A multi-modal heterogeneous data mining algorithm using federated learning

Xianyong Wei

Department of Computer, Shangqiu Polytechnic, Shangqiu, Henan 476000, China

Correspondence
Xianyong Wei, Shangqiu Polytechnic, Shangqiu, Henan 476000, China.
Email: weixianyong2020@163.com

Abstract
In disease diagnosis, the classification accuracy based on multi-modal models is usually higher than single-modal models. However, in the process of multimodal data fusion, how to reasonably solve the heterogeneity problem and better extract the information between different modal data has attracted the attention of scholars. Federated learning is an efficient machine learning method that can expand between multiple parameters or multiple computing nodes. It has been applied successfully in the financial industry and cross-industry cooperation. In this paper, a novel algorithm to disease diagnosis model based on federated learning is proposed. The model not only cleverly solves heterogeneous problems, but also excavates information between different modal data to promote the model to be more robust and discriminative. The experiment results show that the proposed model has better performance than traditional fusion algorithms. Notably, compared with other models, this model converges faster and requires less computation.

1 INTRODUCTION

The rapid development of modern society depends on the continuous innovation of digital technology. The same object will display several types of data in different fields, which is called multimodal data in machine learning. In the business system, companies usually analyse the different modal data to find out the potential needs of customers, such as the US Amazon company and China’s Alibaba [1]. During the shopping process, the website will implicitly recommend products that people potentially need (such as the famous diaper and beer story) to increase sales. In the field of TV, Netflix launched a competition for movie recommendation algorithms by setting up high bonus [2]. By analysing the different types of movies that people have seen, contestants apply machine learning models to recommend new movies to viewers. However, the experiment results show that people cannot design effective systems to recommend movies that the audiences are interested in. Therefore, how to mine effective information from different modal data has always been the focus of scholars.

In this paper, we focus on the multimodal data in medicine [3, 4]. There is a high degree of mismatch between the number of doctors and patients in the China’s medical field, in some departments, a doctor even needs to see more than 100 patient medical records a day. Introducing machine learning methods into the field of disease diagnosis can effectively alleviate this problem. Our goal is to extract valid information by mining data between different modalities to assist the doctors in diagnosing diseases, which can save the diagnosis time. A large number of experimental results show that algorithms based on multimodal data perform better than algorithms based on single modality [5]. Due to the particularity of medical data, heterogeneity problems often occur in modal fusion [6]. How to save the complementary information between different modal data while solving heterogeneous problem has been the research focus of multimodal fusion algorithms. In this paper, we propose an efficient multimodal fusion algorithm to simplify the analysis time of doctors for some common pathological phenomena.

In the issue of multimodal fusion, scholars have proposed many models, including merge method, linear accumulation method [7], kernel function [8] etc. However, these methods are difficult to maintain the independence of each modal data. For example, when the two modal data have remarkable difference in a certain feature area, they may become smooth after linear accumulation, thereby losing the authenticity of the data. And if the data are heterogeneous, the general fusion algorithms...
usually loss their effectiveness. Kernel function can map the original data space to kernel space to solve the heterogeneous problem, but the calculation amount between the kernel matrix is huge. Since the projection space is often infinite, it is also necessary to construct the corresponding kernel function to calculate the original data. Due to the non-convex of the kernel function, the calculation speed of main function is very slow or even difficult to converge. To alleviate the problems, in this paper, we proposed multimodal fusion algorithm based on federated learning (MMF-FL).

Unlike traditional machine learning, where samples need to be centralized into a model for joint training. Federated learning allows each data owner to save the data in the server, learns their own models independently, and then send the gradient descent direction to a third party. Finally, the third party will feedback the latest parameters to the data owner, loop these steps until all models converge. In traditional medical disease diagnosis, the feature space of data of different modalities is usually different, which hinders the effective fusion of modal data. Based on the characteristic of federated learning, it can easily solve the heterogeneous problem in multimodal fusion. Notably, in each iteration, the model of each modal can refer to the gradient descent of other algorithms to guide the next calculation, which can avoid the model from easily falling into the local optimal solution. In this way, federated learning can make the final convergence model more robust and generalized by fully utilizing the potential information between each modal data.

Obviously, for multi-modal fusion, federated learning treats different modal data as the centre instead of concentrating the data into the algorithm for training. Meanwhile, asynchronous training strategy for federated learning is usually used to deep learning models with the huge amount of data. The experimental data in this article uses small sample data, we adopt synchronous training strategies to effectively share different modal training algorithm gradient convergence direction in real time [21]. On the other hand, traditional federated learning model requires different contributors share the same feature space, which is contrary to the heterogeneous problem between different modal data. Federated transfer learning can solve the challenges because it shows better performance in the organizations from the same or related industry, which is contributed to the knowledge propagation of different modal data [22].

In our model, we select least squares (LS) [9] classification algorithm for each modal data. Our model can be roughly divided into four steps: firstly, we train the initial LS model base on sample data and send the gradient to the third party; secondly, we use the aggregation algorithm to analyse these gradient values; thirdly, the third party will feedback the results to each participant; Finally, each modal updates it respective algorithm with the gradients. Overall, the proposed method has the following contributions:

1. In this paper, we first introduce federated learning to multimodal medical disease diagnosis. We proposed a standard multimodal fusion algorithm MMF-FL, which can deal with any form of multimodal data while ignore the heterogeneous problem.
2. Our algorithm can efficiently use the complementary data between modalities to help each other fall into the local optimal solution prematurely by using aggregation algorithms in third party.
3. The experiment results shows that our model achieves the best classification results compared with other traditional multimodal fusion algorithms in the classification task of Alzheimer’s disease. In addition, our algorithm has less calculation and converges faster.

The rest of our paper is arranged as follows. Section 2 lists the related work about multimodal fusion models and federated learning. Section 3 describes the proposed method and the solution procedure. Section 4 presents the experiment results. Finally, Section 5 gives the conclusion.

2 | RELATED WORKS

2.1 | Multimodal data fusion

The data collected based on different devices will take many forms, such as palm print data under several spectrums, pictures obtained by different imaging techniques in disease diagnosis. Studies have shown that model based on multi modal data has better performance than model only based on single modal data in classification tasks. Because the use of complementary information between each modal data, the multi-modal model has stronger robustness and better generalization. The model proposed by Yang et al. [1] linearly combined all the sub-matrices into on matrix together, and then train the classification algorithm. In [2] work, they have given each modal data the same weight value and accumulated them into a matrix. Since the feature space of each modal data is often inconsistent, they filled in the missing features with 0 to unify the feature spaces. In Zhang et al. paper, they decoded all modal data into the same feature space through the hidden layer in the autoencoder (AE) [3, 4], and then linearly accumulated the data. Models in [5, 6] used low-rank learning to extract latent major structure form multimodal data, and trained the model to classify different modal data based on the framework [7]. The work of Alpha integration is mainly at the statistical level. The classification results of samples based different models can be used to obtain the corresponding scalar statistic, which are also referred to as scores. Finally, their algorithm can concentrate the fusion of scores to improve the individual performance [35].

However, these methods have changed the structure information of the original data while lost some feature information that may have important judgment value for sample classification. Our algorithm first independently trains the classification model of each modal data, and then apply the aggregation algorithm to analyse their gradient descent direction. In this way, the model can preserve the integrity of the original data.
2.2 Kernel function

In order to solve the problem of linear inseparability, it is necessary to map the samples from the original space to the high-dimensional space, and then implement the classification task [8]. This usually requires calculating the factorial of the sample space in the high-dimensional space, but the mapping function is often unknown and the target space is often large or even of infinite dimension, which is difficult to complete the operation process. To solve the problem, the inner product of the samples is calculated in the original space with a specific function form to replace the calculation result in the high-dimensional space, which is called kernel function (also known as invisible mapping) [9]. In this way, the samples in the original space become the form of kernel matrix, so the kernel function can effectively solve the heterogeneous problem and improve the classification accuracy. Zhang et al. [10, 11] firstly employed kernel function to AD multimodal data, the experiment results showed that their model achieved better performance than methods based on single modal data. According to the characteristics of disease data, He et al. [12] designed novel kernel function and obtained best classification results compared to traditional kernel method (linear kernel, polynomial kernel, Gaussian kernel etc.).

The huge calculation amount and slow convergence speed in kernel function only suitable for small sample data. Our model starts from the perspective of the gradient descent direction of each modal algorithm, which greatly saves the calculation time of the classification model.

2.3 Federated learning

Federated learning was originally proposed by Google scholars in 2016 to solve the problem of decentralization and confidentiality of Android user data [13, 14]. Each user’s preference for information such as music, videos, picture etc. is very valuable for the analysis of customer group attributes. However, these data are widely dispersed and the EU has issued terrific protection laws on customer privacy, which is almost impossible to form an effective analyse model. The federated learning model proposed by the Google team treats each Android user as a node of the model, it allows device to independently train data on the local server to update the overall model [15, 16]. In this way, algorithm not only protects the privacy of users but also fully mines the valuable information in decentralization data. In each training process, one user calculating the parameters in the local data set, and then feeding back the direction of gradient descents to aggregation algorithm to train the federated learning model.

Because of the effectiveness in distributed data fusion, scholars have actively applied it in various fields. Qiang Yang, the chief artificial intelligence officer of Tencent Weizhong Bank, summarized three general models of federated learning, including horizontal federated learning, vertical federated learning and federated transfer learning [16, 17]. Horizontal federated learning represents the scenario that groups share similar feature space but completely different sample set. For example, the customer groups of banks in two cities are completely different, but the business needs are basically the same. Vertical federated learning is just the opposite of horizontal federated learning, the data sets of the population have a great overlap, but the feature spaces are diverse. For example, residents of one block of the city, the intersection of their feature space, like financial attributes at the banks and the preference in goods, is quietly irrelevant. Federated transfer learning denotes the setting that two data sets differ not only in samples but also in feature space. For examples, the customer information in two businesses in different countries.

In our work, we select vertical federated learning as our algorithm framework [18, 19], and schematic diagram is shown in Figure 1. In disease diagnosis, the same patients have multimodal image data from different imaging technologies. In the algorithm design, we train a feature value from each modal data, and select the maximum value of the modal feature number as the iteration of one cycle. Repeat the above steps until the main function finally converges.

3 PROPOSED METHOD

3.1 Main function

In this module, we will elaborate on the motivation of our algorithm, main function framework and the iteration process. The characteristic of federated learning is equivalent to a community in different data fields, which has no fixed learning framework and convergence direction. With the continuous enrichment of training data sets, the central model will also has better generalization abilities, which makes it can produce greater benefits in business system. When the federated learning model is applied to multi-modal fusion, it does not need to gather all the sample data sets for training. Similar with star topology, it treats each modal data as node data, central model just collect the difference information transmitted form each modal data. Specifically, we use $M$ to denote the maximum feature numbers. In a learning cycle, initialize the parameter $\omega$ and set $M$ as the number of iterations. Each iteration learns a feature value in the modal data, and sends the gradient direction of each model to the central algorithm. The central model will update new gradient direction and feeds it back to each model to continue training, the above operations can be simplified to the framework of Figure 2.

Let $X^{(m)} = (x_{1}^{m}, x_{2}^{m}, \ldots, x_{n}^{m})$ denotes $m$-modal data set, $n$ is the sample numbers. The label vector $Y_{i}$, $y_{i}$ is the label of $i$-th sample. Meanwhile, assuming learning rate $\eta$, regularization parameter $\lambda$, $i$-th modal data regression parameter $\omega^{(m)}$. The main function can be formulated as:

$$\min_{\omega} \sum_{m=1}^{M} \| X^{(m)} \omega^{(m)} - Y \|^{2} + \frac{\lambda}{2} \sum_{m=1}^{M} \| \omega^{(m)} \|^{2}$$ (1)

In our work, we adopt alternating direction method of multipliers (ADMM) [20, 27, 28] to decompose our function. For
FIGURE 1 Architecture of our multimodal fusion algorithm based on federated learning (MMF-FL)

FIGURE 2 The process of MMF-FL learns new features in each iteration. In the multi-modal fusion model based on federated learning, each iteration introduces unlearned features in the modal, because the number of features between the modals are inconsistent, when learning the end of modal data with the largest feature numbers is as a training session arbitrary m-th modal data, the loss is:

\[ L_m = \|X^{(m)} \omega^{(m)} - Y\|^2 + \frac{\lambda}{2}\|\omega^{(m)}\|^2 \]  

(2)

Then gradient is:

\[ \frac{\partial L_m}{\partial \omega^{(m)}} = X^{(m)^T}X^{(m)} \omega^{(m)} - X^{(m)^T}Y + \lambda \omega^{(m)} \]  

(3)

As the solution process of the main function, our MMF-FL algorithm independently train the regression coefficients for each modal data. In each iteration, the classifier only learns a new feature value. Therefore, the form of parameter \( \omega \) changes from \( 1 \times 1 \) to \( 1 \times K \), where \( K \) is the feature number, \( \hat{a} \) is the specific feature vector in the modal data. For different modal data, the values of \( K \) may not be consistent. The evaluation steps for MMF-FL model is listed in Algorithm 1.

3.2 Complexity analysis

As mentioned in the previous sections, the storage space and calculation amount of our model are lower than the traditional machine learning models. In this paragraph, we will analyse the time complexity and space complexity of our algorithm. For each modal data, the maximum amount of calculation of model is equal to the number of feature value, its time frequency is from 1 to \( K \). Therefore, the time complexity of our algorithm is \( O(K) \). For sample data, the model need to calculate the matrix and the transpose of the matrix, which the space complexity is \( O(n \times n) \). In SVM and low rank models, it is usually necessary to perform function operations in the high-dimension space. Meanwhile, due to the non-convex of kernel function and low rank constrains, these models usually require a large number of iterations to obtain a better local optimal solution. Obviously,
our model is a convex function, which can obtain the optimal solution in a limited number of steps and perform operations in the original space.

4 | EXPERIMENTS

4.1 | Data sets

In order to verify the effectiveness of MMF-FL algorithm, we conducted experiments on two multimodal data sets (including 202-Alzheimer’s Disease Neuroimaging Initiative (ADNI) database, 913-ADNI database) [21] with several representative multimodal fusion models. Specifically, there are 50 AD subjects, 53 health control (HC) and 99 mild cognitive impairment (MCI) in the 202-ADNI database with three image modalities, including magnetic resonance imaging (MRI), positron emission tomography (PET) and cerebrospinal fluid (CSF) [29]. MCI is the prodromal stage of AD, and HC is the group of healthy people. In addition, 913-ADNI database is consisted of 160 ADs, 542 MCIs and 211 NCs, which contains five modalities, including serial number (ID), single nucleotide polymorphism (SNP), voxel based morphometry (VBM), fluorodeoxyglucose positron emission tomography (FDG) and F-18 forestair PET scans amyloid imaging (AV45). Further, MCI can be divided into several phases. For example, the MCI in 202-ADNI has two types, MCI converts (MCI-C) and MCI not-converts, the MCI-C will turn AD disease with 18 months while MCI-NC will maintain its original status. Similarly, MCI in 913-ADNI database has three types, like significant memory concern (SMC), early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI), Table 1 shows more details about the 913-ADNI database. Because of the indiscernibility of MCI, in our experiment, we apply the accuracy of distinguishing the different stages of MCI as our important criterion for multimodal fusion models.

4.2 | Experimental settings

In order to verify the effectiveness of the model, we will compare with traditional multimodal fusion algorithms in the aspects of classification accuracy, convergence speed and iteration time. In the classification aspect, we choose the challenging task of correctly classifying the corresponding stage of MCI patients. Six statistical standards are applied into measuring the results, including accuracy (ACC), sensitivity (SEN), specificity (SPE) balanced accuracy (BA), kappa index and area under the receiver operating characteristic curve (AUC). Balanced accuracy normalizes true positive and true negative predictions by the number of positive and negative samples respectively, and divided their sum into two, which can more accurately measure the classification ability of the model to deal with imbalanced data. Kappa is mainly used to measure the consistency between the predicted value and the source data. When it is equal to 1, it means that the two are completely consistent, when it is equal to -1, which corresponds to the opposite situation. For sample allocation, we adopt a ten-fold cross validation strategy, which divides the data set into ten same size mutually exclusive subsets, like \( D = D_1 \cup D_2 \ldots \cup D_{10} \). Before model training, we merged nine subsets into the training set while applied the remaining subset as the test set. When all the subsets are used as the test set, the average of the ten experimental values is used as the final classification result.

In this paper, we compare MMF-FL algorithm with several traditional multimodal fusion models. Support vector machine (SVM) is a common data fusion model, which map original data to high-dimensional space utilizing kernel functions to implement sample classification. Canonical correlation analysis (CCA) algorithm searches for the projection direction to transform high-dimensional data to a one-dimensional vector with the largest correlation coefficients. Zhu et al. [22, 30, 31] firstly applied low-rank constraints and self-paced learning to multimodal data fusion, called SPMRMR. Yuan proposed Latent correlation embedded multi-modal fusion (LCM2F) model [23], they adopted \( \ell_{2,1} \) norm to constrain regression coefficients of all modal data, the sparsity is adaptively controlled by changing the value of \( \rho \). Method in [24, 32] implemented convolutional neural network (CNN) to the feature-level fusion, and used logistic regression algorithm to classify the samples in the fully connected layer. Jiang et al. [25, 33] utilized stacked autoencoder (SAE) to solve heterogeneous problems in multimodal fusion by encoding all modal data into the same framework. In [26, 34] work, they presented long short term memory (LSTM) framework to automatically learn feature representations, which can improve the model performance.

4.3 | Experimental results on ADNI data

Firstly, we verify the effectiveness of our method on the two ADNI databases, which perform three classification
4.4 | Parameter influence

In the experiment, our algorithm has two regularization coefficients, \( \lambda \) and \( \eta \). To verify the stability of our model, which the classification results will not fluctuate greatly with the change of regularization terms. We set the values of parameter \( \lambda \) from \( 10^{-3} \) to 10 and \( \eta \) from \( 10^{-3} \) to 1. In the experiment, we fix one parameter while change the values of another term until all combinations are tested. Specifically, Figure 4 shows the influence of regularization term \( \lambda \) and \( \eta \). It can be seen that in different classification tasks, the classification results of our model does not change greatly in different parameter combinations, which verifies that the MMF-FL algorithm is not sensitive to specific parameters and is only related to the structure of the model.

4.5 | The convergence analysis

Figure 5 shows the convergence of our main function with the iteration numbers. The value of our main function has a significant down trend in limited steps. After a period of program running, our algorithm converge to a determined value while other baselines are still floating in the local optimal solution. As mentioned above, our model will converge faster due to the characteristic of federated learning, obviously the experimental results confirm the assumption.

5 | CONCLUSION

In this paper, we firstly adopt federated learning to multimodal fusion model. Benefits from the decentralized framework of federated learning, our proposed model MMF-FL can easily solve the heterogeneous problem in multi-modal fusion. Meanwhile, in each iteration, we only learn one value of the feature space in modal data. In this way, we can make full use of the regression coefficient of each modal to help each other to guide the subsequent gradient descent directions. The experiment results show that our algorithm not only has better performance than other multi-modal fusion models, but also has less calculation time and amount.

In this paper, our algorithm only conducts experiments on small multi-modal datasets. In future work, we will attempt to combine our model with deep learning networks to verify the effectiveness of MMF-FL in massive multi-modal datasets.

### TABLE 2 MCI-C versus MCI-NC classification results of our method and comparison methods

|     | SVM   | CCA   | SPMRM | LCM2F | CNN   | SAE   | LSTM   | MMF-FL |
|-----|-------|-------|-------|-------|-------|-------|--------|--------|
| ACC | 76.33 ± 0.034 | 74.03 ± 0.021 | 78.89 ± 0.019 | 77.25 ± 0.028 | 78.47 ± 0.056 | 75.20 ± 0.033 | 79.89 ± 0.057 | 81.59 ± 0.071 |
| SEN | 82.44 ± 0.028 | 76.16 ± 0.047 | 78.47 ± 0.035 | 81.53 ± 0.044 | 80.27 ± 0.068 | 74.12 ± 0.027 | 81.95 ± 0.026 | 85.29 ± 0.059 |
| SPE | 76.33 ± 0.019 | 74.03 ± 0.051 | 77.90 ± 0.080 | 76.16 ± 0.037 | 64.46 ± 0.038 | 70.28 ± 0.068 | 74.12 ± 0.087 | 79.62 ± 0.096 |
| BA  | 79.39 ± 0.024 | 75.10 ± 0.049 | 78.19 ± 0.058 | 78.85 ± 0.041 | 72.37 ± 0.053 | 72.20 ± 0.048 | 78.04 ± 0.057 | 82.46 ± 0.078 |
| Kappa | 72.45 | 75.21 | 78.19 | 77.28 | 71.17 | 73.01 | 76.91 | 80.21 |
| AUC | 70.89 ± 0.022 | 72.67 ± 0.034 | 76.91 ± 0.029 | 70.64 ± 0.011 | 72.88 ± 0.042 | 76.40 ± 0.019 | 78.15 ± 0.027 | 82.08 ± 0.028 |

### TABLE 3 SMC versus LMCI classification results of our method and comparison methods

|     | SVM   | CCA   | SPMRM | LCM2F | CNN   | SAE   | LSTM   | MMF-FL |
|-----|-------|-------|-------|-------|-------|-------|--------|--------|
| ACC | 69.50 ± 0.021 | 66.89 ± 0.028 | 73.63 ± 0.059 | 75.81 ± 0.027 | 69.86 ± 0.019 | 69.53 ± 0.097 | 75.07 ± 0.075 | 77.68 ± 0.069 |
| SEN | 56.75 ± 0.035 | 58.63 ± 0.034 | 64.70 ± 0.022 | 68.82 ± 0.038 | 60.94 ± 0.057 | 58.00 ± 0.032 | 65.16 ± 0.062 | 70.12 ± 0.063 |
| SPE | 70.13 ± 0.068 | 73.59 ± 0.019 | 74.18 ± 0.021 | 79.72 ± 0.044 | 66.48 ± 0.051 | 72.87 ± 0.058 | 70.71 ± 0.078 | 80.29 ± 0.095 |
| BA  | 63.44 ± 0.052 | 66.61 ± 0.027 | 69.44 ± 0.022 | 74.02 ± 0.041 | 63.71 ± 0.054 | 65.44 ± 0.040 | 67.94 ± 0.070 | 75.21 ± 0.079 |
| Kappa | 66.37 | 70.38 | 71.52 | 77.38 | 64.57 | 69.27 | 68.71 | 79.2 |
| AUC | 71.40 ± 0.026 | 64.74 ± 0.038 | 74.79 ± 0.051 | 72.93 ± 0.035 | 69.48 ± 0.080 | 64.93 ± 0.049 | 79.40 ± 0.072 | 81.33 ± 0.082 |
TABLE 4  EMCI versus LMCI classification results of our method and comparison methods

| Method | ACC       | SEN       | SPE       | AB        | Kappa     | AUC       |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|
| SVM    | 64.58 ± 0.043 | 62.19 ± 0.029 | 67.52 ± 0.093 | 64.86 ± 0.061 | 65.07 | 68.78 ± 0.045 |
| CCA    | 66.25 ± 0.042   | 69.81 ± 0.036 | 64.29 ± 0.028  | 67.05 ± 0.032  | 64.97 | 69.19 ± 0.014 |
| SPMRM  | 69.82 ± 0.028   | 76.33 ± 0.035 | 65.19 ± 0.090  | 70.76 ± 0.073  | 68.27 | 70.09 ± 0.038 |
| LCM2F  | 70.68 ± 0.053   | 70.28 ± 0.032 | 65.27 ± 0.046  | 67.78 ± 0.039  | 66.21 | 71.09 ± 0.031 |
| CNN    | 72.86 ± 0.034   | 77.19 ± 0.056 | 67.29 ± 0.042  | 72.24 ± 0.049  | 70.19 | 73.14 ± 0.089 |
| SAE    | 71.35 ± 0.053   | 74.51 ± 0.035 | 68.27 ± 0.024  | 71.39 ± 0.030  | 69.17 | 71.68 ± 0.048 |
| LSTM   | 72.15 ± 0.066   | 78.22 ± 0.087 | 69.11 ± 0.055  | 73.67 ± 0.071  | 71.26 | 74.68 ± 0.019 |
| MMF-FL | 75.69 ± 0.033   | 79.68 ± 0.062 | 70.24 ± 0.097  | 74.96 ± 0.080  | 72.05 | 76.81 ± 0.029 |

FIGURE 3 The accuracy of our model and comparison methods in three ADNI classification tasks

FIGURE 4 The results of three ADNI classification tasks with respect to the different parameter settings of $\eta$ and $\lambda$ (a) MCI-C versus MCI-NC, (b) SMC versus LMCI, (c) EMCI versus LMCI
Meanwhile, the multi-task model can help each classification task to better classify the sample. In the next work, we will employ the complementary information in different classification tasks to improve the discrimination ability of MMF-FL algorithm.

**ACKNOWLEDGMENTS**

This work was supported by the Young Backbone Teachers Funding Project of Henan Colleges and Universities (No. 2017GGJS263).

**REFERENCES**

1. Wasinger, R., Krüger, A., Jacobs, O.: Integrating intra and extra gestures into a mobile and multimodal shopping assistant. In: Pervasive Computing, Third International Conference. Pervasive, Munich (2005).
2. Ying, L., Kuo, C.C.-J.: Extracting movie scenes based on multimodal information. In: Electronic Imaging. San Jose, CA (2001).
3. Zhang, D., et al.: Multimodal classification of Alzheimer’s disease and mild cognitive impairment. Neuroimage 55(3), 856–867 (2011).
4. Shi, Y., et al.: Leveraging coupled interaction for multimodal Alzheimer’s disease diagnosis. IEEE Trans. Neural Networks Learn. Syst. 31(1), 186–200 (2019).
5. He, J., et al.: Imbalanced Multi-Modal Multi-Label Learning for Subcellular Localization Prediction of Human Proteins with Both Single and Multiple Sites. PLoS ONE 7(6), e37155 (2012).
6. Cheung, M.W.-L.: Issues in solving the problem of effect size heterogeneity in meta-analytical structural equation modeling: A commentary and simulation study on Yu, Downes, Carter, and O’Boyle (2016). J. Appl. Psychol. 103(7), 787–803 (2018).
7. Ran, D., Perré, S.: Linear-Time Approximation for Maximum Weight Matching. J. ACM 61(1), 1 (2014).
8. Pinheiro, B.C., et al.: Kernel-function-based models for acoustic localization of underwater vehicles. IEEE J. Oceanic Eng. 42(3), 603–618 (2016).
9. Geladi, P., Kowalski, B.R.: Partial least-squares regression: A tutorial. Anal. Chim. Acta 185(1), 1–17 (1986).
10. Yang, B., et al.: Online video recommendation based on multimodal fusion and relevance feedback. In: ACM International Conference on Image & Video Retrieval. Amsterdam (2007).
11. Hori, C., et al.: Attention-based multimodal fusion for video description. In: IEEE International Conference on Computer Vision (ICCV). Venice, Italy, pp. 4203-4212 (2017).
12. Gu, X., et al.: DialogWAE: Multimodal response generation with conditional wasserstein auto-encoder. In: International Conference on Learning Representations. Vancouver, BC, Canada (2018).
13. Hong, C., et al.: Multimodal deep autoencoder for human pose recovery. IEEE Trans. Image Process. 24(12), 5659–5670 (2015).
14. Abavisani, M., Patel, V.M.: Multimodal sparse and low-rank subspace clustering. Inf. Fusion 39, 168–177 (2018).
15. Liu, Z., et al.: Feature self-representation based hypergraph unsupervised feature selection via low-rank representation. Neurocomputing 253, 127–134 (2017).
16. Ramirez, J., et al.: SVM-based speech endpoint detection using contextual speech features. Electron. Lett. 42(7), 426–428 (2006).
17. Bonawitz, K., et al.: Federated optimization: Distributed machine learning for on-device intelligence. In: Proceedings of the 35th Annual Conference of the Neural Information Processing Systems. Long Beach (2018).
18. Konen, J., Mcmahan, H.B., Ramage, D., et al.: Federated optimization: Distributed machine learning for on-device intelligence. NIPS, 1610.02527, (2016).
24. Yang, Q., et al.: Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. 10(2), 12.1–12.19 (2019)
25. Konen, J., et al.: Federated learning: Strategies for improving communication efficiency. In: NIPS Workshop on Private Multi-Party Machine Learning, Barcelona (2016)
26. McMahan, H.B., et al.: Federated learning of deep networks using model averaging. NIPS, 1602.05629v, (2016)
27. Kumar, S., et al.: Federated control with hierarchical multi-agent deep reinforcement learning. NIPS:1712.08266, (2017)
28. Shi, W., et al.: On the linear convergence of the ADMM in decentralized consensus optimization. IEEE Trans. Signal Process. 62(7), 1750–1761 (2014)
29. Wang, H., et al.: Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. MICCAI - International Conference on Medical Image Computing and Computer-Assisted Intervention. Toronto, Canada. vol. 14(Pt 3), p. 115 (2011)
30. Zhu, Q., et al.: Multi-modal AD classification via self-paced latent correlation analysis. Neurocomputing 355, 143–154 (2019)
31. Zhu, Q., et al.: Latent correlation embedded discriminative multi-modal data fusion. Signal Process. 171, 107466 (2020)
32. Hou, J.C., et al.: Audio-visual speech enhancement using multimodal deep convolutional neural networks. IEEE Trans. Emerg. Top. Comput. Intell. 2(2), 117–128 (2018)
33. Jiang, A., et al.: Induction motor fault diagnosis based on multimodal stacked auto-encoder. J. Electron. Meas. Instrum. 32(8), pp. 17–23 (2018)
34. Lang, H.: Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks. In: AVEC’15: Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge. Brisbane, Australia, pp. 73–80 (2015)
35. Soriano, A., et al.: Fusion of scores in a detection context based on alpha integration. Neural Comput. 27(9), 1–28 (2015)

How to cite this article: Wei, X.: A multi-modal heterogeneous data mining algorithm using federated learning. J. Eng. 1–9 (2021).
https://doi.org/10.1049/tje2.12049.