Wearable Sensor based Multimodal Human Activity Recognition Exploiting the Diversity of Classifier Ensemble

Haodong Guo, Ling Chen, Liangying Peng and Gencai Chen
College of Computer Science, Zhejiang University
Hangzhou, China
{hdguo, lingchen, lyoare, chengc}@zju.edu.cn

ABSTRACT
Effectively utilizing multimodal information (e.g., heart rate and acceleration) is a promising way to achieve wearable sensor based human activity recognition (HAR). In this paper, an activity recognition approach MARCEL (Multimodal Activity Recognition with Classifier Ensemble) is proposed, which exploits the diversity of base classifiers to construct a good ensemble for multimodal HAR, and the diversity measure is obtained from both labeled and unlabeled data. MARCEL uses neural network (NN) as base classifiers to construct the HAR model, and the diversity of classifier ensemble is embedded in the error function of the model. In each iteration, the error of the model is decomposed and back-propagated to base classifiers. To ensure the overall accuracy of the model, the weights of base classifiers are learnt in the classifier fusion process with sparse group lasso. Extensive experiments show that MARCEL is able to yield a competitive HAR performance, and has its superiority on exploiting multimodal signals.

Author Keywords
Activity recognition; diversity; classifier ensemble.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Human activity recognition using wearable sensors is one of the most significant and valuable issues of wearable computing, especially for user-centric mobile applications, e.g., healthcare, rehabilitation, and gaming. Traditionally, researchers attempt to recognize human activity only from accelerometer data [1, 2, 3] or other data (e.g., electrocardiogram [4]). However, due to the constraint of single context information, such as noisy data or sensor variations [5], those approaches are hardly to achieve reliable performance in real-world applications.

Existing efforts on utilizing multimodal information for wearable sensor based HAR can be generally divided into two categories, i.e., feature based ensemble [6, 7, 8, 9] and classifier based ensemble [10, 11]. The former category mainly focuses on concatenating multimodal handcrafted features into a long feature vector, and then employs a single classifier to achieve HAR. Multimodal features may bring additional physiological or environmental cues for a pattern, but the feature compatibility problem is left to be solved [10]. For example, physiological data, compared with accelerometer data, always embed with to some extent temporal history. This makes it hard for a single HAR classifier to discriminate the activity patterns across modalities.

The latter category, as an alternative solution, is to combine the base classifiers built on different modalities separately. So the feature compatibility issues among multiple modalities can be avoided [11], and the robustness to noisy data is also improved in classifier ensemble [12]. Nevertheless, researchers in HAR still have to face following problems. First, most methods usually train the modality classifiers to emphasis on coincident modality output [13], while tend to overlook the diversity across modalities. Second, no attention has been paid to the problem of evaluating each feature group from modality data, and selecting relevant features for multimodal HAR.

Moreover, to extract robust features for multimodal data is still a problem shared by both categories. As some modality data (e.g., physiological data) have not been well investigated on feature construction for HAR, the researchers have to either design new features [8] or employ related measures from other research fields [11, 14]. However, developing domain-specific features for each modality is expensive, time-consuming, and requires expertise of the data. It is highly desirable to employ an effective method to acquire and extract representative features.

Recently, as a new feature learning technique that can learn a hierarchy of features tuned to the task at hand, deep networks have been used to achieve state-of-the-art results

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
© 2016 ACM. ISBN 978-1-4503-4461-6/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971708

on a number of benchmark HAR datasets [15, 16, 17]. Meanwhile, diversity learning of classifier ensemble has shown its powerful ability to ensure that all the base classifiers craft uncorrelated errors [18, 19]. In order to build a good ensemble, it is necessary not only to build good base classifiers, but also the base classifiers must be diverse, this means that for the same instance, the base classifiers return different outputs and their errors should be in different instances [20]. It's also worth mentioning that unlabeled data, readily available in real-world tasks, can be utilized to enhance the diversity of classifier ensemble [13, 21].

To exploit the most advances in deep networks and diversity learning, we propose MARCEL (Multimodal Activity Recognition with Classifier Ensemble), a neural network based multimodal activity recognition approach. Aimed at addressing above-mentioned problems, MARCEL consider both diversity and accuracy in the error functions of our HAR model. By using NN as base classifiers, the diversity measure obtained from both labeled and unlabeled data can be decomposed and back-propagated to base classifiers. By such a way, all the networks can be trained simultaneously and interactively on the training dataset. The trained NNs would be able to automatically discover appropriate feature representation from original data. To further investigate the importance of each type of features and select the relevant features among or across modality data, sparse group lasso is utilized to fuse multimodal networks at classifier level.

The main contributions of this work are summarized as follows:

- 1) Propose a wearable sensor based multimodal HAR approach exploiting the diversity of classifier ensemble based on neural network. This method gains model benefits from both the most advances in neural networks and learning to diversity techniques. Thus it would be able to not only automatically discover appropriate feature representation but also utilize unlabeled data to generate diverse base classifiers.
- 2) Combine stacking framework with sparse group lasso to realize classifier level fusion, which can ensure the overall classification performance by selecting the relevant features and assigning different weights to different base classifiers.
- 3) Evaluate our approach on two benchmarked datasets and perform extensive comparison with other methods. The experimental results show that MARCEL is able to yield a competitive HAR performance, and has its superiority on exploiting multimodal signals.

The rest of the paper is organized as follows. Section II presents the existing approaches related with multimodal HAR and learning to diversity. Section III gives the proposed method. Section IV presents experimental results. Finally, Section V concludes the paper.

RELATED WORK

This section gives a review of the previous work related to wearable sensor based HAR, including HAR using single or multiple modality data, diversity learning of classifier ensemble, and structural sparse representation.

HAR Using Single Modality Data

Most of early researches attempt to recognize human activity only from accelerometer data [1, 2, 3, 5, 22, 23, 24] or other data (e.g., electrocardiogram [4]). The common setting of those wearable sensor based HAR systems [1, 3, 22, 24] is to using acceleration sensors located on fixed body part. This setting is experimentally shown to be good at detecting the activity patterns of several daily movements. To apply HAR in a ubiquitous environment, recent researchers are certainly preferable to replace traditional acceleration sensors with smartphones. For example, Kwapisz et al. [2] gave an exhaustive discussion on activity recognition using phone-based accelerometers. CenceMe application [23] is also an activity recognition engine with Nokia N95 phone.

The major flaw embedded in those approaches is that acceleration or physiological data only is hardly to achieve reliable performance in real-world applications, since the sensor variability (e.g., rotation and translation) sometimes greatly degrades the performance of wearable sensor based HAR [5]. Thus, how to utilize multimodal information for activity recognition becomes a hot research topic.

In addition, effective feature extraction from different modality data is not a trivial thing for HAR researchers. Currently, some modality data still have not been well investigated on feature construction for HAR. Hence, the researchers have to either design new features [8] or employ related measures from other research fields [14]. It is difficult to decide what features are important for the application at hand since the choice of features is highly problem-dependent. To address this limitation, automatically learning universal feature representation from raw data (e.g., deep networks) is an alternative solution [15, 16, 17].

Multimodal HAR

Existing efforts on utilizing multimodal information for wearable sensor based HAR can be generally divided into two categories, i.e., feature based ensemble [6, 7, 8, 9, 14, 25] and classifier based ensemble [10, 11].

Feature based ensemble is a straightforward consideration that combines multimodal information in feature level. Features extracted from different modality data are firstly concatenated into a long feature vector. For example, Kunze et al. [6] combined the features of accelerometers and gyroscopes. Similarly, combining accelerometers with physiological sensors [7, 8, 9], microphone [14], or location sensors [25] is also investigated to improve HAR performance. Then, a single classifier, such as support vector machines (SVM), Multiple Kernel Learning (MKL)

[26], and dynamic Bayesian network [9], is fed with the generated long feature vectors. Due to the feature compatibility issues arising from different time shifts, window length configurations, and sampling frequencies, a single classifier, however, cannot perform well with features from different sensors.

Classifier based ensemble aims to improve the generalization ability and accuracy of activity recognition by combining the base classifiers built on different modalities separately. Li et al. [10] fused the classification scores provided by SVM and gaussian mixture models built on different modality data. While Guo et al. [11] employed an adaptive stacking framework in classifier level ensemble. Due to uniform expressions (i.e., predictive scores) outputted from different classifiers, fusing in classifier level is proved to be able to address the feature compatibility problems. More importantly, classifier level fusion can adjust the classifier ensemble weights according to their classification accuracy [11]. Most classifier based ensemble methods in HAR focus their efforts on minimizing the classification loss. However, to overlook the diversity across modalities and only optimizing accuracy would inevitably lead the over-fitting of the base classifiers, i.e., generating correlated errors.

Learn to Diversity of Ensemble

Diversity learning of classifier ensemble has shown its powerful ability to ensure that all the individual classifiers craft uncorrelated errors. There exist two ways to exploit diversity of classifier ensemble, i.e., only optimizing diversity measure [18] and trading off between diversity and accuracy measures [21, 27, 28]. Since no research evidence has theoretically shown the clear correlation between diversity and accuracy in ensembles, most work in literature tend to make the tradeoff between diversity and accuracy measures, instead of only optimizing diversity measure. For instance, Yin et al. [27] attempted to utilize diversity for classifier selection and combination heuristically and iteratively. Yu et al. [29, 30] proposed the diversity regularized machine, which efficiently generates an ensemble of assorted SVMs. These work has effectively justify the usefulness of diversity learning to improve the performance of classifier ensemble.

To further exploit the diversity measure, some other researches are very insightful. Chen et al. [28] took diversity as regularized item to improve neural network ensembles. Unlabeled data was used in [13, 21] to enhance the diversity of classifier ensemble. Both of those methods make it possible for other researchers to utilize the diversity measure.

Structural Sparsity based Feature Evaluation

Sparse representation based feature selection can efficiently obtain a small subset from a large real world dataset, and meanwhile presents the importance of each type of features as sparsity coefficients [31, 32]. The popular lasso algorithm [33] aimed at encouraging the sparsity of feature coefficients by adding an additional L1-norm penalty on the widely used least squares loss. However, these lasso based methods mainly focused on the sparsity of the single basic element in the feature vector.

To utilize the group property of the concatenated group features (e.g., features extracted from multimodal data), structured sparsity based techniques have been proposed recently. For example, some researches extended the L1-norm in lasso to L1/Lq-norm (q>1) which facilitates group sparsity [34]. The group structure of the entire feature vector is considered in group lasso [35]. Zhao et al. [36] presented a group lasso model with an L2-norm regularization to further extend group lasso to sparse group lasso, which yields sparsity in intra-group and inter-group simultaneously. With sparse group lasso, not only some feature groups can be dropped but also some features within the remaining groups can be removed.

Inspired by the literatures, MARCEL exploit the diversity of classifier ensemble based on neural network in activity recognition. MARCEL firstly constructs a multimodal neural network ensemble, which would be able to not only automatically discover appropriate feature representation but also utilize unlabeled data to generate diverse base classifiers. To further investigate the importance of each type of features and select the relevant features among or across modality data, the latest sparse group lasso technique is employed in MARCEL.

Relationship with Other Similar Methods

Various types of deep networks have been successfully applied in behavior modeling, such as, RBM for skill assessment [16], CNN for activity recognition [17], and DNN for audio sensing [37]. In our work, neural networks are only taken as base classifiers to learn modality-specific feature representation.

To further utilize the benefit obtained from neural networks, many work used ensemble methods to improve the performance of a single model. Deng et al. [38] performed stacking for speech recognition at the most straightforward frame level and then fed the combined output to a separate HMM decoder. This kind of design is good at combining speech recognition error patterns generated from different networks, but is hardly to exploit multimodal data come from diverse sensors. In [39], a sparse filtering approach for neural networks classifier was used to decrease the length of neural network input. Differently, we use sparse group lasso technique to select the relevant features among or across modality data.

Diversity measure is used to improve the performance of classifier ensemble in our work. While, Zhang et al. [40] used diversity to achieve model compression rather than aiming to produce an accurate yet diverse ensemble of models.

METHODOLOGY

Wearable sensor based multimodal HAR, as an ensemble learning process, naturally bears two goals simultaneously, maximizing the generalization performance and minimizing the classification loss of base classifiers. To achieve the goals, MARCEL has two distinct process stages, as shown in Figure 1. Firstly, based on neural network ensemble, both the diversity and accuracy measures are employed to train a good ensemble (with accurate and diverse base classifiers). The diversity of base classifiers is able to enhance their generalization performance by minimizing uncorrelated errors. Secondly, the base classifiers are incorporated by classifier level fusion to ensure the overall classification performance.

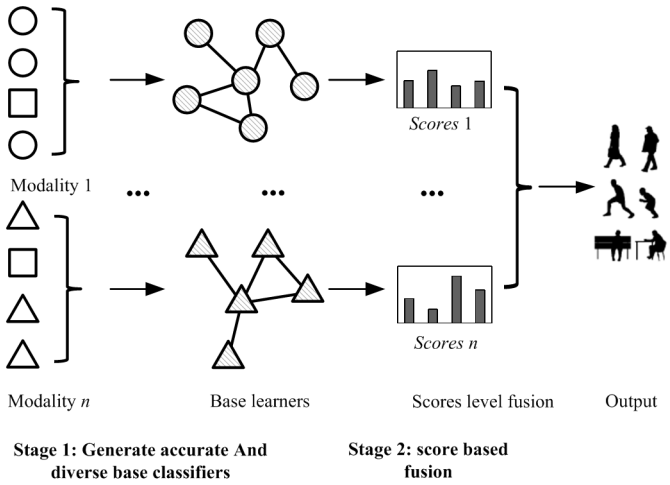

Figure 1. The workflow of MARCEL.

Problem Definition

Let $\mathcal{X}$ and $\mathcal{Y}$ denote the space of inputs and the set of class labels, respectively. $\mathcal{Y}=\{y_i \mid 1 \le i \le c\}$, where c is the number of class labels. Given training data set $\mathcal{TS}=\mathcal{L}\cup\mathcal{U}$, where $\mathcal{L}=\{(x_i,y_i) \mid 1 \le i \le L\}$ contains L labeled training examples and $\mathcal{U}=\{x_i \mid L+1 \le i \le L+U\}$ contains U unlabeled training examples, $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. In addition, we use $\tilde{\mathcal{L}}=\{x_i \mid 1 \le i \le L\}$ to denote the unlabeled dataset derived from $\mathcal{L}$.

Suppose the classifier ensemble is composed of m base classifiers $\boldsymbol{f}=\{f_k \mid 1 \le k \le m\}$, corresponding to different modal data. MARCEL maximizes the diversity and accuracy measures of the classifiers on the labeled data $\mathcal{L}$, as well as the diversity of the classifiers on the unlabeled data $\mathcal{D}=\tilde{\mathcal{L}}\cup\mathcal{U}$. The loss function of MARCEL is as follows:

$$E(\boldsymbol{f},\mathcal{L},\mathcal{D}) = E_y(\boldsymbol{f},\mathcal{L}) - \lambda \cdot E_d(\boldsymbol{f},\mathcal{D}) \quad (1)$$

where, the first term $E_y(\boldsymbol{f},\mathcal{L})$ corresponds to the classification loss for label prediction (e.g., logistic); the second term $E_d(\boldsymbol{f},\mathcal{D})$ corresponds to the diversity loss of $\boldsymbol{f}$ on a specified data set $\mathcal{D}$ (e.g., $\mathcal{D}=\mathcal{U}$); the non-negative parameter λ controls the trade-off between the two terms. Similar loss functions with the combination of a classification loss and a penalty term (i.e., diversity) have been well investigated in ensemble learning [13, 41]. Furthermore, $e(f_k,\mathcal{L})$ denotes the empirical loss of the k-th base classifiers f_k on the labeled dataset $\mathcal{L}$. Then, $E_y(\boldsymbol{f},\mathcal{L})$ can be calculated by:

$$E_y(\boldsymbol{f},\mathcal{L}) = \frac{1}{m}\cdot\sum_{k=1}^{m} e_k(f_k,\mathcal{L}) \quad (2)$$

Many diversity measures are extensively acknowledged in classifier ensemble, e.g., Disagreement [42], Q-Statistics, Double Fault [43], Kappa [44], and Prediction Confidence [21]. However, until now no diversity measure has research evidences theoretically showing its correlation with the ensemble accuracy except negative correlation learning (NCL) [28]. In this paper, $E_d(\boldsymbol{f},\mathcal{D})$ is calculated by NCL as follows:

$$E_d(\boldsymbol{f},\mathcal{D}) = \frac{1}{m}\cdot\sum_{k=1}^{m} e_d(f_k,\mathcal{D}) \quad (3)$$

where,

$$e_d(f_k,\mathcal{D}) = \frac{1}{2|\mathcal{D}|}\cdot\sum_{\boldsymbol{x}\in\mathcal{D}} (f_k(\boldsymbol{x})-\tilde{\boldsymbol{f}})^2 \quad (4)$$

And $\tilde{\boldsymbol{f}} = \frac{1}{m}\cdot\sum_{k=1}^{m} f_k$ is the combination formulation of $\boldsymbol{f}$.

The first goal of MARCEL is to train the target ensemble $\tilde{\boldsymbol{f}}$, which consists of accurate and diverse base classifiers, by minimizing the loss function in Equation (1) as follow:

$$\tilde{\boldsymbol{f}} \leftarrow \arg\min_{\boldsymbol{f}} E(\boldsymbol{f},\mathcal{L},\mathcal{D}) \quad (5)$$

Multimodal Neural Network Ensemble with Diversity

Existing efforts on utilizing multimodal information for activity recognition have to encounter the feature extraction step. However, some modality data still have not been thoroughly investigated on feature construction for HAR. Hence, the researchers have to either design new features [8] or employ related measures from other research fields [14]. To address the problem, MARCEL uses NN to implement the base classifiers. It has been extensively demonstrated that the most advances in neural networks (i.e., deep neural networks) can automatically discover appropriate feature representation.

As shown in Figure 2, for each modality of neural network ensemble, we employ a deep neural network with M layers of hidden units. Thus, the base classifier f_k ($1 \le k \le m$) is modeled as:

$$f_k(\boldsymbol{x}) = \mathcal{S}(\boldsymbol{w}_k^M \cdot h_k^M(\boldsymbol{x}) + b_k^M) \quad (6)$$

where, $\boldsymbol{w}_k^M$ and b_k^M are the weights and bias value for the output (M-th) layer, respectively. $\mathcal{S}$ is a non-linear squashing function, e.g., $sigmoid(t) = 1/(1+e^{-t})$ and $tanh(\cdot)$. Typically the j-th layer ($1 \le j \le M$) can be defined as:

$$h_k^j(\boldsymbol{x}) = \mathcal{S}(\boldsymbol{w}_k^{j-1} \cdot h_k^{j-1}(\boldsymbol{x}) + b_k^{j-1}),\ j > 1 \tag{7}$$

where,

$$h_k^1(\boldsymbol{x}) = \mathcal{S}(\boldsymbol{w}_k^1 \cdot \boldsymbol{x} + b_k^1) \tag{8}$$

Here, we take a standard fully connected deep neural network as an example. Other advance NN designs embedded with prior knowledge about a particular problem (i.e., convolutional networks [45, 46]) can also be employed in MARCEL.

Correspondingly, the first term $E_y(\boldsymbol{f}, \mathcal{L})$ in Equation (1) is set to be the mean square error function $MSE(f_k, \mathcal{L})$ on the labeled dataset $\mathcal{L}$, which is commonly used to measure the empirical loss of NN. By decomposing the loss function $E(\boldsymbol{f}, \mathcal{L}, \mathcal{D})$ into each neural network, the loss function of f_k is as follows:

$$e_k(\mathcal{L}, \mathcal{D}) = MSE(f_k, \mathcal{L}) - \lambda e_d(f_k, \mathcal{D}) \tag{9}$$

where,

$$MSE(f_k, L) = \frac{1}{2|\mathcal{D}|}\sum_{i=1}^{L}(f_k(\boldsymbol{x}_i) - y_i)^2 \tag{10}$$

Given the parameters of the neural networks, a forward propagation is first performed for one or a mini-batch of multimodal example(s), then the errors propagate backwards from the output nodes to the input nodes regarding the network's modifiable weights, and finally the weights are updated by a gradient descent step [46]. Taking derivatives of $e_k(\mathcal{L}, \mathcal{D})$ with respect to $\boldsymbol{w}_k$, the gradient is as follows:

$$\nabla \boldsymbol{w}_k = \frac{\partial e_k(\mathcal{L}, \mathcal{D})}{\partial \boldsymbol{w}_k} = \frac{\partial MSE(f_k, \mathcal{L})}{\partial \boldsymbol{w}_k} - \lambda \frac{\partial e_d(f_k, \mathcal{D})}{\partial \boldsymbol{w}_k} \tag{11}$$

where,

$$\frac{\partial MSE(f_k, \mathcal{L})}{\partial \boldsymbol{w}_k} = \frac{1}{|\mathcal{D}|} \cdot \sum_{i=1}^{L}(f_k(\boldsymbol{x}_i) - y_i)\frac{\partial f_k(\boldsymbol{x}_i)}{\partial \boldsymbol{w}_k} \tag{12}$$

$$\frac{\partial e_d(f_k, \mathcal{D})}{\partial \boldsymbol{w}_k} = (1 - \frac{1}{M}) \cdot \frac{1}{|\mathcal{D}|} \cdot \sum_{\boldsymbol{x} \in \mathcal{D}}(f_k(\boldsymbol{x}) - \tilde{\boldsymbol{f}}(\boldsymbol{x}))\frac{\partial f_k(\boldsymbol{x})}{\partial \boldsymbol{w}_k} \tag{13}$$

Note that we can deduce $\partial f_k(\boldsymbol{x})/\partial \boldsymbol{w}_k$ via classical back-propagation algorithm, which depends on the structure of the modality-specific NN. Therefore, we can evaluate the required derivatives $\nabla \boldsymbol{w}_k$, and apply the gradient descent step. In addition, Equation (13) is defined on unlabeled dataset, e.g., $\mathcal{D}=\tilde{\mathcal{L}} \cup \mathcal{U}$. Therefore, the diversity loss $E_d(\boldsymbol{f}, \mathcal{D})$ would be obtained from unlabeled data. The multimodal NN learning process of MARCEL is described in Algorithm 1. For every iteration, there are two kinds of typical steps, i.e., forward propagation (lines 4-6) and backward propagation (lines 7-11). Both labeled and unlabeled data are used to increase the diversity of multimodal NNs (lines 8-11). After this process, MARCEL not only can learn an optimal combination of multimodal representation, but more importantly the ensemble error (related with diversity and accuracy) is back-propagated to effectively and adaptively train the modality-specific neural networks.

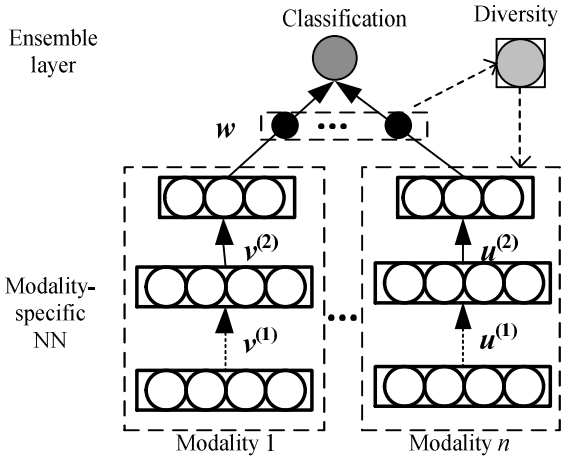

Figure 2. The NN architecture of MARCEL.

Algorithm 1 Multimodal NN learning process of MARCEL

Input: the training data set $\mathcal{TS} = \mathcal{L} \cup \mathcal{U}$,
the trade-off parameter $\lambda > 0$,
the number of iterations $\mathcal{T} > 0$.
Output: the set of modality-specific NNs: ***MNN***

1. initialize ***MNN*** as given or default network structures
2. $\tilde{\mathcal{L}} \leftarrow$ remove the labels of $\mathcal{L}$
3. **repeat**
4. **Each** f_k in ***MNN*** executes **Forward Propagation**:
5. use (6),(7) and (8).
6. calculate $\tilde{\boldsymbol{f}} = \frac{1}{m} \cdot \sum_{k=1}^{m} f_k$
7. **Each** f_k in ***MNN*** executes **Backward Propagation**:
8. If $\mathcal{D}=\tilde{\mathcal{L}}$
9. $\nabla \boldsymbol{w}_k \leftarrow$ use (11),(12) and (13)
10. Else if $\mathcal{D}=\mathcal{U}$
11. $\nabla \boldsymbol{w}_k \leftarrow$ use (11)(13) where $\frac{\partial MSE(f_k, L)}{\partial \boldsymbol{w}_k} = 0$
12. **until** converge or reach max iterations $\mathcal{T}$
13. **return** ***MNN***

Multimodal Feature Group Evaluation

To investigate the importance of each type of features and select the relevant features among or across modality data, we exploit the sparse group lasso for grouped feature selection.

Let $\mathcal{V} = \{(v_i, y_i) \mid 1 \le i \le L\}$ denote L labeled training examples, where $v_i = \{v_i^t \mid 1 \le t \le p\}$ is the extracted feature vectors obtained in previous section, y_i is the corresponding label vectors. Suppose the features vectors consist of k non-overlapping groups (corresponding to k modality data) and the length of the l-th feature group is G_l. In addition, we use $\alpha_j = \{\alpha_j^t \mid 1 \le t \le l\}$ to denote the coefficient vector for label y_j, α_j^t the coefficient vector for feature group G_l. Thus, the grouped feature selection problem for label y_j can be formulated as the following optimization task:

$$f(\alpha_j) = \min e(\alpha_j) + \lambda\varphi(\alpha_j) \tag{14}$$

where $e(\alpha_j)$ is the loss function, and $\varphi(\alpha_j)$ is the penalty terms. Here, we use the logistic loss in this problem:

$$e(\alpha_j) = \sum_{i=1}^{n} \log(1 + \exp(-y_j^i(\alpha_j^T v_i + c))) \tag{15}$$

where c is the intercept (scalar). The penalty term is formulated as:

$$\lambda\varphi(\alpha_j) = \lambda_1 \left\|\alpha_j\right\|_1 + \lambda_2 \sum_{l=1}^{k} w_l \left\|\alpha_j^l\right\|_2 \tag{16}$$

where λ_1 and λ_2 are regularization parameters, and w_l is the weight for the l-th sample. The first part of Equation (16) is a common L1-norm penalty, and the last part penalty is to encourage sparsity on the group level of features. These penalty terms can lead to sparsity in both inter-group and intra-group features. In other words, we can select the relevant features among or across modality data by filtering out features with small weights. In this paper, we use an efficient algorithm [47] with a linear time complexity to solve the optimization problem.

Multiple Modalities Combination with Selected Features

In the first process stage of MARCEL, accuracy and diversity measures are employed to train a good ensemble. Note that the ensemble is calculated by sum up the outputs of all the base classifiers, which is used for traditional classifier ensemble. All classifiers of the ensemble are built on the exactly same modality dataset. However, in MARCEL, multimodal networks are built on modality-specific datasets respectively. Simply summing up the outputs of all the base classifiers would degrade the final classification performance. Therefore, the second process stage of MARCEL aims to adjust the classifier ensemble weights according to the classification accuracy of the base classifiers. For instance, acceleration-NN is usually effective to discriminate "walking" and "running", and therefore it should has larger ensemble weight than other modality-specific NNs (e.g., electrocardiogram-NN).

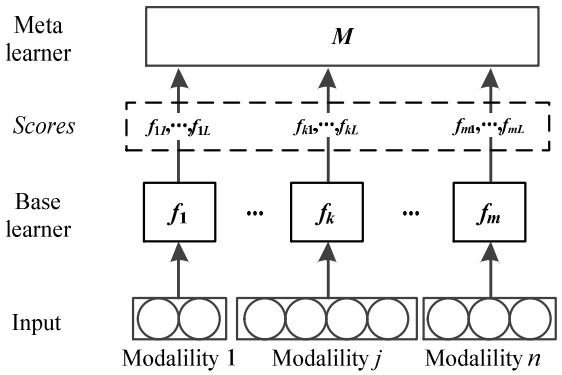

Figure 3. The classifier level fusion framework of a-stack.

Classifier level fusion is employed to ensure the overall classification performance. Considering the compatibility issues arising from different time shifts, window length configurations, and sampling frequencies, fusion multiple modalities in feature level would weaken the utility of other modality features. While fusing in classifier level, the problem can be addressed, because the outputs from different classifiers have uniform expressions, i.e., predictive scores. More importantly, classifier level fusion can adjust the classifier ensemble weights according to their classification accuracy. We utilize a-stack, an adaptive stacking framework [11], to combine the predictions of the base classifiers $\boldsymbol{f} = \{f_k \mid 1 \le k \le m\}$ obtained from the first stage of MARCEL.

Given training data set $\mathcal{TS} = \mathcal{L}$, and the base classifiers $\boldsymbol{f} = \{f_k \mid 1 \le k \le m\}$ with f_k: $\mathcal{X} \to \mathcal{Y}$, the output label is $y \leftarrow \arg\max f_{k,l}(x)$, $f_{k,l}(x)$ is the predictive score returned by the classifier f_k when the input x is labeled with y. And the score $f_{k,l}(x)$ is usually the conditional probability $p_i(y|x)$ in statistical machine learning methods. As shown in Figure 3, A-stack tries to find a meta classifier M: $\mathcal{S} \to \mathcal{Y}$, where $\mathcal{S}$ = {$f_{11}(x)$, ..., $f_{1L}(x)$, ..., $f_{k1}(x)$, ..., $f_{kj}(x)$, ..., $f_{kL}(x)$,…, $f_{m1}(x)$, ..., $f_{mL}(x)$}, $f_{kj}(x)$ is the predictive scores returned by the kth base classifier for the class label j for $1 \le k \le m$ and $1 \le j \le L$. To combine multimodal information properly, the base classifiers are built on their data separately and then fuse their scores in a meta level classifier.

EXPERIMENTS

In this section, we present the experiments to evaluate the proposed method. Firstly, we describe the experimental setup and the utilized datasets. Secondly, we investigate the impact of parameters (e.g., the diversity control parameter λ) on the classification performance. Thirdly, experiments are further conducted to show whether unlabeled data benefits the performance of activity recognition. Then, we study the effects of other popular diversity measures (e.g., disagreement and double-fault measures) on the performance of MARCEL. Finally, we extensively compare

our method with several state-of-the-art activity recognition methods.

Experimental Setting and Datasets

Two benchmarked real-world datasets on activity recognition are utilized in our experiments. To the best of our knowledge, they are the latest available wearable sensor based multimodal datasets with complete annotation process. Both of which contain multiple modality data (e.g., physiological and inertial data). The statistics of the two datasets are summarized in Table 1.

1) PAMAP2 Dataset [48] contains data of 18 different human activities (e.g., walking, cycling, and playing soccer, etc.), performed by 9 subjects wearing 3 inertial measurement units (IMU) and a heart rate monitor. The three IMUs are attached over the chest, dominant wrist, and dominant ankle of subjects, respectively. Each IMU contains two 3-axis accelerometers, a 3-axis gyroscope, and a 3-axis magnetic sensor, all sampled at 100Hz. The heart rate chest strap records heart rate values with approximately 9Hz.

2) MHEALTH [49] dataset comprises inertial and physiological sensor recordings for 10 subjects while performing 10 human activities. Sensors placed on the subject's chest, right wrist, and left ankle are used to measure the motion experienced by diverse body parts, i.e., acceleration, angular velocity, and magnetic field orientation. The sensor positioned on the chest also provides 2-lead ECG measurements. All sensing modalities are recorded at a sampling rate of 50Hz.

	PAMAP2	MHEALTH
# of activity	18	10
# of subjects	1 Female/8 Males	10 volunteers
Sensors	3 IMUs and a heart rate monitor	3 IMUs and a 2-lead ECG monitor
# of modalities	4(acceleration, angular velocity, magnetic field, and heart rate)	4(acceleration, angular velocity, magnetic field, and ECG)
Labeled/total duration	27248.27/ 38504.91 seconds	6863.9 / 24314.9 seconds

Table 1. The statistics of the utilized datasets.

We employ the datasets to perform background activity recognition task [49], which categorizes the activities into 6 classes, including lying, sitting/standing, walking, running, cycling, and other (the remaining activities except the above six classes). The idea behind the definition of this task is that users always perform meaningful activities, and ignoring these other activities would limit the applicability of activity recognition methods. More importantly, the complexity of the classification problem would significantly increase in this way [48]. Thus, the task is employed to evaluate our method.

A sliding window strategy with fixed overlap is utilized to extract subsequences from the original data. Based on the experience gained in [11, 48, 49], the window size and sliding step are set as 5.12s and 1s, respectively, which are fixed in the following experiments. The data from the two datasets were processed in our activity recognition chain separately. Note that the raw subsequences extracted by sliding window are used as the input of MARCEL. Thus, the input layer size of modality-specific networks is same as the window size, and the output layer size of base classifier is same as the total class number (i.e., 6). Another significant parameter of MARCEL is the hidden layer number. In practice, we find that setting 3 hidden layers is usually enough, and larger hidden layer number does not bring noticeable improvements. The size of each hidden layer is set to be a random value between the input layer size and the output layer size. Logistic regression is employed as the meta level classifier of MARCEL. For the sparse group lasso algorithm, we use a linear time complexity algorithm implemented in SLEP toolbox [50].

The commonly used performance measure, accuracy, is used for evaluating classification algorithm. All results are evaluated by Leave-One-Subject-Out (LOSO) validation, which is regarded as the standard evaluation strategy of activity recognition methods. LOSO validation can ensure subject independent in the evaluation process.

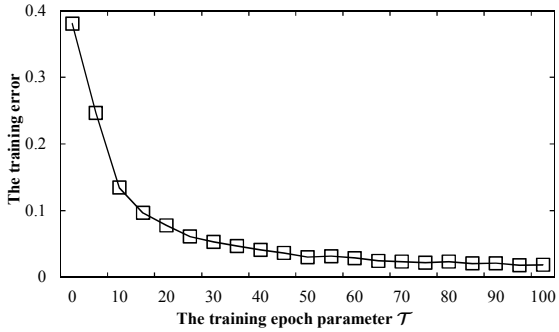

Figure 4. The impact of training epoch parameter $\mathcal{T}$ on MHEALTH.

Experiment 1: the impact of parameters

To investigate the impact of the trade-off parameter λ on the performance of MARCEL, we firstly tune the maximum gradient descent steps $\mathcal{T}$ of base networks with only classification loss (without diversity measure). The setting containing all modalities in MHEALTH (i.e., "Acc+Gyro+Magn+ECG") is used and the trade-off parameter λ is set to 0, where only classification loss is utilized to train the networks. Figure 4 gives the training error of MARCEL's base networks varying along with the epoch of the training stage. The training error is gradually stable when the epoch is large enough (epoch>50). This usually indicates that the parameters of the network have

been close to converge. Thus, we take $\mathcal{T}$ =50 in following experiments.

Then, we increase the value of λ from 0 to 2 in this experiment, with a fixed step of 0.1. The setting containing all modalities in MHEALTH (i.e., "Acc+Gyro+Magn+ECG") is selected to demonstrate the impact of parameter λ. The data of three subjects is used for the LOSO validation. The classification results on the dataset are presented in Figure 5. It can be observed that in every dataset, there is a value of λ under which the best classification performance can be achieved.

The parameter λ is used to control the diversity of the classifier ensemble. It can be found that when λ increases, there is an increasing phase for the classification performance and then a decreasing phase is followed. This might be because when λ is very small, all base classifiers tend to generate correlated errors. In an extreme case, ensembling all "same" base classifiers would not lead to performance improvement. On the other hand, when λ is very large, the diversity measure is dominant over the accuracy measure. In an extreme case, the final decision is "confused" when all base classifiers vote different opinions. Fortunately, there is a preferable balance point, i.e., λ=0.7.

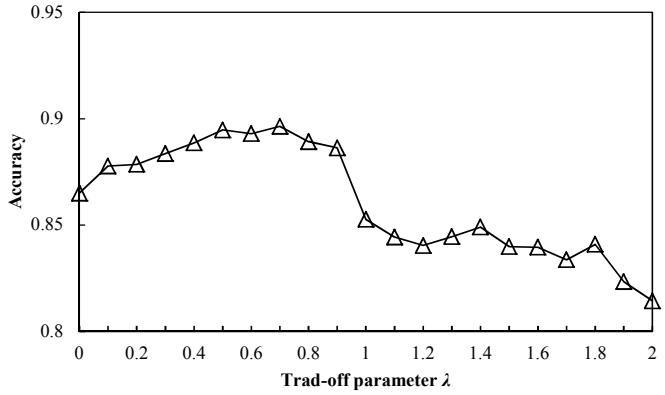

Figure 5. The impact of the diversity control parameter λ on MHEALTH.

Experiment 2: the feature extraction capability of the base classifiers in MARCEL

To evaluate the capability of the multimodal neural networks in feature extraction, we compare the performance of the base classifiers in MARCEL with that of traditional handcrafted features (THF) on two multimodal datasets. For MARCEL, only classification loss is utilized to train the networks (i.e., λ=0). The employed features of THF include time-domain features, e.g., mean, variance, standard deviation, median, maximum, minimum, root mean square, correlation between two axes, zero crossing rate, skewness, and kurtosis, as well as frequency-domain features, e.g., entropy and spectral entropy. Both compared methods use SVM as the final classifier.

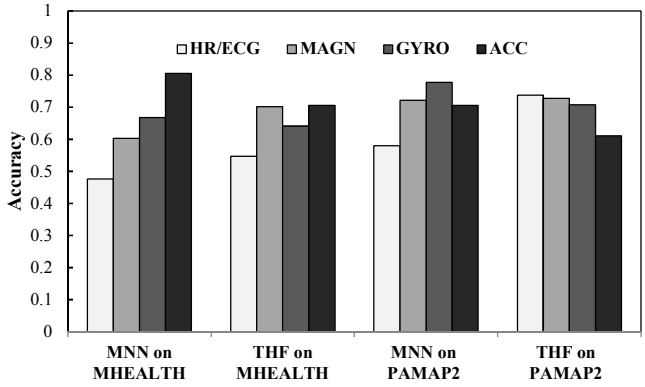

Figure 6. The performance of base classifiers (MNN) and traditional handcrafted features (THF) on both given datasets (MHEALTH and PAMAP2).

As shown in Figure 6, the result of MNN built on different single modality data is comparable in certain respects with that of THF on two multimodal datasets. It suggests that the base classifiers in MARCEL have the ability to automatically learn discriminative feature representations for HAR.

Experiment 3: the benefit from unlabeled data

To quantify the benefits obtained from the diversity of classifier ensemble in multimodal activity recognition, especially from unlabeled data, LOSO validation is adapted by splitting out partial labeled training data as unlabeled training data $\mathcal{U}$, thus $\mathcal{TS} = \mathcal{L} \bigcup \mathcal{U}$. The splitting rate of unlabeled data is defined as $\eta = |\mathcal{U}| / (|\mathcal{L}| + |\mathcal{U}|)$ .The unlabeled training data $\mathcal{U}$ aims to augment the diversity among base classifiers in MARCEL. And the best trade-off parameter is used (λ=0.7). Figure 7 reports the average performance improvement obtained from unlabeled data under various unlabeled data sizes, i.e., η . The range of η is {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}.

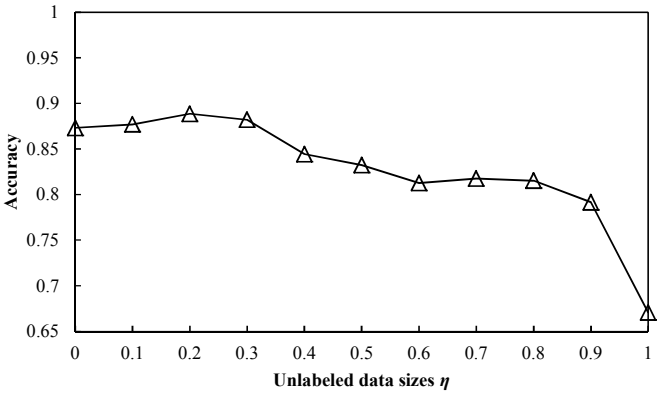

Figure 7. The benefit obtained from unlabeled data under various unlabeled data sizes on MHEALTH.

As shown in Figure 7, following tendencies could be discerned from the results:

1) The performance increases gracefully with the increase of unlabeled data sizes (η range from 0.1 to 0.3). This suggests that the unlabeled data do help to improve the performance in some way.

2) However, the performance begins to degrade when η is greater than 0.4. This indicates that the diversity measure can only ensure the overall performance in some way, which justifies the necessary to optimize diversity and accuracy measures in the training process simultaneously. Because the clear correlation between diversity and accuracy in ensembles is still uncertain [27].

Experiment 4: other diversity measures

To evaluate the usefulness of different diversity measures, we adapt three typical diversity measures (i.e., Disagreement, Double Fault, and Prediction Confidence [21]) on our method. The benefit obtained from the three measures is compared with NCL, which is used in this paper. The trade-off parameter λ is set to 0.7. Figure 8 reports the average accuracy improvement (by subtracting the accuracy under η=0, λ=0) obtained from different diversity measures under unlabeled data $\eta = 0.3$.

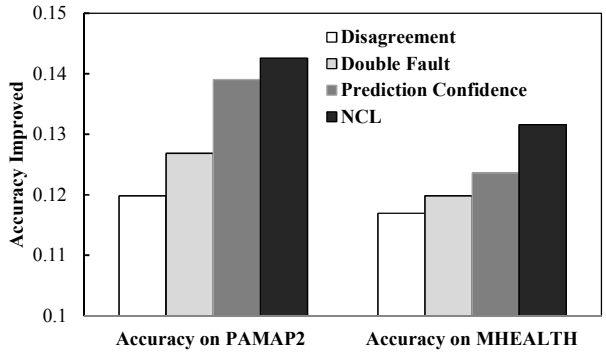

Figure 8. The average performance improvement obtained from different diversity measures on both given datasets (MHEALTH and PAMAP2).

As shown in Figure 8, all of those diversity measures help improve the classification performance of multimodal activity recognition. However, NCL tends to achieve the best result, followed by Prediction Confidence. This might be because most existing diversity measures (e.g., Disagreement and Double Fault) are calculated based on the binary (correct/incorrect) outputs of the base classifiers. Differently, both NCL and Prediction Confidence are able to utilize the prediction difference calculated based on the concrete output (score). However, none of those diversity measures has research evidences theoretically showing its correlation to the ensemble accuracy except NCL (the theoretical analysis is in [28]).

Experiment 5: comparison with other methods

To evaluate the effectiveness and show its competitive performance of the proposed approach, MARCEL is compared with the other three state-of-the-art activity recognition approaches (i.e., FEM [7], CEM [11], and MKL [26]). These comparative methods are elaborately chosen for fair comparisons. The comparison between classical feature based ensemble method (i.e., FEM) and MARCEL aims to test the ability to learn useful feature representations. While, the comparison between classifier level ensemble method (i.e. CEM) and MARCEL intends to show the ability of diversity learning to ensure that all the base classifiers craft uncorrelated errors. MARCEL is also compared with the MKL method to show its superiority on multimodal HAR.

In addition, two baseline methods are used to demonstrate the gain comes from MARCEL structure. 1) FEM2+SVM. Different from FEM, FEM2 is to use modality-specific features based on sensor type rather than the same set of features for all modalities. So, an additional feature selection step is utilized in FEM2 to discover modality specific features in its training stage. Sequential floating forward selection algorithm [45] is chosen to select out features for each modality. The comparison between FEM2+SVM (a non-NN approach) and the proposed method intends to show the gain obtained from the use of NNs and the modifications to the training of the ensemble. 2) SingleNN. The comparison between a single large NN (SingleNN) with MARCEL is to quantify the benefit coming from the ensemble of diversity learning.

In the comparative studies, only labeled samples are utilized, i.e., $\mathcal{TS} = \mathcal{L}$, $\mathcal{U}=\varnothing$. Total six kinds of combination settings are presented for each dataset (Acc, Gyro, Magn, HR/ECG denote the four kinds of modalities as shown in Table 1). For example, "Acc+Magn" means both acceleration and magnetic field are utilized to recognize activities. For MARCEL, the best trade-off parameter is used (λ=0.7) for both datasets. The employed features of FEM include time-domain features, e.g., mean, variance, standard deviation, median, maximum, minimum, root mean square, correlation between two axes, zero crossing rate, skewness, and kurtosis, as well as frequency-domain features, e.g., entropy and spectral entropy. For MKL method, we set an independent kernel for each modality feature group. A relatively optimal kernel was allocated to each modality. For SingleNN, all modalities data are used as the input and its size is similar to the sum of all ensemble base deep nets (in terms of layers and units). For the other compared methods, default parameters suggested in literatures are adopted.

The classification performances of these methods on two datasets with six kinds of combination settings are shown in Table 2. On each setting of these methods, the mean as well as the standard deviation of accuracy are recorded. Furthermore, to statistically measure the significance of performance difference, pairwise *t*-tests at 95% significance level are conducted between the methods. From Table 2, following tendencies could be discerned for all settings:

	Methods					
Modalities in datasets	**MARCEL**	**FEM+SVM**	**AStack(CEM)**	**FEM+MKL**	**FEM2+SVM**	**SingleNN**
PAMAP2:						
Acc	**0.724±0.116**	0.610±0.044*	0.702±0.235*	0.705±0.122*	0.676±0.080*	0.628±0.131*
Acc+HR	**0.795±0.114**	0.690±0.070*	0.711±0.229*	0.734±0.119*	0.705±0.200*	0.671±0.093*
Acc+Magn	**0.828±0.098**	0.772±0.084*	0.807±0.092*	0.808±0.077*	0.754±0.216*	0.656±0.257*
Acc+Gyro	**0.805±0.118**	0.693±0.213*	0.717±0.055*	0.754±0.099*	0.737±0.214*	0.731±0.141*
Acc+Gyro+Magn	**0.828±0.115**	0.764±0.221*	0.810±0.086	0.816±0.072	0.798±0.186*	0.720±0.137*
Acc+Gyro+Magn+HR	**0.848±0.086**	0.765±0.227*	0.811±0.085*	0.819±0.072*	0.802±0.216*	0.793±0.225*
MHEALTH:						
Acc	**0.869±0.064**	0.641±0.045*	0.735±0.061*	0.811±0.061*	0.752±0.073*	0.778±0.058*
Acc+ECG	**0.902±0.042**	0.711±0.036*	0.790±0.033*	0.846±0.086*	0.801±0.099*	0.791±0.048*
Acc+Magn	**0.917±0.049**	0.708±0.032*	0.732±0.034*	0.876±0.066	0.790±0.086*	0.884±0.044*
Acc+Gyro	**0.900±0.083**	0.671±0.048*	0.700±0.040*	0.878±0.054*	0.722±0.072*	0.775±0.074*
Acc+Gyro+Magn	**0.919±0.056**	0.708±0.033*	0.712±0.038*	0.896±0.097	0.821±0.024*	0.865±0.092*
Acc+Gyro+Magn+ECG	**0.923±0.057**	0.707±0.033*	0.748±0.045*	0.906±0.080*	0.882±0.007*	0.891±0.099*
win/tie/loss	**/**	**12/0/0**	**11/1/0**	**9/3/0**	**12/0/0**	**12/0/0**

Table 2. The classification accuracy of all compared methods (mean±std.); Total six kinds of combination settings are presented for each dataset (Acc, Gyro, Magn, HR/ECG denote the four kinds of modalities as shown in Table 1); for example, "Acc+Magn" means both acceleration and magnetic field are utilized to recognize activities;* indicates whether MARCEL is statistically superior to the compared method (pairwise t-test at 95% significance level).

1) First, the classification performance of CEM is high than that of FEM in most cases but still statistically inferior to that of MARCEL. This result justifies the conclusion of [11] that classifier level fusion is able to ensure the overall classification performance by avoiding feature compatibility issues. However, CEM usually over-fits to generate correlated errors. On the contrary, MARCEL can ensure that all the individual classifiers craft uncorrelated errors by exploiting the diversity of classifier ensemble.

2) Second, the superiority of MARCEL over the compared methods (i.e., FEM, CEM, and MKL) increases along with the number of modalities (from one to four kinds of modalities). This indicates that, as the number of modalities increases, both the feature compatibility issues (in FEM and MKL) and the classifier correlated errors (in CEM) deteriorate gradually. MARCEL could effectively utilize the diversity of classifier ensemble to benefit activity recognition from the multimodal data.

3) Third, the superiority of MARCEL over the baseline methods (i.e., FEM2 and SingleNN) exists in all cases. The results compared with SingleNN demonstrate that MARCEL benefits a lot from the ensemble of diversity learning. The performance of MARCEL obtained from the use of NNs and the modifications to the training of the ensemble is also justified by comparing with FEM2.

CONCLUSIONS

Effectively utilize multimodal information is a promising issue in wearable sensor based HAR. Existing efforts try to combine multiple modalities by feature or classifier ensemble. However, the feature construction and classifier combination strategy are still intractable. To address those problems, we propose MARCEL, a novel activity recognition approach exploiting the diversity of classifier ensemble based on neural network. Inspired by leaning to diversity techniques, our approach exploits the diversity of the base classifiers to construct a good ensemble for multimodal HAR. Firstly, the diversity of classifier ensemble is embedded in the error function of neural network (NN) and back-propagated to enhance the diversity and accuracy of the learned multimodal representation. Then, the base classifiers are incorporated by classifier level fusion to adjust the classifier ensemble weights. Extensive experiments show that MARCEL is able to yield a competitive classification performance in human activity recognition task, and has its superiority on exploit the benefit of multimodal signals.

Currently, the applicability of the model to actual hardware remains uncertain. Next, we will extend our work in the following directions: Firstly, evaluate the wearable performance of MARCEL to show its applicability in actual hardware; Secondly, utilize unsupervised techniques (e.g., stacked convolutional auto-encoders) to further improve the performance of our model; Thirdly, investigate the effects of multiple diversity ensemble (combining multiple diversity measures) on the performance of HAR.

ACKNOWLEDGMENTS

This work was funded by the Ministry of Industry and Information Technology of China (No. 2010ZX01042-002-003-001), China Knowledge Centre for Engineering Sciences and Technology (No. CKCEST-2014-1-5), the National Natural Science Foundation of China (Nos. 60703040 and 61332017), the Science and Technology Department of Zhejiang Province (Nos. 2011C13042 and 2015C33002).

REFERENCES

1. Bao, L., Intille, S. Activity recognition from user-annotated acceleration data. In *Proceedings of International Conference on Pervasive Computing* (2004), 1-17.
2. Kwapisz, J., Weiss, G., Moore, S. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* (2011), 12(2), 74-82.
3. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L. Activity recognition from accelerometer data. In *Proceedings of AAAI* (2005), 1541-1546.
4. Pawar, T., Chaudhuri, S., Duttagupta, S. P. Body movement activity recognition for ambulatory cardiac monitoring. *IEEE Transactions on Biomedical Engineering* (2007), 54(5), 874-882.
5. Guo, H., Chen, L., Chen, G., Lv, M. Smartphone-based activity recognition independent of device orientation and placement. *International Journal of Communication Systems* (2015).
6. Kunze, K., Lukowicz, P. Dealing with sensor displacement in motion-based onbody activity recognition systems. In *Proceedings of International Conference on Ubiquitous Computing* (2008), 20-29.
7. Lara, Ó.D., Pérez, A.J., Labrador, M.A., Posada, J.D. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and Mobile Computing* (2012), 8(5), 717-729.
8. Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., Korhonen, I. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine* (2006), 10(1), 119-128.
9. Tapia, E.M., Intille, S.S., Haskell, W., Larson, K., Wright, J., King, A., Friedman, R. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Proceedings of IEEE International Symposium on Wearable Computers* (2007), 37-40.
10. Li, M., Rozgic, V., Thatte, G., Lee, S., Emken, B.A., Annavaram, M., Narayanan, S. Multimodal physical activity recognition by fusing temporal and cepstral information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2010), 18(4), 369-380.
11. Guo, H., Chen, L., Shen, Y., Chen, G. Activity recognition exploiting classifier level fusion of acceleration and physiological signals. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (2014), 63-66.
12. Goswami, G., Mittal, P., Majumdar, A., Vatsa, M., Singh, R. Group sparse representation based classification for multi-feature multimodal biometrics. *Information Fusion* (2015).
13. Zhang, M.L., Zhou, Z.H. Exploiting unlabeled data to enhance ensemble diversity. *Data Mining and Knowledge Discovery* (2013), 26(1): 98-129.
14. Subramanya, A., Raj, A., Bilmes, J., Fox, D. Recognizing activities and spatial context using wearable sensors. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence* (2006).
15. Plötz, T, Hammerla, N., Olivier, P. Feature learning for activity recognition in ubiquitous computing. In *Proceedings of International Joint Conference on Artificial Intelligence* (2011), 22(1): 1729.
16. Ladha, C., Hammerla, N. Y., Olivier, P., Plötz, T. ClimbAX: skill assessment for climbing enthusiasts. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2013), 235-244.
17. Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of International Joint Conference on Artificial Intelligence* (2015), 25-31.
18. Hu, Q., Li, L., Wu, X., Schaefer, G., Yu, D. Exploiting diversity for optimizing margin distribution in ensemble learning. *Knowledge-Based Systems* (2014), 67, 90-104.
19. Yin, X. C., Huang, K., Yang, C., Hao, H. W. Convex ensemble learning with sparsity and diversity. *Information Fusion* (2014), 20, 49-59.
20. Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C. I., Kuncheva, L. I. Diversity techniques improve the performance of the best imbalance learning ensembles. *Information Sciences* (2015), 325, 98-117.
21. Zhang, M., Zhou, Z. Exploiting unlabeled data to enhance ensemble diversity. In *Proceedings of IEEE International Conference on Data Mining* (2010), 619-628.
22. Krishnan, N., Colbry, D., Juillard, C., Panchanathan, S. Real time human activity recognition using tri-axial accelerometers. In *Proceedings of Sensors, Signals and Information Processing Workshop* (2008).
23. Miluzzo, E., Lane, N. D., Fodor, K., Peterson, R., Lu, H. Sensing meets mobile social networks: The design, implementation and evaluation of the CenceMe application. In *Proceedings of ACM Conference on Embedded Network Sensor Systems* (2008), 337-350.
24. Gupta, P., Dallas, T. Feature Selection and Activity Recognition System using a Single Tri-axial Accelerometer. *IEEE Transactions on Biomedical Engineering* (2014), 61(6), 1780-1786.

25. Zhu, C., Sheng, W. Realtime recognition of complex human daily activities using human motion and location data. *IEEE Transactions on Biomedical Engineering* (2012), 59(9), 2422-2430.
26. Althloothi, S., Mahoor, M. H., Zhang, X., Voyles, R. M. Human activity recognition using multi-features and multiple kernel learning. *Pattern Recognition* (2014), 47(5), 1800-1812.
27. Yin, X., Huang, K., Hao, H., Iqbal, K., Wang, Z. A novel classifier ensemble method with sparsity and diversity. *Neurocomputing* (2014), 134, 214-221.
28. Chen, H. Diversity and regularization in neural network ensembles. *School of Computer Science University of Birmingham, PhD Thesis October* (2008).
29. Yu, Y., Li, Y.F., Zhou, Z.H. Diversity regularized machine. In *Proceedings of International Joint Conference on Artificial Intelligence* (2011), 1603-1608.
30. Li, N., Yu, Y., Zhou, Z.H. Diversity regularized ensemble pruning. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (2012), 330-345.
31. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A. Y. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of International Conference on Machine Learning* (2007), 759-766.
32. Bhattacharya, S., Nurmi, P., Hammerla, N., Plötz, T. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing* (2014), 15, 242-262.
33. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996, 267-288.
34. Liu, J., Ji, S., Ye, J. Multi-task feature learning via efficient l 2, 1-norm minimization. In *Proceedings of the twenty-fifth Conference on Uncertainty in Artificial Intelligence* (2009), 339-348.
35. Yuan, M., Lin, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 68(1), 49-67.
36. Zhao, L., Hu, Q., Wang, W. Heterogeneous Feature Selection with Multi-Modal Deep Neural Networks and Sparse Group Lasso. *IEEE Transactions on Multimedia* (2015), 17(11), 1936-1948.
37. Lane, N D., Georgiev, P., Qendro, L. DeepEar: robust smartphone audio sensing in unconstrained acoustic environments using deep learning. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), 283-294.
38. Deng, L., Platt, J C. Ensemble deep learning for speech recognition. In *Proceedings of INTERSPEECH* (2014), 1915-1919.
39. Romaszko, L. A deep learning approach with an ensemble-based neural network classifier for black box icml 2013 contest. *Workshop on Challenges in Representation Learning, ICML* (2013).
40. Zhang, X., Povey, D., Khudanpur, S. A Diversity-Penalizing Ensemble Training Method for Deep Learning. In *Proceedings of INTERSPEECH* (2015).
41. Opitz, D.W. Feature selection for ensembles. In *Proceedings of AAAI/IAAI* (1999), 379-384.
42. Skalak, D.B. The sources of increased accuracy for two proposed Boosting algorithms. In *Proceedings of AAAI, Integrating Multiple Learned Models Workshop* (1996), 1129, 1133-1133.
43. Shipp, C.A., Kuncheva, L. Relationship between combination methods and measures of diversity in combining classifiers. *Information Fusion* (2002), 3, 135-148.
44. Kuncheva, L., Whitaker, C. Measures of diversity in classifier ensembles. *Machine Learning* (2003), 51(2), 181-207.
45. LeCun, Y., Bottou, L., Bengio, Y., Haffner P. Gradient-Based Learning Applied to Document Recognition. In *Proceedings of the IEEE* (1998), 86(11), 2278-2324.
46. LeCun, Y., Bottou, L., Orr, G., Müller, K. Efficient backprop. *Neural networks: Tricks of the Trade* (2012), 9-48.
47. Liu, J., Ye, J. Moreau-Yosida regularization for grouped tree structure learning. In *Proceedings of Advances in Neural Information Processing Systems* (2010), 1459-1467.
48. Reiss, A., Stricker, D. Introducing a new benchmarked dataset for activity monitoring. In *Proceedings of IEEE International Symposium on Wearable Computers* (2012), 108-109.
49. Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., ... Villalonga, C. mHealthDroid: A Novel Framework for Agile Development of Mobile Health Applications. In *Proceedings of Ambient Assisted Living and Daily Activities* (2014), 91-98.
50. Liu, J., Ji, S., Ye, J. SLEP: Sparse learning with efficient projections. *Arizona State University* (2009), 6, 491.
51. Pudil, P., Ferri, F. J., Novovicova, J., Kittler, J. Floating search methods for feature selection with nonmonotonic criterion functions. In *Proceedings of IAPR International Conference on Pattern Recognition* (1994), 2, 279-283.