IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 13, 2018, accepted September 11, 2018, date of publication September 20, 2018, date of current version October 17, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871349

Domain Specific Learning for Sentiment Classification and Activity Recognition

HONG-BO WANG, (Member, IEEE), YANZE XUE, XIAOXIAO ZHEN, AND XUYAN TU
School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Hong-Bo Wang (foreverwhb@ustb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572074 and in part by the China Scholarship Council under Grant 201706465028.

ABSTRACT A deep neural network, while avoiding its complex process of feature selection, requires sufficient training samples to learn those connection weights of adjacent layers. However, in many real applications, not enough training samples are available in all cases. This paper suggests a universal-to-domain-specific learning method based on recurrent neural network for cross-domain sentiment classification and activity recognition. In the situation of having only a small amount of training samples that is in available, the structure of its network model can be adjusted flexibly according to the needs of a target domain classification or recognition. Where there are two points of our concern as follows: 1) the fine-tune and regular constraints can increase its training efficiency by updating in a small local area, namely, sharing these parameters between input and hidden layers with a target domain and 2) then, a linear output network moves on its implementing amelioration from subtlety as an exploration or exploitation in order to mitigate the phenomenon of over-fitting. Aiming at an actual situation, this domain-specific learning model with a slide window of instances and features is designed and implemented for a good long-short term memory. Finally, the two strategies are applied into IMDB reviews, Amazon product reviews, and human activities recognition collected by the built-in gyroscope sensors data, and the experimental results verify their validity.

INDEX TERMS Personalized activity recognition, recurrent neural networks, sentiment classification.

I. MOTIVATION

In order to get a higher accuracy and reliability of classification and prediction, traditional machine learning made two assumptions. Firstly, the training examples and test data used for learning should be independent. Secondly, there are enough examples to train the model. However, it is difficult for new and emerging areas to provide enough of samples in a short period of time, and the small amount does not cover all its history moments. Thus a large amount of data annotation for each area is a necessary step, which not only consumes manpower and resources but also wastes those different but related records that have been annotated.

Transfer method uses not only marked examples but also unlabeled data, so it is neither supervised nor unsupervised, nor is it equivalent to semi-supervised learning, but a new method of machine learning. It has been applied and popularized in many fields such as wireless positioning [1], text classification [2], spam filtering [3], visual recognition [4], bearing fault diagnosis [5], humanoid robot learning [6] and so on. Transfer learning is widespread in human activities. For example, if a person can play tennis, he can quickly learn how to play badminton.

On the other hand, deep neural network(DNN) has a shallow-to-deep structure, which follows the rules from these common features to those specific attributes. Freezing common features does not greatly improve the accuracy of its model [7], but the FINETUNE can promote its generalizing ability among features in the fields of voice interaction [8], image recognition [9] and intelligent recommendation [10]. However, many scenes for medical, education and daily life [11] do not perform well, so further research is needed.

It is well known that the more features share in two domains, the easier their common knowledges migrate. The under-fitting and under-conforming problems require a deep learning, whilst the process of modulating and multiplexing in a deep model require transfer learning. At present, there are two main directions for the study of deep transfer learning: one is field adaptation, and the other is multi-task learning. The former is to quickly and efficiently develop a system with better performance for a new task, field or probability distribution by maintaining and utilizing knowledge learned from one or more similar scenarios. The latter pays more emphasis on improving its model by jointly learning. This paper mainly focuses on the field adaptation.

The main contributions of this paper are listed as follows. (1) Lots of experiments based on the reviews on IMDB movie and Amazon product compare the effectiveness of SRCONLY, FINETUNE and LSTM-LON three transfer models, and analyses their related parameters (learning rate, regularization coefficient and so on) of the LSTM-LON. (2) A similarity degree judgment using a short and long memory network is helpful to make the distribution of data in its source and target areas clear, which is closely related with the performance of transfer learning. (3) Since the recognition accuracy of each wearer based on the general model is not all accurate, a personalized model is proposed. The validity of FINETUNE and LSTM-LON transfer learning methods in human activity recognition model is verified in practical application.

The remainder of this paper is organized as follows. Section II outlines the related and current work in transfer learning and recurrent neural network. Section III presents a prototype implementation of the flexible emotion classification model based on recurrent neural networks(RNN), LSTM-LON model and its simulation analysis. Section IV reports the experiments that evaluate the effectiveness and scalability in application of human activity recognition. Finally, section V concludes the paper with a discussion of future work.

II. RELATED WORK

In order to make full use of the pivot feature to match a training set as much as possible to its related test set, Blitzer *et al.* [12] introduced the structural correspondence learning algorithm (SCL) into cross-domain affective analysis. Firstly, the selection of hub feature is replaced by the calculation of mutual information [13], which is a measure of the degree of relevance between the feature and domain, so as to improve the accuracy of classification. Pan *et al.* [14] proposed the spectral feature alignment (SFA) to construct a matrix based on the co-occurrence information of the words between the source and its target area, by a transformation of the co-occurrence matrix, the common feature space in two areas constructed the feature transfer of emotion classification. Tan *et al.* [15] applied the naive Bayesian semi-supervised learning method based on the maximum expectation to the cross-domain sentiment analysis, where its feature space was constructed by an entropy of frequent co-occurrence, and the samples of the source and target areas were represented in a common. In the sample space, the adaptive Naive Bayesian classifier is constructed by the maximum expectation. Shi *et al.* [16] proposed a cross-domain active transfer learning method, which selects out-of-domain labeled samples by the size of likelihood bias. This learning method correctly predicts in-domain data, and high-likelihood-biased labeled samples are utilized directly, while those with low bias are selected through active learning. Liao *et al.* [17] gave a method to estimate the degree of mismatch between each sample in the source area and a small amount of tagged data in the target area and apply that information to logistic regression. Zhuang *et al.* [18] proposed three regularization techniques based on semi-supervised learning, manifold regularization [19], entropy regularization [20] and expectation regularization [21], they proposed a hybrid learning method based on regularization. The method first derives a classifier from the source domain training and then optimizes it in the target domain data through hybrid regularization. Dai *et al.* [22] extended a Boosting learning to transfer learning and proposed the TrAdaBoost algorithm, where the weight of the samples in its target domain that is conducive to its model training is enhanced.

Many variants of RNN focuses on processing and predicting those sequence data, such as bidirectional RNNs [23], echo state networks (ESN) [24], gated recurrent neural networks (GRNN) [25], long and short time memory networks (LSTM) [26] and so on. Nowadays, the recognition of human body movements based on the wearable sensor is mainly carried by the smart phones. The accelerometers, gyroscopes, magnetometers and other sensors built in the smart bracelet generate different three-dimensional data for various behaviors, and these data in turn provide us lot of more accurate training samples for recording our service location [27]. The difficulty in recognizing human activities based on wearable sensors is that the data generated by the sensors varies greatly, when different people perform the same activities, for example, an individual walks at different step-sizes and even different speeds [26]. These pose a challenge to the accurate activity recognition models based on wearable devices. It is still able to identify more accurately such actions as *sitting*, *standing−on*, *upstairs* and *downstairs*.

III. DOMAIN SPECIFIC LEARNING BASED ON RNN

A. EMOTION CLASSIFICATION BASED ON LSTM

As a kind of RNN, LSTM also has such a chain structure, where each unit consists of a forgotten gate, an input gate, an output gate, and a storage unit of information, which can meet the processing of sequence data. Our LSTM model for sentiment classification includes a word embedding layer, a LSTM layer and a softmax layer, as Figure 1 shows its structure in detail.

1) WORD EMBEDDING LAYER

This layer maps each word w_i appearing in the corpus into a k-dimensional vector $x_i \in R^k$. A dictionary required in this process can be generated or pre-trained at random. Such a comment containing n words $[w_1, w_2 \ldots, w_n]$ can be represented by a matrix $X \in R^{n\times k}$, where $X = [x_1, x_2 \ldots, x_n]^T$.

2) LSTM LAYER

The LSTM layer is the core of our network model, which applies its N LSTM units for extracting those feature wanted,

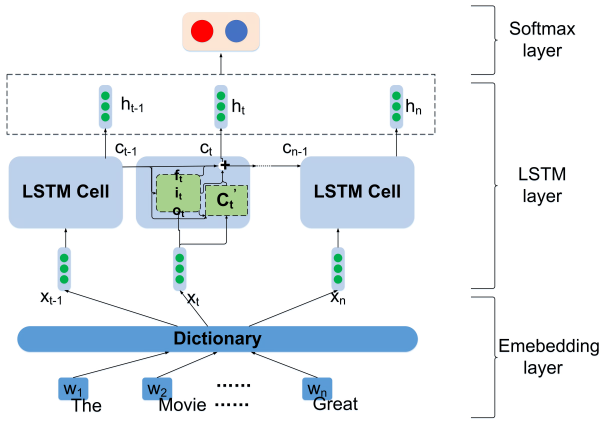

FIGURE 1. LSTM model with three layers for emotion classification.

and we can use its output of the last LSTM unit as those high-level features of the entire review. For time step t, provided X of Equation (1), the formulas of forget gate f_t, input gate i_t, output gate o_t, cell state c_t and output h_t of LSTM unit defines their functions as following in Equation (2)-(6), respectively:

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{1}$$

$$f_t = \sigma\left(W_f * X + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_i * X + b_i\right) \tag{3}$$

$$o_t = \sigma\left(W_o * X + b_o\right) \tag{4}$$

$$c_t = f_t \odot X + i_t \odot tanh\left(W_c * X + b_c\right) \tag{5}$$

$$h_t = o_t \odot tanh\left(c_t\right) \tag{6}$$

Where, a symbol $\odot$ denotes the inner product operation, and another σ represents its activation of sigmoid function.

3) SOFTMAX LAYER

Let h_N being a high-level semantic feature extracted by the LSTM layer, its following output y obtained after the Softmax layer's weight W and the offset term b is shown in Equation (7).

$$y = W * h_N + b \tag{7}$$

After the normalization of Softmax, a sentence may label as a training example with positive (noted as 0) or negative (noted as 1) emotion, as its probability follows the distribution $p(j|X)$ in Equation (8).

$$p(j|X) = softmax\left(y_j\right) = \frac{e^{y_j}}{\sum_{m=0}^{1} e^{y_m}} \tag{8}$$

Where, its criteria of $j = 0$ represents a positive emotion example, $j = 1$ for a negative one.

4) LOSS FUNCTION

As shown in Equation (9), the cross information entropy $L(y)$ may acquire their related degree of all sentences in the training set. This just meets the iterative calculation requirements of the classification model loss function.

$$L\left(y\right) = \sum_{j=0}^{1}\left(-y_j' log\left(y_j\right)\right) \tag{9}$$

Where y_j' represents the value of the j-th position in its real emotion tag of the sentence X.

In fact, the iterative computing accumulates the whole entropy $Loss$ of all training samples during conceiving of its appropriate model, as shown in Equation (10).

$$Loss = \sum_{(X,y')\in D} L\left(y\right) = \sum_{(X,y')\in D}\sum_{j=0}^{1}\left(-y_j' \cdot log\left(y_j\right)\right) \tag{10}$$

Where, $\left(X, y'\right) \in D$ represents a training sample (X, y') subjected to a training set D.

B. LSTM-LON MODEL

It can be seen from the previous section that some essential and semantic features can be learnt from a training set by the chain structure of LSTM, but these features has a domain dependency. It seems plausible to apply these features into a target area through some transformation. This idea inspired our research into a combination of LSTM and LON (abbreviation for linear output network). The LSTM-LON model adds a fully-connected layer before the softmax layer in the LSTM model. An adaptive layer with the same number of LSTM cells is added after the sigmoid layer. Assuming that the weights and biases of its adaptation layer denotes as W_{adap} and b_{adap}, respectively, and its transformed eigenvectors $\hat{z}$ may obtain as shown in Equation (11).

$$\hat{z} = \sigma(W_{adap} * h_N + b_{adap)} \tag{11}$$

Where its activation is a sigmoid function: $\sigma\left(x\right) = \frac{1}{1+e^{-x}}$.

In order to mitigate the pressure of over-fitting in its training stage, $L2$ regularization has been integrated in the new adaptation layer and its loss function of LSTM-LON model. The part of its loss function recorded as $Loss_{adap}$, as shown in Equation (12).

$$Loss_{adap} = ||W||^2 + ||b||^2 \tag{12}$$

The weighted sum of information entropy of all labeled sentences in the training set and their regular items is obtained from their loss functions in the entire LSTM-LON model. Set $Loss_{adap}$ before the coefficient of λ, $Loss$ function's expression is as follows in Equation (13).

$$\begin{aligned} Loss &= \sum_{(X,y')\in D} L\left(y\right) + \lambda \times Loss_{adap} \\ &= \sum_{(X,y')\in D}\sum_{j=0}^{1}\left(-y_j' \cdot log\left(y_j\right)\right)\lambda \times Loss_{adap} \end{aligned} \tag{13}$$

In order to improve Adam's optimization, before computing an adaptation layer, its parameters of each layer need initialize. The pre-trained LSTM emotion classification model corresponds to the parameters of each layer, and its specific implementation process is shown in Algorithm 1 below.

Algorithm 1 LSTM-LON Sentiment Classification

Input: Source Training Set D_U^S, Target Training Set D_U^T, Target Test Set D_L^T, LSTM model Φ, LSTM-LON model Ψ, iteration numbers n_1, n_2
Output: the accuracy y on D_L^T

$i, j \leftarrow 0$
while $i < n_1$ **do**
  Select a mini-batch r_{src} of data from D_U^S
  Feed r_{src} to model Φ and optimized by Eq.(10)
  Update model Φ
  $i \leftarrow i + 1$
end while
Save the updated model Φ_{tr}
Initialize model Ψ using the parameters before Softmax layer in model Φ_{tr}
while $j < n_2$ **do**
  Select a mini-batch r_{tgt} of data from D_U^T
  Feed r_{tgt} to model Ψ and optimized Eq.(13)
  Update model Ψ using using parameter update algorithm
  $j \leftarrow j + 1$
end while
Save the updated model Ψ_{tr}
for all word sequence $X \in D_L^T$ **do**
  Test X on model Ψ_{tr}
  Label X as $l_X \leftarrow argmax(p(y|X))$
end for
Compute accuracy y
return y

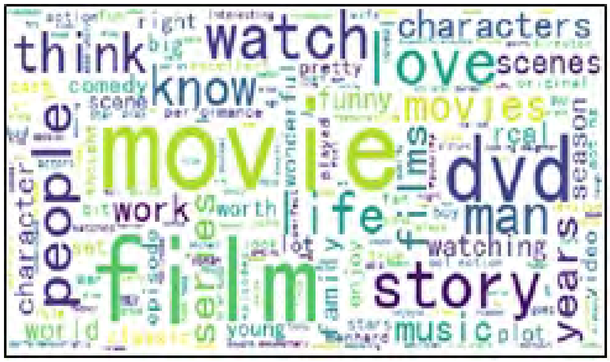

FIGURE 2. Counting word frequencies (positive examples) of DVD fields.

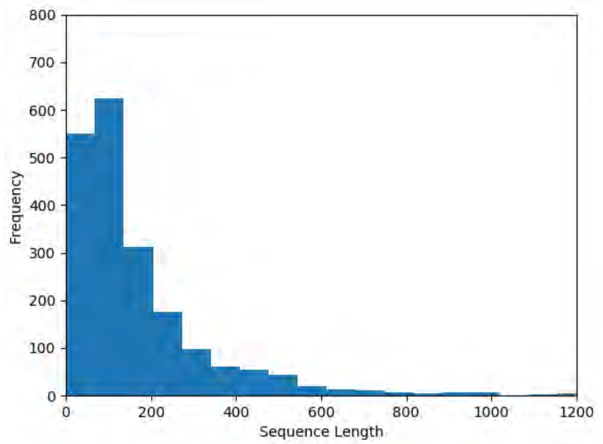

FIGURE 3. Distribution of length of comments in various fields of book.

C. SIMULATION

1) EXPERIMENTAL DATA

This section uses the IMDB movie reviews as a source area, and take the book and DVD data of Amazon Review as its target areas for LSTM learning. The number of comments in the three review sets, its word distribution and the related length distribution of its comments are shown in Table 1.

TABLE 1. Comparison of the number of each comment set.

	Book	DVD	IMDB movie reviews
Positive example	1000	1000	12500
Negative example	1000	1000	12500

Figure 2 shows the word frequency(positive examples) of DVD fields. The larger a word font, the more often it appears in the field. We can see that the top five words appearing in the field of books are *book*, *read*, *story*, *author* and *people*; then the words of *movie*, *film*, *dvd*, *story* and *people* occupy the top five in the field of DVD. Meanwhile the top 5 in IMDB movie reviews are *movie*, *film*, *comedy*, *story* and *director*, respectively. From this perspective, the reviews in IMDB movie and DVD areas are more similar.

Figure 3 shows the distribution of comment length in each area. It can be seen that there are differences in various fields. The average length of comments in the field of books is 172.8 words, while that in DVD field is 166.8 words, and it is 233 words in IMDB.

2) EXPERIMENTAL STEPS

The code in our experiments is using Python3.6.0 with Tensorflow neural network framework. The running environment is under the Ubuntu operating system, Inel Core i5-3210M @ 2.50GHz, GPU NVIDIA GeForce GT 635, 4GB RAM.

After a preliminary analysis of the experimental data set, the following procedure is in detail.

Step (1): training LSTM model for emotion classification with the IMDB dataset, we can get an neural network model with the accuracy of more than 85%.

Step (2): conducting transfer learning experiments with 800 positive and 80 negative examples in the field of books and DVDs with four strategies (SRCONLY, TGTONLY, FINETUNE and LSTM-LON).

In step (1), its hyper-parameters are set as follows: 50 neutrons of its embedding layer, 64 LSTM units, 250 slides of its sequence, 24 batches. While its learning rate is 0.001. Its curves of training accuracy and loss function of LSTM on IMDB review dataset are shown in Fig. 4 and Fig. 5, respectively.

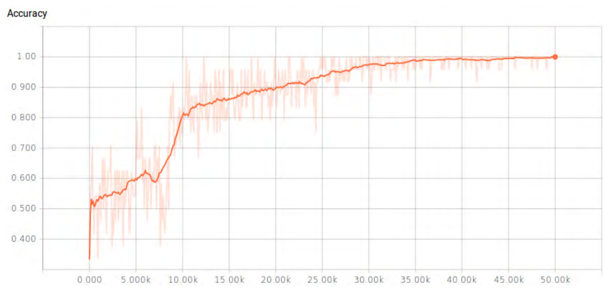

FIGURE 4. Curve of training accuracy of LSTM on IMDB review dataset.

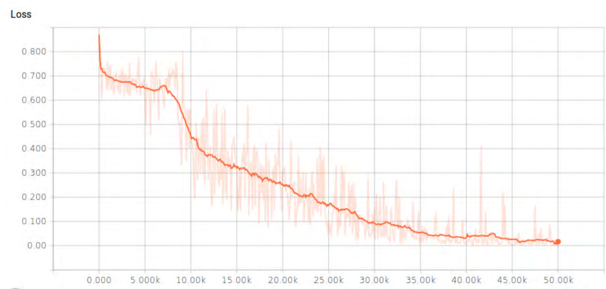

FIGURE 5. Curve of training loss function of LSTM on IMDB review dataset.

It can be seen from Fig. 4 that its training accuracy of LSTM model for emotion classification has converged since 10,000 iterations, and it reaches a good score of 90% after 20,000 iterations. Surprising, its accuracy reached more than 95% in 50,000 iterations. Fig. 5 illustrates that its loss function dropped rapidly before 7,000 iterations, which verified the proposed algorithm.

D. EXPERIMENTAL ANALYSIS

1) ACCURACY COMPARISON

It can be seen from Table 2 that the LSTM-LON or FINTUNE methods are more accurate than SRCONLY. Especially, on the task $IM \rightarrow B$, compared with SRCONLY, the accuracy of FINETUNE is reduced by 8.75%. The accuracy of LSTM-LON on $IM \rightarrow B$ and $IM \rightarrow D$ tasks are more 9.12% and 17.91% higher than that of SRCONLY and FINETUNE, respectively. For the task $IM \rightarrow D$, the LSTM-LON achieves 78.33% accuracy, while FINETUE has only 70.25%, which is about 1.5% higher than that of SRCONLY.

TABLE 2. Comparison of training accuracy of various transfer methods.

Tasks	SRCONLY	FINETUNE	LSTM-LON
$IM \rightarrow B$	0.6167	0.5292	**0.7083**
$IM \rightarrow D$	0.6875	0.7025	**0.7833**
Average	**0.6521**	**0.6185**	**0.7458**

On the other hand, among the three methods (SRCONLY, FINETUE, LSTM-LON), a task $IM \rightarrow D$ gets a higher accuracy than another task $IM \rightarrow B$. This is probably due to their data distributions exit a more similarity between the IM and IMD domains.

2) LOSS COMPARISON

Transfer losses often evaluate their performance of various algorithms. This section designs some experiments for comparing their transfer losses from different methods (SRCONLY, FINETUNE, and LSTM-LON) on tasks $IM \rightarrow B$ and $IM \rightarrow D$ as shown in Figure 6.

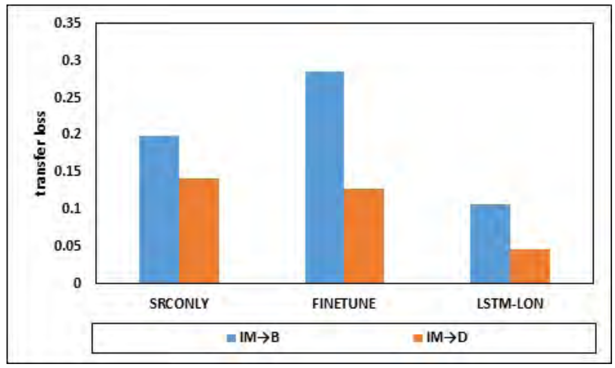

FIGURE 6. Comparison of transfer loss from different methods (SRCONLY, FINETUNE, and LSTM-LON).

As can be seen from Figure 6, compared with the SRCONLY method, the transfer loss of the FINETUNE changes significantly with different tasks. On the task $IM \rightarrow D$, FINETUNE method only decreases by 0.015 compared with the SRCONLY, especially when the task $IM \rightarrow B$, on the contrary, Transfer Loss has risen by 0.0875 than the SRCONLY. The LSTM-LON method performs at a better and stable trend, where its transfer loss is not only the lowest, but also its decline is by 0.0916 and 0.0908, respectively, comparing with the SRCONLY and FINETUNE.

E. PARAMETERS ANALYSIS

In this section, the comparison of the three configures of parameters for tasks $IM \rightarrow D$ exits in its training. (1) The learning rate is 0.00025 and λ is 0; (2) The learning rate is 0.00001 and λ is 0.0001; (3) The learning rate is 0.00001 and λ is 0.00001; In total, it iterates 12,000 times, the LSTM-LON produces a new record about its accuracy on this test set every 1000 intervals for tackling tasks $IM \rightarrow D$. Finally, its precision curve plots in Figure 7.

Figure 7 shows three representative patterns of parameters in the experiment. It can be seen that when the learning rate is set as 0.00001, taking λ out of assembly causes the model to stop learning prematurely, e.g., λ takes 0.0001. On the contrary, if its λ is 0.00001, the accuracy obtained in early iteration is not ideal, but after 10000 iterations, it will get a satisfactory performance.

When there is no regular constraint on its adaptation layer in a LSTM-LON model, that is, when λ equals 0, a more significant improvement of accuracy with fewer iterations may come up, but with the increase in its iteration, its accuracy will not be promoted.

IV. APPLICATION

As the mentioned above, although a model of human activity recognition based on wearable devices obtained after training

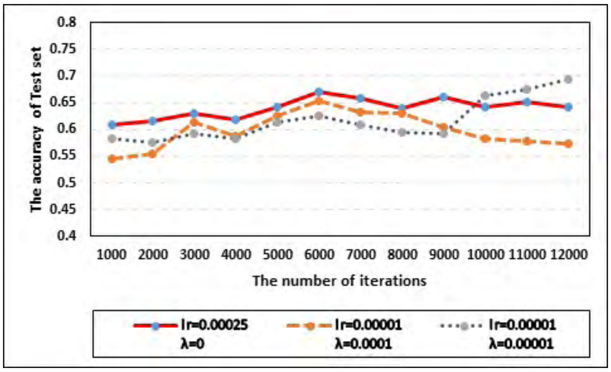

FIGURE 7. Curve of accuracy under three different patterns of parameters.

a large number of samples has a high accuracy, it is apparently inappropriate to apply the universal model directly to everyone. The generalized model of human activity will be evolved into a personalized activity recognition.

A. HUMAN ACTIVITY RECOGNITION

A personalized model includes a two-stage process. In the first stage, we need train a generalized model on a large training set of human activity data collected by sensors built into the smart phone through a RNN. In the second stage, each individual need cultivate its features on a private activity dataset through a transfer learning.

1) UNIVERSAL ACTIVITY RECOGNITION MODEL

Figure 8 shows the basic steps of the universal event recognition model, which need undergo about four steps below. (1) Collecting the active sensor data from clients' smart device. (2) Sampling these collected data through a window-slide (segmentation and overlapping) technology. The sensor data in each sampled window is recorded as a sequence labeled with the corresponding activity type. (3) Setting up a model for long-short-term memory network recognition. (4) Training the network model with annotated series.

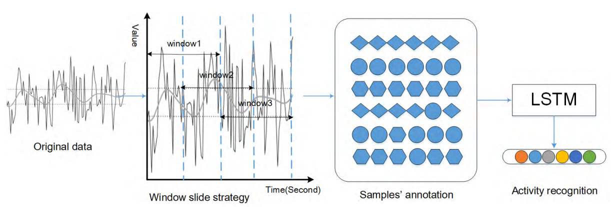

FIGURE 8. Activity recognition based on a sequence labeled with LSTM.

This paper uses two layers of LSTM network in step (4) as a classification for activity recognition. The model diagram is shown in Figure 9.

The model uses the normalized sensor samples with a sequence $X = [x_1, x_2, \cdots, x_N]$, where N is its length of the sequence. The output h_N of the last neuron after the input of LSTM in the second layer is classified into human activities

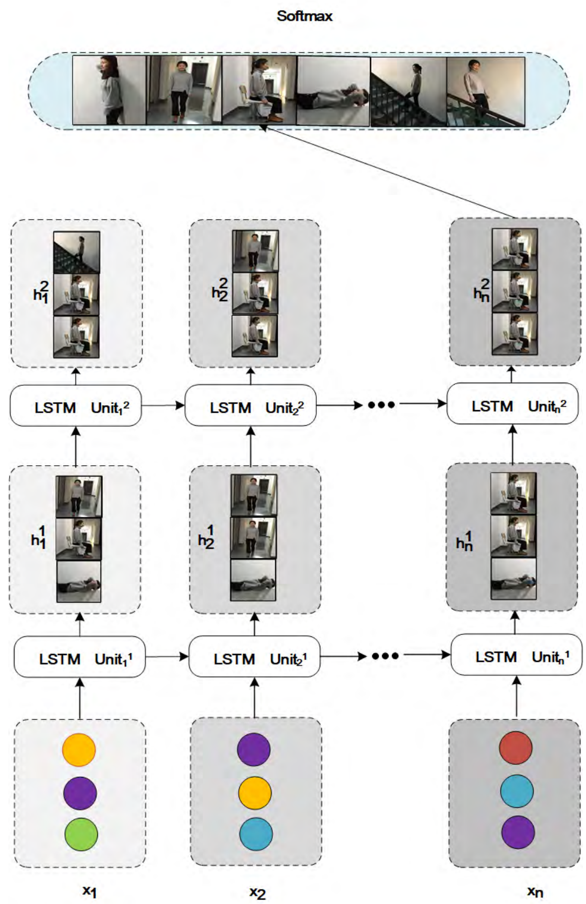

FIGURE 9. LSTM-oriented Universal Activity Recognition Model.

through its sofmax layer of six neurons characterized by h_N. Assuming that the weights of its Softmax layers are W_{out}, its output y is expressed as follows in Equation (14).

$$y = W * h_N + b \tag{14}$$

The probability that its input sequence X is labeled with a human activity label is as follows in Equation (15).

$$p(j|X) = softmax\left(y_j\right) = \frac{e^{y_j}}{\sum_{m=0}^{5} e^{y_m}} \tag{15}$$

Where $j = 0, 1, 2, 3, 4,$ or 5, the activity denotes as walking, upstairs, downstairs, sitting, standing, lying down, respectively.

General training loss function $Loss$ is divided into two parts of the output Softmax layer, at the same time cross entropy function $Loss_1$ and the model parameters $Loss_2$ denote as follows in Equation (16), (17) and (18).

$$Loss_1 = \sum_{(X,y')\in D} L\left(y\right) = \sum_{(X,y')\in D} \sum_{j=0}^{1} \left(-y'_j.log\left(y_j\right)\right) \tag{16}$$

$$Loss_2 = ||W||^2 + ||b||^2 \tag{17}$$

$$Loss = Loss_1 + \lambda Loss_2 \tag{18}$$

Where, *Loss* is optimized by a variant Adam's optimization function during its training stage, and its gradient value at each iteration is updated by the back-propagation strategy until the goal is touched.

2) PERSONALIZED LEARNING

The subsection uses LSTM as a generalized model and customizes a personalized model via the two transfer learning processes, which are based on the LSTM model pre-trained with those larger sensor data, considering with the difference from a target data, it is necessary to fine-tune the high-level network structures and its related parameter initialization. As for the specific personalized activity recognition, it has been described in detail as shown in Algorithm 2.

Algorithm 2 Personalized Activity Recognition

Input: Source Training Set D_U^S: the normalized sensor samples $X = [x_1, x_2, \cdots, x_N]$, where N is the length of the sequence. Target Training Set D_U^T, Target Test Set D_L^T, LSTM model Φ, LSTM-LON model Ψ, iteration numbers n_1, n_2, error ϵ.
Output: h_N and the accuracy y on D_L^T

$i, j \leftarrow 0$
while $i < n_1$ **do**
  Select a mini-batch r_{src} of generalized data from D_U^S
  Feed r_{src} to model Φ and optimized by Eq.(14)
  Update model Φ using related parameters strategy
  $i \leftarrow i + 1$
end while
Save the updated model Φ_{tr}
Initialize model Ψ using the parameters before Softmax layer in model Φ_{tr}
while $j < n_2$ or $h_N > \epsilon$ **do**
  Select a mini-batch r_{tgt} of personalized data from D_U^T
  Feed r_{tgt} to model Ψ and optimized Eq.(18)
  Update model Ψ
  $j \leftarrow j + 1$
end while
Save the updated model Ψ_{tr}
for all sensor data sequence $X \in D_L^T$ **do**
  Test X on model Ψ_{tr}
  Label X as $l_X \leftarrow argmax(p(y|X))$ by Eq.(15)
end for
Compute accuracy y and h_N
return y and h_N

B. SIMULATION EXPERIMENT

1) EXPERIMENTAL DATA

The samples of sensor data in the experiments comes from smart phone-based recognition of human activities and postural transitions data set. This sample data is recorded *X, Y, Z* axis data collected by the accelerometer inertia sensor built into a smart phone. During the collection, 30 volunteers aged 19-48 were asked to complete six daily human activities including three static behaviors (standing, sitting and lying down) and three dynamic movements (walking, going upstairs, downstairs), where its frequency is 50Hz, and its sliding window size is 128, a total of 10299 marked samples. The number of samples for those different behaviors is as follows in Table 3.

TABLE 3. Number of samples marked by each activity.

Behaviour	Label	#.Samples
walking	1	1722
upstairs	2	1544
downstairs	3	1406
sitting	4	1777
standing on	5	1906
lying down	6	1944
sum	10299 samples	

Different habits of a behavior can reflect on its sensor data. We plot them provided by different volunteers in the training and test set shown in Figure 10 and Figure 11, where these dots of different colors in the picture represent different volunteers.

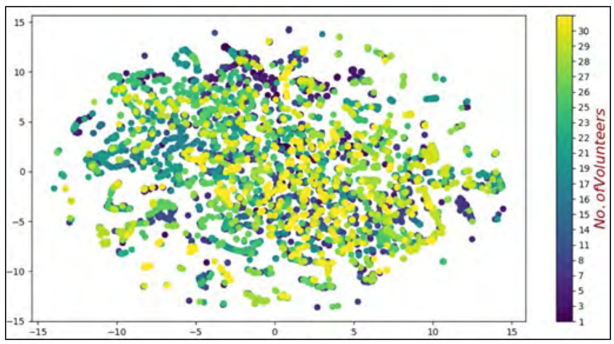

FIGURE 10. Data distribution training set (different volunteers in colors).

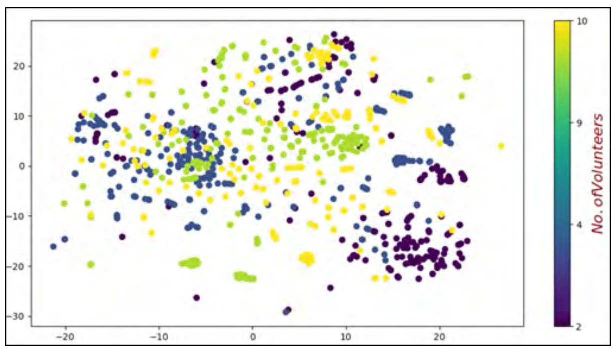

FIGURE 11. Data distribution of Test set (different volunteers in colors).

The training set contains sensor data from 21 volunteers and the test set contains sensor data from 9 volunteers. It can be seen that the distribution of 9 volunteers in the test is not uniform. For example, their sensor data provided by volunteers No.2, No.4 and No.9 is very different (No.2 volunteer often distributes in bottom right of the figure; No.4 and No.9 locate in the map above and left side). This led to the traditional model having a big difference in identifying performance among the nine volunteers.

2) EXPERIMENTAL STEPS AND PARAMETER SETTINGS

The experiment is divided into two parts: general model(GM) and personalized model(PM) training.

(1) Generalized model of training

The generalized model has trained on a dataset containing 21 sets. After its learning, the universal model got an accuracy of 91.6% on a test set containing 9 volunteers. The parameters set as follows. Its learning rate is 0.0025 in 300 tries with 1500 batch-size, and the regular coefficient is 0.001. The behavior of volunteers in each test set an initial value. Taking volunteer No.2, No.10, No.18 and No.24 as an example, the recognition accuracy of these four volunteers during the GM training is as follows.

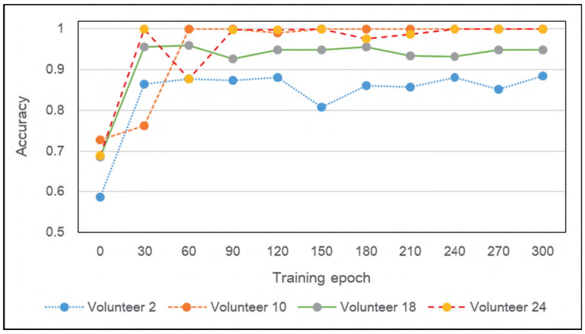

FIGURE 12. Accuracy curve of activity recognition four volunteers.

As we can see from Figure 12, although the accuracy of volunteers No.2 and No.18 is 88.41% and 94.79%, respectively. But other two volunteers have all achieved 100% recognition of their activities.

(2) Personalized model training

Sensor data collected from volunteers No.2 and No.18 is divided into a target training set and a test set according to the ratio of 2:1. The generalized model is trained by the FINETUNE method, and the LSTM-LON method for sensor data collected from volunteers No.2 and No.18, in which λ and μ were 0.0015. The final accuracy is shown below.

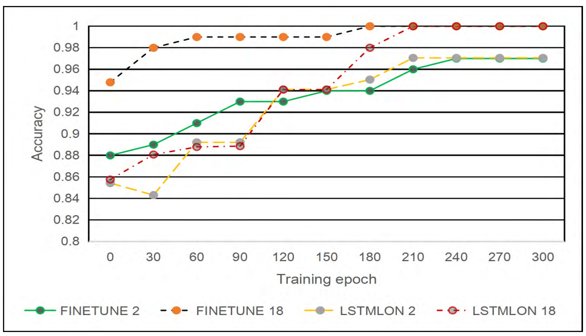

FIGURE 13. Curves of FINETUNE and LSTMLON training on two volunteers.

As can be seen from Fig. 13, the FINETUNE method achieves convergence faster, while LSTM-LON, although slow in convergence, can be seen slightly better at final accuracy than FINETUNE, For example, the final accuracy on NO.2 was 97.05% and the FINETUNE method was 97%.

V. CONCLUSION

This paper suggests an emotion classification model based on LSTM. By comparing the SRCONLY, FINETUNE and LSTM-LON strategies on test data of Amazon.com Books and DVDs reviews, we analyses the parameters of learning rate and regularization coefficient in this method. Two kinds of transfer learning models (FINETUNE and LSTM-LON) deal with the human movement sensor samples((from UCI machine learning knowledge base) to verify their validity in sentiment classification and human activity recognition. The long-term goal of this project is to combine the deep transfer learning with the big data collected from the Internet of things, so as to provide better service for our daily life.

REFERENCES

[1] Z. Sun, Y. Chen, J. Qi, and J. Liu, "Adaptive localization through transfer learning in indoor Wi-Fi environment," in *Proc. 7th Int. Conf. Mach. Learn. Appl.*, San Diego, CA, USA, Dec. 2008, pp. 331–336.

[2] B. C. Do and Y. A. Ng, "Transfer learning for text classification," in *Proc. 18th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Y. Weiss, B. Schölkopf, and J. C. Platt, Eds. Cambridge, MA, USA: MIT Press, 2005, pp. 299–306.

[3] J.-N. Meng, H.-F. Lin, and Y.-H. Yu, "Transfer learning based on SVD for spam filtering," in *Proc. Int. Conf. Intell. Comput. Cogn. Inform. (ICICCI)*, Jun. 2010, pp. 491–494.

[4] M. Xiao and Y. Guo, "Feature space independent semi-supervised domain adaptation via kernel matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 54–66, Jan. 2014.

[5] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14347–14357, 2017.

[6] A. Farchy, S. Barrett, P. MacAlpine, and P. Stone, "Humanoid robots learning to walk faster: From the real world to simulation and back," in *Proc. Int. Conf. Auto. Agents Multi-Agent Syst.*, 2013, pp. 39–46.

[7] S. K. Sahu and A. Anand. (2017). "What matters in a transferable neural network model for relation classification in the biomedical domain?" [Online]. Available: https://arxiv.org/abs/1708.03446

[8] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[9] W. Ouyang, X. Wang, C. Zhang, and X. Yang, "Factors in finetuning deep model for object detection with long-tail distribution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 864–873.

[10] L. Zhao, S. J. Pan, E. W. Xiang, E. Zhong, Z. Lu, and Q. Yang, "Active transfer learning for cross-system recommendation," in *Proc. AAAI*, 2013, pp. 1205–1211.

[11] J. C. Davila, A.-M. Cretu, and M. Zaremba, "Wearable sensor data classification for human activity recognition based on an iterative learning framework," *Sensors*, vol. 17, no. 6, p. 1287, 2017.

[12] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 120–128.

[13] D. Bollegala, D. Weir, and J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1719–1731, Aug. 2013.

[14] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 751–760.

[15] S. Tan, X. Cheng, Y. Wang, and H. Xu, "Adapting naive Bayes to domain adaptation for sentiment analysis," in *Advances in Information Retrieval—ECIR* (Lecture Notes in Computer Science), vol. 5478, M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, Eds. Berlin, Germany: Springer, 2009, pp. 337–349.

[16] X. Shi, W. Fan, and J. Ren, "Actively transfer domain knowledge," in *Machine Learning and Knowledge Discovery in Databases—ECML PKDD* (Lecture Notes in Computer Science), vol. 5212, W. Daelemans, B. Goethals, and K. Morik, Eds. Berlin, Germany: Springer, 2008, pp. 342–357.

[17] X. Liao, Y. Xue, and L. Carin, "Logistic regression with an auxiliary data source," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 505–512.
[18] F. Zhuang, P. Luo, Q. He, and Z. Shi, "Inductive transfer learning for unlabeled target-domain via hybrid regularization," *Chin. Sci. Bull.*, vol. 54, no. 14, pp. 2470–2478, 2009.
[19] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[20] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 529–536.
[21] G. S. Mann and A. McCallum, "Simple, robust, scalable semi-supervised learning via expectation regularization," in *Proc. ACM 24th Int. Conf. Mach. Learn.*, 2007, pp. 593–600.
[22] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 193–200.
[23] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schlüter, and H. Ney. (2016). "A comprehensive study of deep bidirectional LSTM RNNs for acoustic modeling in speech recognition." [Online]. Available: https://arxiv.org/abs/1606.06871
[24] J. Herbert, "Echo state network," *Scholarpedia*, vol. 2, no. 9, p. 2330, 2007.
[25] R. Dey and F. M. Salem. (2017). "gate-variants of gated recurrent unit (GRU) neural networks." [Online]. Available: https://arxiv.org/abs/1701.05923
[26] P. Zhou *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2016, pp. 207–212.
[27] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. R. Reyes-Ortiz, "Energy efficient smartphone-based activity recognition using fixed-point arithmetic," *J. Universal Comput. Sci.*, vol. 19, no. 9, pp. 1295–1314, 2013.

HONG-BO WANG (M'07) received the Ph.D. degree in computer application technology from the University of Science and Technology Beijing, China, in 2007. He is currently an Associate Professor and a M.S. Candidate supervisor with the Department of Computer Science and Technology, School of Computer and Communication Engineering, University of Science and Technology Beijing. He is focusing on transfer learning, microservices, and cognition scheduling. He received the 2015 National College Challenge Cup Smart City Special Competition First Prize, the Smart iCase Gold Medal Team Coaches, the Best Paper Reward at the 15th IEEE International Conference on Cognitive Informatics and Cognitive Computing (Stanford University), one of the best team winner in hAQUAthon 2017 in U.K., and the Gold Reward Team Coach at the Google Android Mobile Internet Innovation Challenge Finals in China in 2017.

YANZE XUE is currently pursuing the M.S. degree with the School of Computer and Communication Engineering, University of Science and Technology Beijing. His research area is machine learning and natural language processing.

XIAOXIAO ZHEN is currently pursuing the M.S. degree with the Department of Computer Science and Technology, School of Computer and Communication Engineering, University of Science and Technology Beijing. Her research areas are artificial intelligence and machine learning.

XUYAN TU is currently a Professor and a Ph.D. Supervisor with the University of Science and Technology Beijing. His research interests include artificial intelligence and large system cybernetic. He was the Chairman of the China Association for Artificial Intelligence.