{i,j}$ has a high negative correlation with $QoSTD^{i,j}$, although the absolute values of correlation coefficients for TCMD3 is relatively small. In particular, the values of $ave_score^{i,j}$ is dramatically increased, and the values of $loss^{i,j}$ is dramatically decreased, when the values of $QoSTD^{i,j}$ is larger than 0.6. Accordingly, it is verified that setting the value 0.6 as a threshold is available. If the value of $QoSTD^{i,j}$ is less than 0.6, the prediction of State j using the training data set labelled by TCMD i should be rejected. Thus, we can deduce that $QoSTD^{i,j}$ can be used as a criterion to judge whether the quality of the training data set meets the requirements necessary to predict the given state. Furthermore, this shows that we can estimate the performance of the classification by analyzing

the values of $QoSTD^{i,j}$ even though the classifier isn't trained by the subjectively labelled training data.

On the other hand, we note that the values of $QoSTD^{i,j}$ differs by TCMD. For TCMD1 and TCMD4, the values is comparatively high. This implies $QoSTD^{i,j}$ can be used to measure the ability of a TCMD to use empirical knowledge to create consistent labels, although the data labelled by them are generally ambiguous.

- Examining the influence of eliminating a single mode on the perception of status

In practice, there is the possibility that data cannot be acquired for some modes; this might occur if some sensors did not function properly or were not used. Thus, we used $QoSTD^{i,j}$ to explore how the loss of a mode would influence the perception of the status for a given level of reliability. In this paper, the training data contained five modes, as shown in Table 1. The $QoSTD^{i,j}$ values based on the thirteen Zheng's scores, as labelled by TCMD1, and with one mode eliminated, are shown in Fig. 6; $t_z = 1$. Note that, here, the features of each of the five modes were eliminated in turn when extracting the eigenfeatures; these are compared to those in which all modes were used.

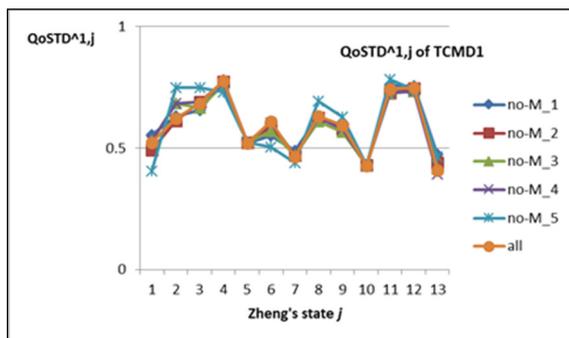


Figure 6. Values of $QoSTD^{i,j}$ when a single mode is eliminated

From Fig. 6, we can see that most of the $QoSTD^{i,j}$ values were not changed, with the case of eliminating mode 5 (Face) being the exception. In that case, some values of $QoSTD^{i,j}$ increased and others decreased. Thus, the features of mode 5 are sensitive, and eliminating mode 5 will likely affect the quality of the training data. However, the $QoSTD^{i,j}$ values tended to increase when they were larger than 0.6; thus, based on the above analysis, we can deduce that eliminating any one of the five modes would not greatly reduce the quality of the prediction.

The values of ave_QoSTD^i for each of the four TCMDs when a single mode is eliminated are shown in Table 2, and these are compared with those obtained using all of the modes.

Table 2. Values of ave_QoSTD^i when a single mode is eliminated

ave_QoSTD^i	i=1(TCMD1)	i=2(TCMD2)	i=3(TCMD3)	i=4(TCMD4)
no-M1	0.7326	0.4089	0.4475	0.803
no-M2	0.6929	0.3644	0.393	0.8337
no-M3	0.7301	0.3681	0.3934	0.7951
no-M4	0.7337	0.404	0.4	0.8354
no-M5	0.7025	0.4095	0.5854	0.8197
all	0.73	0.41	0.4	0.8

From Table 2, we can see that it is not clear whether eliminating a single mode reduces the quality of the training data. This needs to be confirmed by calculating the ave_QoSTD^i values. Based on the analysis presented above, if these values are larger than 0.6, we can estimate that eliminating the corresponding mode would not degrade the quality of the perception of the status.

- Examining the influence of utilizing only a single mode on the perception of status

The values of ave_QoSTD^i based on only a single mode are shown in Table 3 for each of the four TCMDs. These values are compared to the values obtained when all of the modes were used.

Table 3. Values of ave_QoSTD^i utilizing one mode

ave_QoSTD^i	i=1(TCMD1)	i=2(TCMD2)	i=3(TCMD3)	i=4(TCMD4)
M1	0.6597	0.5704	0.4699	0.69
M2	0.6964	0.5578	0.5423	0.55
M3	0.6994	0.4109	0.3913	0.7
M4	0.6174	0.5057	0.4061	0.62
M5	0.731	0.3748	0.405	0.69
all	0.73	0.41	0.4	0.8

In the instance of TCMD1, the values of ave_QoSTD^i do not decrease very much when only a single mode is used. We can estimate that when the training data are labelled by TCMD1, the reliability of the Zheng's prediction is not degraded if only the features of a single mode are utilized as the training data.

In the instance of TCMD4, the values of ave_QoSTD^i show a large variation. We can estimate that when the training data are labelled by TCMD4, the reliability of the Zheng's prediction is degraded if only the features of a single mode are utilized as the training data.

For TCMD2 and TCMD3, the values of ave_QoSTD^i greatly increased for some modes, although the values were less than 0.6. It seems that it may be possible to improve the prediction of the Zheng's by using the features of only a single mode or by eliminating some of the features of some of the modes, when the quality of the full training data set is not very good.

On the basis of above experimental results and analysis, we conclude that the reliability of the status perception could be estimated by analyzing the values of $QoSTD^{i,j}$. The larger $QoSTD^{i,j}$ is, the better the quality of the prediction is. Accordingly, we can set a threshold value for the $QoSTD^{i,j}$ to

ensure a given level of the quality of the training data set, and this will improve the reliability of the subjective perception. The training data for which the value of $QoSTD^{i,j}$ is less than the threshold are considered to be of insufficient quality, and they are not used to predict the state.

On the other hand, by comparing the values of $QoSTD^{i,j}$ when all of the modes were utilized, we can determine whether a change in the data modality will affect the quality of the status perception. If the values of $QoSTD^{i,j}$ are greatly reduced when the features of some modes are removed, this indicates that this combination will reduce the quality of the prediction and should thus be avoided.

7. Conclusion

This paper defined the $QoSTD^{i,j}$ metric as a way to measure the quality of training data subjectively labelled by observers (i), which was used to improve the prediction of states (j) of an individual's health status. We also presented a method for determining an appropriate threshold for the $QoSTD^{i,j}$, so that unsupported predictions will not be made. Furthermore, the SVM model was used to predict the defined states subjectively labelled by observers, and the averages of SVM scores $ave_score^{i,j}$ and classification losses $loss^{i,j}$ was used to evaluate the prediction performance.

We used the training data set of TCM Zheng's that was presented in [4] and [5] to examine the effectiveness of using $QoSTD^{i,j}$ to improve the reliability of a prediction of an individual's status. Experimental results show that there are a high positive correlation between $QoSTD^{i,j}$ and $ave_score^{i,j}$ for each of the Zheng's (j) labelled by each of the TCMDs (i), and a negative correlation between $QoSTD^{i,j}$ and $loss^{i,j}$. Especially, there is a definite threshold for the $QoSTD^{i,j}$, beyond which the $ave_score^{i,j}$ is dramatically increased, and $loss^{i,j}$ is dramatically decreased. We thus verified that $QoSTD^{i,j}$ can be used to assure the reliability of the state prediction. Also, this shows that we can estimate the performance of the classification by analyzing the values of $QoSTD^{i,j}$ even though the classifier isn't trained by the subjectively labelled training data.

We also analyzed the influence of multimodal data on the perception of the status. The experimental results verify that we can utilize $QoSTD^{i,j}$ to judge whether it will affect the quality of the prediction if some modes are eliminated.

As an area of future work, we intend to utilize other well-known machine learning classifiers to predict the defined states subjectively labelled by observers, so as to verify the effectiveness of $QoSTD^{i,j}$ in improving the reliability of the status perception furthermore.

8. Acknowledgments

This work was supported by research funds from Iwate Prefecture University. The author would like to thank Prof. Shaozi Li and Prof. Feng Guo for their cooperation in data collection, system implementation, and experiments.

9. References

- [1] R. W. Picard, "Affective computing", The MIT Press, 2000.
- [2] Ying Dai, et al. (Ed.), "Kansei engineering and soft computing: theory and practice", Engineering Science Reference, IGI Global, 2010.
- [3] A. Vinciarelli, et al. "A survey of personality computing", IEEE Trans. on affective computing, vol. 5, no. 3, pp. 273-291, 2014.
- [4] Feng Guo, Ying Dai, et al. "Inferring individuals' sub-health and their TCM syndrome based on the diagnosis of TCM doctors", Proc. of IEEE SMC 2010, pp. 3717-3724, 2010, Turkey.
- [5] Yi Wang, Ying Dai, et al. "Sensitive-based information selection for predicting individual's sub-health on TCM doctors' diagnosis data", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 23, No. 5, pp. 749-760 (2011).
- [6] Koichi Sugiyama, "An introduction to Multivariate data analysis", Asakura Bookstore (in Japanese).
- [7] Ying Dai, "Evaluating the effect of different Mode's attributes on the subjective classification in the case of TCM", Proc. of IEEE CIMSIm, 2013, pp. 171-176, 2013, Korea.
- [8] Y. Dai, "Predicting person's Zheng states using the heterogeneous sensor data by the semi-subjective teaching of TCM doctors", Proc. of IEEE SMC 2014, pp. 636-641, 2014, USA.
- [9] J. Wagner, et al. "Smart sensor integration: A framework for multimodal emotion recognition in real-time", Proc. of IEEE ACII 2009, pp. 1-8, 2009.
- [10] R. Olfati-Saber, et al. "Consensus filters for sensor network and distributed sensor fusion", Proc. of IEEE CDC-ECC 2005, pp. 698-6703, 2005.
- [11] Amar Mitiche, "Multiple sensor integration/fusion through image processing: A review", Optical Engineering, 25(3), doi:10.1117/12.7973834, 1986.

[12]P. R. Cadesch, "Subjective inference with multiple evidence", *Artificial Intelligence*, Vol. 28, Issue 3, pp. 333-341, 1986.

[13]T. Yanaru, "An emotion processing system based on fuzzy inference and subjective observations", *Proc. of IEEE Artificial Neural Networks and Expert Systems*, pp. 15-20, 1995.

[14] Yanwei Fu, et al. "Robust subjective visual property prediction from crowdsourced pairwise labels", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 1, pp. 1, PrePrints, 2015, doi:10.1109/TPAMI.2015.2456887.10.

[15]Touradj Ebrahimi, "Predicting subjective sensation of reality during multimedia consumption based on EEG and peripheral physiological signals", *Proc. of IEEE ICME 2014*, pp. 1-6, 2014, China.

[16]Ahmed Mohammed Mikaeil, et al. "Machine learning to data fusion approach for cooperative spectrum sensing", *Proc. of IEEE YBERC 2014*, pp. 429-434, 2014, China.

[17]Yilin Tan, et al. "Utilizing concept correlations for effective imbalanced data classification", *Proc. of IEEE IRI 2014*, pp. 561-568, 2014, USA.

[18] Jin-Xing Liu, et al. "RPCA-based tumor classification using gene expression data", *IEEE trans. on Computational Biology and Bioinformatics*, vol. 12, no. 4, pp. 964-970, 2015.