

Quality Assessment of Training Data with Uncertain Labels for Classification of Subjective Domains

Ying Dai

Faculty of Software and Information Science, Iwate Prefectural University, Takizawa, Japan

Email: dai@iwate-pu.ac.jp

How to cite this paper: Dai, Y. (2017) Quality Assessment of Training Data with Uncertain Labels for Classification of Subjective Domains. *Journal of Computer and Communications*, 5, 152-168.

<https://doi.org/10.4236/jcc.2017.57014>

Received: April 18, 2017

Accepted: May 21, 2017

Published: May 24, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In order to improve the performance of classifiers in subjective domains, this paper defines a metric to measure the quality of the subjectively labelled training data (QoSTD) by means of K-means clustering. Then, the QoSTD is used as a weight of the predicted class scores to adjust the likelihoods of instances. Moreover, two measurements are defined to assess the performance of the classifiers trained by the subjective labelled data. The binary classifiers of Traditional Chinese Medicine (TCM) Zhengs are trained and retrained by the real-world data set, utilizing the support vector machine (SVM) and the discrimination analysis (DA) models, so as to verify the effectiveness of the proposed method. The experimental results show that the consistency of likelihoods of instances with the corresponding observations is increased notable for the classes, especially in the cases with the relatively low QoSTD training data set. The experimental results also indicate the solution how to eliminate the miss-labelled instances from the training data set to re-train the classifiers in the subjective domains.

Keywords

Quality Assessment, Subjective Domain, Multimodal Sensor Data, Label Noise, Likelihood Adjusting, TCM Zheng

1. Introduction

Recently, much research is aimed at predicting the status of individuals in subjective domains, including their emotional states, their health, and their personality by using a set of training data, acquired from a variety of sensors and interpreted or labelled by a first-person or a third-person [1] [2] [3]. As described by C. E. Brodley in [4], labeling noises of training data set are emergent in such

domains for several reasons including data-entry error, inadequacy of the information used to label each object, especially, uncertainty of the states. In [5], statistical taxonomy of label noise inspired by [6] is summarized. There are three kinds of models defined: noisy completely at random model, noisy completely at random model, noisy random model, and noisy not at random. All of these models assume that the true classes are existed, and whether the labelling error occurs is told by introducing a binary variable. However, in the above domains of emotion, health, or personality, the absolute ground truth is unknown. The subjects including first-person and third-person subjectively provide labels as their choices, so that the class-uncertain label noise naturally appears, whoever providing these labels, a skillful expert or a nonprofessional. For example, in the domain of Traditional Chinese Medicine (TCM), the status of health is described fundamentally by 13 Zhengs which are diagnosed by TCM doctors based on the information acquired from five senses. Further, a disease severity of these 13 Zhengs are scored according to the subjective observation. Accordingly, scores of 13 Zhengs labelled by TCM doctors are ambiguous, because the absolute ground truth of the 13 Zhengs is unknown.

Many methods are proposed to deal with label noise. In the literature, there exist three main approaches to take care of label noise [5]. A first approach focuses on building algorithms that are robust to label noise; second, the quality of training data is tried to improve by identifying mislabeled data; eventually, label noise-tolerant learning algorithms aim at building a label noise model simultaneously with a classifier, which uncouples both components of the data generation process and improves results of the classifier. All of these methods are based on the underlying premise that the label errors are independent of the ground truth of the classes, that is, the ground truth of the classes is existed although the observation labels are contaminated for some reasons. However, it is obvious that the above premise is not available for the domains being of subjectivity, because the observation labels are also influenced by the uncertainty of the classes.

In order to deal with the above issue about the label noise of training data caused by the subjective labelling, especially in the domain that the ground truth is uncertain, we defined a metric QoSTD that is intended to measure the quality of the training data with uncertain labels for the classification of subjective domains, so as to predict the states of person's emotion, health, and so on. QoSTD is an aggregation of two components which reflect the ability of clustering and partitioning of the training data set. The training data include the features extracted from the multimodal sensor data of subjects, the subjective scores of various items in a first-person questionnaire, and observation scores of classes in a subjective domain which are provided by third-persons. By using this metric, we can analyze the influence of subjectively labelled data on the quality of the training data, and we can estimate the sufficiency of the training data for the classification. When QoSTD for a particular class is less than a predetermined value, this indicates that the training data for this class can't satisfy the performance of classification.

We trained binary classifiers for the states based on the support vector machine (SVM) model and the discrimination analysis (DA) model, so as to validate the relation of QoSTD with the performance of classification. Furthermore, the QoSTD is used as weights of predicted class scores to adjust the likelihoods of the instances without the absolute ground-truth. To evaluate the effectiveness of QoSTD in dealing with the label noise brought by subjective labelling, we used TCM Zheng training data set that was used in [7] and [8] for experiments. The experimental results show that the proposed method improved the consistency of likelihoods of instances with the corresponding observations notably for the classes, especially in the cases with the relatively low QoSTD training data set. The experimental results also indicated the solution how to eliminate the miss-labelled instances from the training data set to re-train the classifiers in the subjective domains.

2. Related Works

The literature contains many studies on the classification in the presence of label noise [4]. In [5], a method to identifying and eliminating mislabeled training instances for supervised learning is proposed. The paper focuses on the issue of determining whether or not to apply filtering to a given data. However, for the work described in the paper, the data were artificially corrupted. Therefore, the application of this method to relatively noise free datasets should not significantly impact the performance of the finally classification procedure. Moreover, the authors indicated that a future direction of this research would be to extend the filter approach to correct labeling errors in training data. However, it is difficult to judge labeling errors in the subjective domains, because the absolute ground truth is unknown. In [9], authors propose to use a unsupervised mixture model in which the supervised information is introduced, so as to compare the supervised information given by the learning data with an unsupervised modeling. For this model, the probability that the j th cluster belongs to the i th class is introduced to measure the consistency between classes and clusters. However, it is not possible to obtain an explicit solution of the above probability for the classes of the subjective domains. In [10], a self-training semi-supervised support vector machine (SVM) algorithm and its corresponding model selection method is proposed to train a classifier with small training data. The model introduces the Fisher ratio which represents the separability of a corresponding set. It is obvious that the above parameter is not available for the classes of the subjective domains in the case that the ground truth of the classes is unknown. In [11], the quality of class labels in medical domains is considered. However, the ground truth of the training data used in the experiments is assumed to be certain, and those were corrupted artificially to analyze the impact of inputted noise on the classification.

Moreover, 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. 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. However, all of these methods didn't consider how to deal with the effects of training data's mislabeling on the classification.

On the other hand, various methods have been proposed that utilize TCM to infer the health status of an individual as a means of auto-diagnosing. References [7] and [8] 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 [19] and [20] analyzes the effect of multimodal sensor data on the Zheng classification. However, all of these papers don't consider in introducing the metric QoSTD as the weights of the predicted class scores of the instances, so as to improve the reliability of the classification in the subjective domain.

3. Measuring the Quality of Subjectively Labelled Training Data

Because the target is the status' classification in the subjective domains, the data used as training data are generally diverse. For an instance, the data used to extract features or attributes maybe include the data measured by sensors or other equipment, or the data from the first-person questionnaires; with direct observation to the object, the states of the instance are labelled by third-persons for supervised learning. Although the kinds of obtained data are heterogeneous, all of the features extracted from the different modes are handled in the same way as the features of different modes. For example, the histogram, shape, and the texture of an 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_s^m , which are normalized for each data set. Here, s , m , and n indicate indices of the sample, the mode, and a certain mode's feature. The combined features of all of training samples yield a matrix A with the size of $S * M$. Here, S and M are the number of samples and total features, respectively. On the other hand, the labelled state scores

from the third-person for each instance 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.

Eigen feature vectors of the instance is obtained by calculating the eigenvalues and eigenvectors of $A' * A$; this is based on the method of principle component analysis (PCA) [21]. With ranking the eigenvalues in a descending order, the corresponding top P eigenvectors are selected to form a matrix U with the size of $M * P$. Then, the matrix of eigenfeatures regarding the samples is computed by the following equation.

$$EF = \{ef_{s,p} : s \in [1, S], p \in [1, P]\} = A * U \quad (1)$$

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

The eigenfeature vector is then used to represent the instance. 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 QoSTD is defined to measure 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. QoSTD is calculated not only based on the partition of the training data, but also the clustering ability of those. We call these two metric as the partition and the clustering. These determine the performance of the classification regarding the training data. Let the score of State j for Samples labelled by Observer i be denoted as z_s^{ij} . In the training data set, those that have scores larger than the value of 0 for state j are considered being labelled as state j , and compose the data set $PZ^{ij} = \{z_s^{ij} : z_s^{ij} \geq 0, s \leq S\}$, 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. One of them with the more samples labelled by State j is assumed as the positive cluster of State j , denoted as $PC^{ij} = \{pc_s^{ij}, s \leq S\}$, and the other is the negative cluster, denoted as $NC^{ij} = \{nc_s^{ij}, s \leq S\}$. Accordingly, the partition of the data set for State j labelled by Observer i is defined as

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

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

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

where $\#$ indicates the number of data points; $\#(PZ^{ij} \cap PC^{ij})$ indicates the number of the samples which are labelled as State j and clustered into the positive cluster of State j . So, the larger the values of par^{ij} and clu^{ij} is, the better the separability of the training data set for State j is. If these values are equal to 1, this means that the training data are completely separable. Accordingly, the quality of the training data set for classifying State j labelled by Observer i is defined as $QoSTD^{ij}$ by the following expression, which is an aggregation of par^{ij} and

clu^{ij} :

$$QoSTD^{ij} = w_1 par^{ij} + w_2 clu^{ij}. \quad (4)$$

Here, w_1 and w_2 are the weights of partition and clustering, reflect the importance of the partition and the clustering ability of the training data in the classification. In the case that these two factors are equivalently important, both are set to 0.5. The value of $QoSTD^{ij}$ is equal to 1, if the training data set is completely separable for the State j which are labelled by Observer i .

Figure 1 shows the example of 120 data points of $ef_{s,1}$ and $ef_{s,2}$. **Figure 1(a)** is the instances' scatter regarding State j_1 labelled by an observer, and **Figure 1(b)** is the distribution regarding State j_2 . The dark blue points indicate the corresponding positive instances belonged to that state, and the light blue points indicate the negative instances not belonged to that. **Figure 1(c)** is the clustering of the data points by K-means. The instances of cluster 1 are indicated by light orange points, and the instances of cluster 2 are indicated by dark orange points. We assumed that the cluster with the more positive instances is the positive cluster, and the cluster with the more negative instances is the negative cluster. For the case of (a), it is obvious that the cluster 1 is regarded as the negative cluster, and the cluster 2 is regarded as the positive cluster according to the results of Fig. (a), and (c). So does the case of (b).

Based on the definition of QoSTD, and combining the results of instances' clustering in **Figure 1(c)**, it is obvious that the quality of the data points' distribution in **Figure 1(a)** is better than that in **Figure 1(b)** for training the classification model, although the class of positive instances and the class of negative instances are overlapped either in the case of (a) or in the case of (b). In fact, the value of $QoSTD^{ij}$ regarding the case (a) is 0.78, and the value of that regarding the case (b) is 0.43. So, we think that the larger the values of $QoSTD^{ij}$ 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 completely into two classes with a positive or negative State j .

4. Using QoSTD for Classification

As mentioned above, the metric $QoSTD^{ij}$ could be used to judge the quality of training data for the classification. Accordingly, the value of $QoSTD^{ij}$ is considered to be used as a weight of the predicted class scores of the instances regarding States j .

For 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 EF is obtained by Equation (1). Then, the value of $QoSTD^{ij}$ for State j labelled by Observer i is calculated using Equations (2), (3), and (4).

The following is the scheme that trains classifiers utilizing $QoSTD^{ij}$.

Generally, existing supervised learning algorithm, for example, discrimination

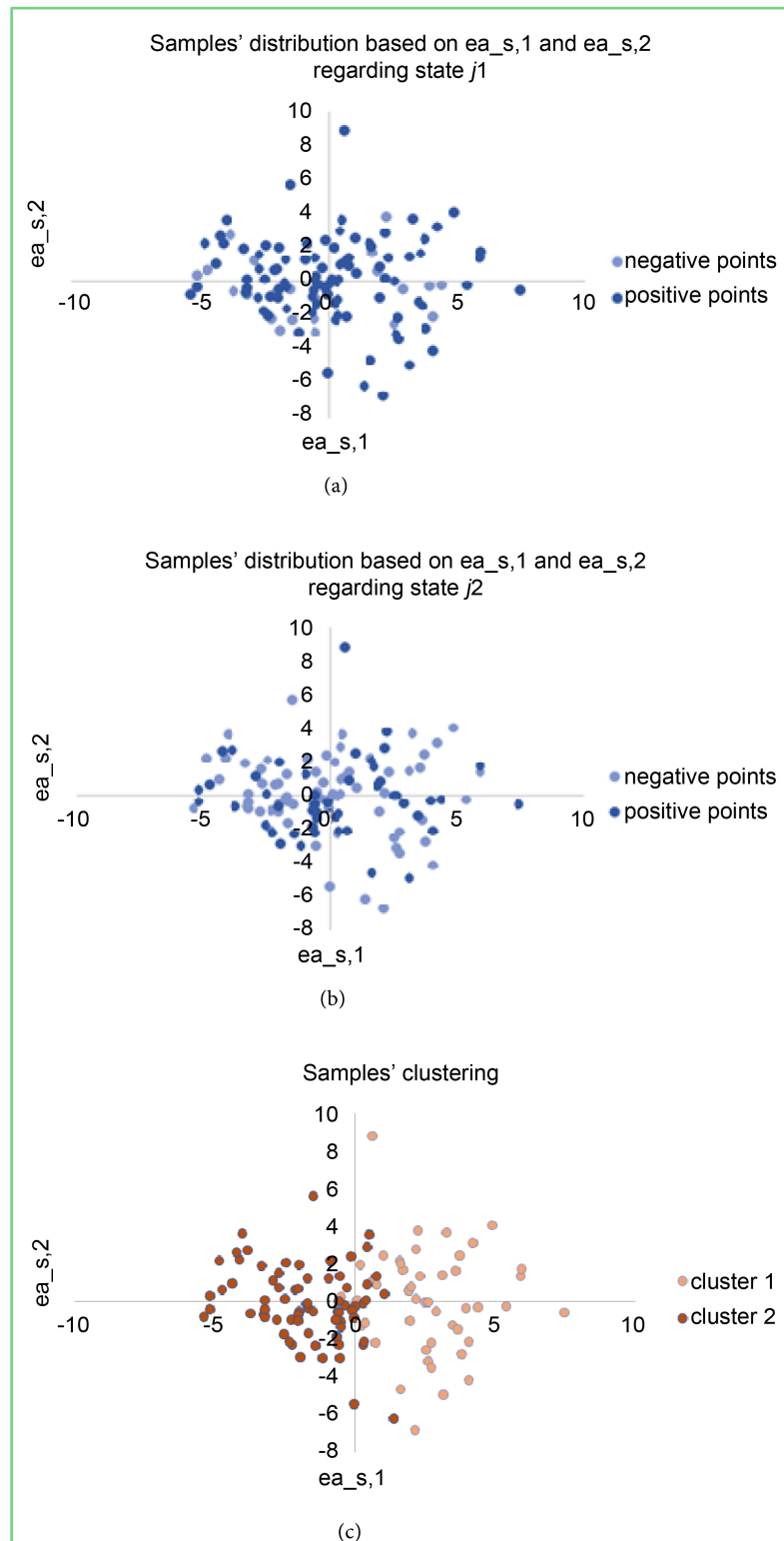


Figure 1. Samples' distribution and clustering. (a) Samples' distribution regarding state $j1$. (b) Sample's distribution regarding state $j2$. (3) Samples' clustering.

analysis (DA), support vector machine (SVM), or decision tree (DT), could be utilized to train binary classifiers of State j with the training data labelled by ob-

server i . With using the trained classification model, the predicted class score belonging to State j is generated for the response to the instance s , which is denoted as $score_s^{ij}$. However, considering that the quality of training data influences the performance of classification, $QoSTD^{ij}$ is utilized as a weight of $score_s^{ij}$ to adjust the scores of prediction. The corresponded computation is as the below.

$$score_r_s^{ij} = QoSTD^{ij} * score_s^{ij} \quad (5)$$

where, $score_r_s^{ij}$ denotes the adjusted score of instance s belonging to State j labelled by observer i .

Then, the likelihood of the instance belonging to State j is calculated by the Equation (6).

$$l_{rs}^{ij} = \frac{1}{1 + \exp(-a * score_r_s^{ij})} \quad (6)$$

where, l_s^{ij} indicates the likelihood of instance s belonging to State j labelled by observer i ; the parameter a is the slope parameter.

For the instance s , if the value of l_{rs}^{ij} is more than a threshold T_max , it is assigned to the positive lass of State j ; if that value is less than a threshold T_min , it is assigned to the negative class of State j ; otherwise, whether the instance is belonged to State j is uncertain. Then, the uncertain instances are eliminated from the training data set, and the classification model is trained again with the refined training data.

Two measurements, Con^{ij} and $Recall^{ij}$, are introduced to assess the performance of classifying the classes without the absolute ground-truth. Con^{ij} , which is defined by the following Equation (7), reflects the consistency of the labelled score of the assigned instances from the training data with the likelihood of those. Let the labelled score that is larger than the value of 0 is denoted as pz_s^{ij} , the likelihoods of the assigned instance is denoted as $l_ra_s^{ij}$, and the number of the assigned instances is denoted as $S1^{ij}$. Then,

$$Con^{ij} = \frac{\sqrt{\sum_{s=1}^{S1^{ij}} (pz_s^{ij} - \overline{pz}^{ij})(l_ra_s^{ij} - \overline{l_ra}^{ij})}}{\sqrt{\|pz_s^{ij} - \overline{pz}^{ij}\|^2 \|l_ra_s^{ij} - \overline{l_ra}^{ij}\|^2}} \quad (7)$$

On the other hand, $Recall^{ij}$, which is defined by Equation (8), reflects the rate of the number of the assigned instances to the all.

$$Recall^{ij} = \frac{S1^{ij}}{S^{ij}} \quad (8)$$

It is obvious that the larger the values of Con^{ij} and $recall^{ij}$ are, the better the performance of the classifiers for the classification.

Let the object value of Con^{ij} is Con_Obj^{ij} , and that of $Recall^{ij}$ is $Recall_Obj^{ij}$. Then, the whole training procedure is as the below.

Step 1

Constructing the binary classification model;

Step 2

Calculating the adjusted likelihood $l_{r_s}^{ij}$ of instances by Equations (5) and (6);

Step 3

If $T_{\min} < l_{r_s}^{ij} < T_{\max}$, the instance s is not assigning to State j ; otherwise, the instance s is assigned to the positive or negative class of State j according to the likelihood;

Step 4

Eliminating the unassigned instances from the training data set;

Step 5

Calculating Con^j and $Recall^j$ by Equations (7) and (8), and repeating the procedure from Step 1 to Step 4, until the limited rounds or $Recall_Obj^j$ are reached;

Step 6

Finding the maximal value of Con^j , and the corresponding round. The binary classification model constructed in this round is used as the final model, if this is larger than Con_Obj^j , and the corresponding $Recall^j$ is larger than $Recall_Obj^j$; if not, the final classification model can't be determined, and the training procedure is given up.

After constructing the binary classification model, a new instance could be assigned to the positive class of State j , if its $l_{r_s}^{ij}$ is larger than T_{\max} ; however, it is assigned to the negative class of State j , if its $l_{r_s}^{ij}$ is less than T_{\min} . Otherwise, which the instance is belonged to is uncertain.

5. Training Data Set

In this study, the real-world training data set that was used in [7] [8] is utilized for predicting the individual's health status represented by the states of TCM's thirteen Zhengs (Clod syndrome, Pyretic syndrome, Deficiency of vital energy, Qi stagnation, Blood asthenia, Blood stasis, Jinxu, Phlegm retention, Heart syndrome, Lung syndrome, Spleen syndrome, Liver syndrome, Kidney syndrome), so as to validate the effectiveness of the proposed method. This dataset 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, and tongues are shown in **Figure 2**.

The extracted features were combined with the above feelings and physical

states to form the matrix A . Each of these items is the modes of 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 instances from 32 individuals in the dataset, each of which includes 71 features and the corresponding thirteen Zheng scores labelled by the four TCMDs.

The matrix EF of eigenfeature vectors of the instances is obtained by Equation (1) with calculating the eigen values and eigen vectors of $A' * A$. Then, the matrix EF is used to train the binary classifiers of TCM Zheng.

Although two kinds of classification models are trained In order to verify the above statement that $QoSTD^j$ can be utilized as the weight of the predicted class score to improve the performance of the classifiers, especially, in the case that the training data are subjectively labelled and the ground truth is uncertain, all of existing supervised classification models are available. For these two kinds of classifiers, one is SVM model that is trained by utilizing the MATLAB (Mathworks, Natick, MA, USA) function *fitcsvm*. The kernel function here is a polynomial of order three. The other is DA model that is trained by MATLAB function *fitcdiscr*.

Based on the above binary classification model, the class scores of the instances belonging to Zheng j are obtained by using the MATLAB function *predict*. Then, the class scores are used to calculate the likelihood measures of the corresponding instances by Equations (5) and (6). For the SVM model, the value of parameter a in the Equation (6) is set as 1/100, and for the DA model, the value of that is set as 50. Moreover, $T_{max} = 0.99$, and $T_{min} = 0.01$. For a instance s , it is assigned to Zheng j , if the calculated likelihood is larger than 0.99;

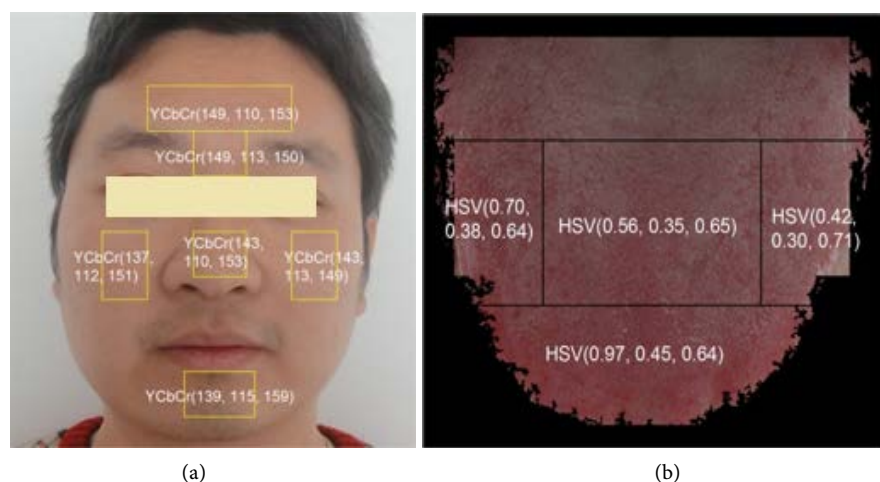


Figure 2. Features extracted from face and tongue. (a) Face. (b) Tongue.

Table 1. Modes and the number of features.

Symbol	M_1	M_2	M_3	M_4	M_5
Mode	Feeling	Physicals	Eye	Tongue	Face
Features	9	15	9	20	18

if the likelihood is less than 0.01, the instance s is not belonged to Zheng j ; otherwise, the assigned post of the instance s is uncertain. The training procedure is repeated with the refined training data set, until the limited rounds or $Recall_Obj^{ij}$ are reached.

6. Experimental Results and Analysis

6.1. About $QoSTD^{i,j}$

Figure 3 shows the values of $QoSTD^{ij}$ for all thirteen Zhengs, as labelled by four TCMDs.

From **Figure 3**, 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 it is certain that the $QoSTD^{ij}$ can be used as a criterion for judging the quality of the subjectively labelled training data. If the $QoSTD^{ij}$ is less than a threshold, the following learning procedure should be given up, so as to ensure the performance of the classification.

6.2. About Adjusting the Predicted Class Scores

As described in Section 3, the predicted class scores of the instances are adjusted by introducing $QoSTD^{i,j}$ as the weights of those scores. For exploiting how adjusting the predicted class scores improve the performance of the classification, another measurement that reflects the consistency of the labelled score of the assigned instances from the training data with the likelihoods of those in the case without adjusting the class scores is introduced. This measurement Con_ori^{ij} is calculated by the Equation (9).

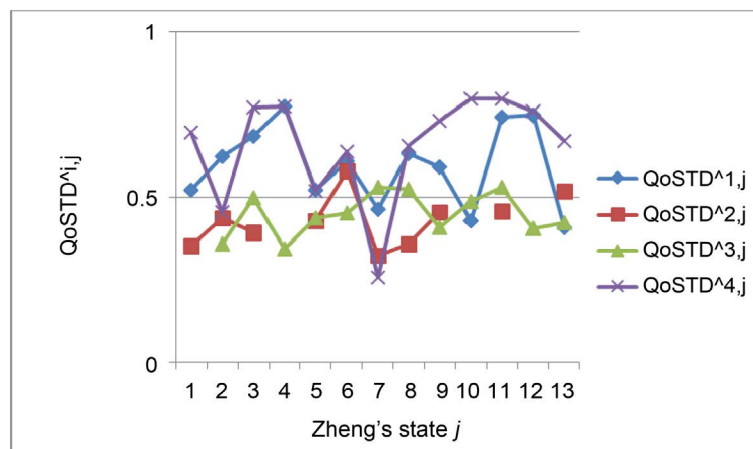


Figure 3. Quality of the training data set for all thirteen Zhengs.

$$Con_ori^{ij} = \frac{\sqrt{\sum_{s=1}^{S1^{ij}} (pz_s^{ij} - \overline{pz_s^{ij}})(l_a_s^{ij} - \overline{l_a_s^{ij}})}}{\sqrt{\|pz_s^{ij} - \overline{pz_s^{ij}}\|^2} \sqrt{\|l_a_s^{ij} - \overline{l_a_s^{ij}}\|^2}} \quad (9)$$

where, $l_a_s^{ij}$ indicates the likelihoods of the assigned instances in the case that the class scores are not adjusted by the Equation (5).

Table 2 shows the values of $QoSTD^{ij}$ and the first round's results of Con^{ij} and Con_ori^{ij} of the instances in the above training data set regarding thirteen Zhengs by using the DA-based and the SVM-based binary classifiers, while the scores of the instances are labelled by the TCM doctor identified as 1 (TCMD1). **Table 3** shows the corresponding results, while the scores of the instances are

Table 2. Results based on TCMD1.

TMCD1		DA			SVM		
Zheng	QoSTD	Con_ori	Con	increased rate	Con_ori	Con	increased rate
1	0.52	0.16	0.39	1.47	0.46	0.49	0.06
2	0.62	0.82	0.83	0.02	0.80	0.79	0.00
3	0.69	0.84	0.85	0.02	0.81	0.81	0.00
4	0.78	0.82	0.83	0.00	0.82	0.82	0.00
5	0.52	0.10	0.16	0.50	0.45	0.45	0.00
6	0.61	0.63	0.60	-0.04	0.63	0.63	0.01
7	0.46	0.23	0.36	0.59	0.44	0.46	0.06
8	0.63	0.83	0.83	0.00	0.83	0.83	0.00
9	0.59	0.77	0.82	0.06	0.70	0.70	0.00
10	0.43	0.01	0.10	13.40	0.29	0.37	0.30
11	0.74	0.90	0.90	0.00	0.91	0.91	0.00
12	0.75	0.81	0.82	0.00	0.81	0.81	0.00
13	0.41	0.10	0.31	2.05	0.53	0.55	0.03

Table 3. Results based on TCMD3.

TMCD3		DA			SVM		
Zheng	QoSTD	Con_ori	Con	increased rate	Con_ori	Con	increased rate
1							
2	0.36	0.12	0.39	2.27	0.49	0.50	0.04
3	0.50	0.31	0.32	0.04	0.66	0.66	0.00
4	0.35	0.19	0.33	0.77	0.45	0.50	0.10
5	0.44	0.31	0.39	0.23	0.39	0.42	0.07
6	0.45	0.15	0.24	0.66	0.27	0.29	0.07
7	0.53	0.59	0.67	0.12	0.62	0.62	0.00
8	0.53	0.67	0.45	-0.33	0.68	0.68	0.00
9	0.41	0.37	0.07	-0.81	0.42	0.47	0.12
10	0.49	0.00	0.02	106.00	0.36	0.38	0.06
11	0.53	0.77	0.79	0.02	0.69	0.69	0.00
12	0.41	0.24	0.46	0.92	0.53	0.55	0.05
13	0.42	0.28	0.38	0.39	0.50	0.52	0.05

labelled by the TCM doctor identified as 3 (TCMD3). For **Table 3**, the results about Zheng 1 are empty, because there were not any instances labelled by TCMD3 for Zheng 1 in the training data set.

From the results of **Table 2** and **Table 3**, we can see that the values of Con^{ij} is increased for almost of the Zhengs, compared with the results of Con_{ori}^{ij} . In the case of TCMD1, compared with the corresponding values of Con_{ori}^{ij} , the values of Con^{ij} for all of the Zhengs are gained with the SVM classification model; those rise for twelve of thirteen Zhengs with DA model. In the case of TCMD3, similar with the case of TCMD1, except two Zhengs with DA model, the values of Con^{ij} are increased for the Zhengs with neither SVM nor DA model. Especially, the increased rates are relatively notable for the almost Zhengs in the case that the $QoSTD^{ij}$ is less than 0.5.

Moreover, it is observed that the most of Con^{ij} are larger with SVM modal compared with DA model; however, the most of increased rates of Con^{ij} to Con_{ori}^{ij} are larger with DA model compared with SVM model. This means that the DA-based classifiers are more sensitive to $QoSTD^{ij}$ than the SVM-based classifiers, although the SVM-based classifiers seem to have the better classification ability.

Accordingly, we can say that adjusting the predicted class scores with $QoSTD^{ij}$ as the weights really improves the performance of the classifiers trained, especially in the cases that the classifiers are trained by the data set that is with the low values of $QoSTD^{ij}$, whatever the classification models which are used to train the classifiers.

6.3. About Re-Training

As described in Section 4, the training produce is repeated for constructing the classifiers with the refined training data set until the limited rounds or $Recall_Obj^{ij}$ are reached. In our experiments, the limited rounds is set as 20, Con_Obj^{ij} is set as 0.7, and $Recall_Obj^{ij}$ is set as 0.1. **Table 4** shows the results of maximal Con^{ij} , the corresponding round, and the difference of maximal and first round Con^{ij} regarding thirteen Zhengs, while the scores of the instances are labelled by the TCM doctor identified as 1 (TCMD1). **Table 5** shows the corresponding results, while the scores of the instances are labelled by the TCM doctor identified as 3 (TCMD3).

From the results of **Table 4** and **Table 5**, we can see that the differences of maximal and first-round Con^{ij} is larger or equal to 0, whatever the cases of TCM doctors and the models used to training the classifiers. Especially, in most cases, these values are larger than 0. Moreover, the differences of maximal and first-round Con^{ij} is relatively high in the case of TCMD3 which corresponds to the relatively low $QoSTD^{ij}$. So we can deduce that eliminating the unsigned examples in the training data set and re-training the binary classifiers with the refined training data set can really improve the performance of the classifiers for the classification, especially in the case that $QoSTD^{ij}$ is relatively low. However, it is noted that for some of Zhengs in **Table 4** and **Table 5**, the maximal values

Table 4. Results based on TCMD1.

TMCD1		DA			SVM		
Zheng	QoSTD	max	max-1th	max order	max	max-1th	max order
1	0.524	0.391	0.000	1	0.553	0.067	4
2	0.623	0.835	0.005	2	0.798	0.004	9
3	0.685	0.850	0.000	1	0.815	0.003	2
4	0.777	0.827	0.000	2	0.818	0.000	1
5	0.521	0.156	0.000	1	0.500	0.053	3
6	0.610	0.668	0.065	2	0.634	0.002	2
7	0.464	0.361	0.000	1	0.577	0.113	7
8	0.633	0.833	0.001	2	0.834	0.000	1
9	0.594	0.819	0.004	2	0.700	0.004	2
10	0.429	0.098	0.000	1	0.577	0.206	10
11	0.743	0.903	0.000	1	0.909	0.000	1
12	0.750	0.818	0.001	2	0.812	0.000	1
13	0.410	0.310	0.004	2	0.616	0.070	10

Table 5. Results based on TCMD3.

TMCD3		DA			SVM		
Zheng	QoSTD	max	max-1th	max order	max	max-1th	max order
1							
2	0.36	0.895	0.503	4	0.590	0.085	6
3	0.50	0.772	0.511	3	0.851	0.187	6
4	0.35	0.328	0.000	1	0.689	0.193	5
5	0.44	0.707	0.322	5	0.622	0.207	5
6	0.45	0.244	0.000	1	0.633	0.340	5
7	0.53	0.828	0.159	2	0.794	0.175	6
8	0.53	0.897	0.448	5	0.858	0.181	6
9	0.41	1.000	0.931	3	0.797	0.328	6
10	0.49	1.000	0.979	3	0.766	0.388	6
11	0.53	0.878	0.120	2	0.759	0.072	6
12	0.41	0.884	0.425	5	0.777	0.223	6
13	0.42	0.918	0.536	5	0.811	0.289	6

of Con^{ij} do not reach the value of Con_Obj^{ij} that is set as 0.7, although they raise after re-training. It is indicated that the provided training data set cannot make the corresponding classifiers achieve the required performance for these Zhengs. In such cases, constructing the classifiers for these states should be given up.

It is also noted that the round of re-training that make Con^{ij} maximal is different. For example, for Zheng 2 in the case of TCMD1, the maximal $Con^{1,2}$ is

0.84, and the corresponding round is 2th round by DA model; the maximal $Con^{1,2}$ is 0.80, and the corresponding round is 9th round by SVM model. For Zheng 2 in the case of TCMD3, the maximal $Con^{1,2}$ is 0.90, and the corresponding round is 4th round by DA model; the maximal $Con^{1,2}$ is 0.60, and the corresponding round is 6th round by SVM model. It is obvious that this matches the issue described in [18] that discarding an uncertain instance in the training data set maybe influence the performance of the classification because that is an exception rather than an error for the small training data set. So, we deduce that we can't say that re-training the classifiers with the refined training data set consecutively must improve the performance of the classification. The proposed solution regarding the above issue is to find the round that make Con^{ij} maximal and satisfy the condition regarding $Recall_Obj^{ij}$, so as to adopt the classifier trained in this round.

As a whole, we used the real-world training data set to train the classification models. This training data set involved in thirteen Zheng's labelled by TCM doctors with the label noises that were caused because the absolute ground-truth is unknown, while the current research in the literature almost utilizes the artificial corrupted training data set. The experiments verified that the $QoSTD^{ij}$ is relevant to the performance of classifying the classes without the absolute ground-truth. There is the high positive correlation between $QoSTD^{ij}$ and Con^{ij} , introducing the measurement of $QoSTD^{ij}$ as the weights of the predicted class scores to adjust the likelihoods of the instances really improved the performance of the classifiers.

7. Conclusions

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) without the absolute ground-truth. The $QoSTD^{ij}$ was used as the weights of the predicted class scores to adjust the likelihoods of the instances. Moreover, two measurements of Con^{ij} and $Recall^{ij}$ were defined in order to assess the performance of the classifiers trained by the subjective labelled data in a more suitable way. The training procedure was repeated by the refined training data set, until the object values of Con^{ij} and $Recall^{ij}$ were reached.

For verifying the effectiveness of the proposed method, real-world training data set was used to train the classifiers based on the DA and SVM classification models. This training data set involved in thirteen Zheng's labelled by TCM doctors with the label noises that was caused because the absolute ground-truth is unknown. The experimental results showed the effectiveness of the proposed method in improving the performance of the classifiers for the instances without the absolute ground truth. Furthermore, the proposed method indicated the solution how to eliminate the instances with the label noises from the training data set.

As an area of future work, we intend to utilize the other training data set in

the field of emotion, personality, and so on, to train the classifiers based on the proposed method, to verify the effectiveness of our method in improving the classification in the subjective domains.

Acknowledgements

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

References

- [1] Picard, R.W. (2000) Affective Computing. The MIT Press, Cambridge, MA.
- [2] Dai, Y., *et al.* (Ed.) (2010) Kansei Engineering and Soft Computing: Theory and Practice. Engineering Science Reference, IGI Global.
- [3] Vinciarelli, A. and Mohammadi, G. (2014) A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, **5**, 273-291. <https://doi.org/10.1109/TAFFC.2014.2330816>
- [4] Brodley, C.E. and Friedl, M.A. (1999) Identify Mislabeled Training Data. *Journal of Artificial Intelligence Research*, **11**, 131-167.
- [5] Frenay, B. and Verleysen, M. (2014) Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, **25**, 845-869. <https://doi.org/10.1109/TNNLS.2013.2292894>
- [6] Schafer, J.L. and Graham, J.W. (2002) Missing Data: Our View of the State of the Art. *Psychological Methods*, **7**, 147-177. <https://doi.org/10.1037/1082-989X.7.2.147>
- [7] Guo, F., Dai, Y., Li, S. and Ito, K. (2010) Inferring Individuals' Sub-Health and Their TCM Syndrome Based on the Diagnosis of TCM Doctors. 2010 *IEEE International Conference on Systems Man and Cybernetics (SMC)*, Istanbul, 10-13 October 2010, 3717-3724.
- [8] Wang, Y., Dai, Y., Guo, F. and Li, S. (2011) 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*, **23**, 749-760. <https://doi.org/10.3156/jsoft.23.749>
- [9] Bouveyron, C. and Girard, S. (2009) Robust Supervised Classification with Mixture Models: Learning from Data with Uncertain Labels. *Pattern Recognition*, **42**, 2649-2658.
- [10] Li, Y., Guan, C., Li, H. and Chin, Z. (2008) A Self-Training Semi-Supervised SVM Algorithm and Its Application in an EEG-Based Brain Computer Interface Speller System. *Pattern Recognition Letters*, **29**, 1285-1294.
- [11] Pechenizkiy, M., Tsymbal, A., Puuronen, S. and Pechenizkiy, O. (2006) Class Noise and Supervised Learning in Medical Domains: The Effect of Feature Extraction. 19th *IEEE International Symposium on Computer-Based Medical Systems*, Salt Lake City, UT, 22-23 June 2006, 708-713. <https://doi.org/10.1109/cbms.2006.65>
- [12] Cadesch, P.R. (1986) Subjective Inference with Multiple Evidence. *Artificial Intelligence*, **28**, 333-341.
- [13] Yanaru, T. (1995) An Emotion Processing System Based on Fuzzy Inference and Subjective Observations. 2nd *New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, Dunedin, 20-23 November 1995,

- 15-20. <https://doi.org/10.1109/annes.1995.499429>
- [14] Fu, Y., *et al.* (2015) Robust Subjective Visual Property Prediction from Crowd-sourced Pairwise Labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 563-577. <https://doi.org/10.1109/TPAMI.2015.2456887>
- [15] Kroupi, E., Hanhart, P., Lee, J.S., Rerabek, M. and Ebrahimi, T. (2014) Predicting Subjective Sensation of Reality during Multimedia Consumption Based on EEG and Peripheral Physiological Signals. 2014 *IEEE International Conference on Multimedia and Expo (ICME)*, Chengdu, 14-18 July 2014, 1-6. <https://doi.org/10.1109/ICME.2014.6890239>
- [16] Mikaeil, A.M., Guo, B. and Wang, Z. (2014) Machine Learning to Data Fusion Approach for Cooperative Spectrum Sensing. 2014 *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Shanghai, 13-15 October 2014, 429-434. <https://doi.org/10.1109/CyberC.2014.80>
- [17] Tan, Y., *et al.* (2014) Utilizing Concept Correlations for Effective Imbalanced Data Classification. 2014 *IEEE 15th International Conference on Information Reuse and Integration (IRI)*, Redwood City, CA, 13-15 August 2014, 561-568.
- [18] Liu, J.-X., Xu, Y., Zheng, C.H., Kong, H. and Lai, Z.H. (2015) RPCA-Based Tumor Classification Using Gene Expression Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **12**, 964-970. <https://doi.org/10.1109/TCBB.2014.2383375>
- [19] Dai, Y. (2013) Evaluating the Effect of Different Mode's Attributes on the Subjective Classification in the Case of TCM. 2013 *5th International Conference on Computational Intelligence, Modelling and Simulation (CIMSIm)*, Seoul, 24-25 September 2013, 171-176. <https://doi.org/10.1109/cimsim.2013.35>
- [20] Dai, Y. (2014) Predicting Person's Zheng States Using the Heterogeneous Sensor Data by the Semi-Subjective Teaching of TCM Doctors. 2014 *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, San Diego, CA, 5-8 October 2014, 636-641. <https://doi.org/10.1109/SMC.2014.6973980>
- [21] Sugiyama, K. (1999) An Introduction to Multivariate Data Analysis. Asakura Bookstore. (In Japanese)



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
 A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
 Providing 24-hour high-quality service
 User-friendly online submission system
 Fair and swift peer-review system
 Efficient typesetting and proofreading procedure
 Display of the result of downloads and visits, as well as the number of cited articles
 Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jcc@scirp.org

