WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM

Zhenghua Chen, Le Zhang, Chaoyang Jiang, Zhiguang Cao, and Wei Cui

Abstract—Human activity recognition can benefit various applications including healthcare services and context awareness. Since human actions will influence WiFi signals, which can be captured by the channel state information (CSI) of WiFi, WiFi CSI based human activity recognition has gained more and more attention. Due to the complex relationship between human activities and WiFi CSI measurements, the accuracies of current recognition systems are far from satisfactory. In this paper, we propose a new deep learning based approach, i.e., attention based bi-directional long short-term memory (ABLSTM), for passive human activity recognition using WiFi CSI signals. The BLSTM is employed to learn representative features in two directions from raw sequential CSI measurements. Since the learned features may have different contributions for final activity recognition, we leverage on an attention mechanism to assign different weights for all the learned features. Real experiments have been carried out to evaluate the performance of the proposed ABLSTM for human activity recognition. The experimental results show that our proposed ABLSTM is able to achieve the best recognition performance for all activities when compared with some benchmark approaches.

Index Terms—Human activity recognition, WiFi, CSI, ABLSTM

1 Introduction

Human activity recognition is of great importance for many applications, such as healthcare services and context awareness. To maintain healthy conditions for elders, the long-term monitoring of their daily activities is compulsory [1]. Other possible healthcare applications include the detection of falls [2] and the recognition of some specific diseases, such as Parkinson's [3]. Moreover, the recognition of human activities in buildings can be exploited for building control systems to provide a comfortable indoor environment with high energy efficiency [4].

To recognize various human activities, a number of sensors have been employed in prior works. Camera-based human activity recognition systems can be found in [5], [6]. The merit of camera based systems is the capability of detecting some tiny movements of the human body. However, these systems often suffer from some issues, such as the influence of the illumination condition and privacy concerns. Wearable sensors are also popular for human activity recognition due to the high recognition accuracy [7]. But the systems based on wearable sensors require users to take extra devices for activity recognition, which is inconvenient and obstructive for users. Another widely employed sensor for human activity recognition is the modern smartphone. Since many sensors, such as accelerator, gyroscope and barometer, are embedded in smartphones, they can be treated as a power sensing platform for human activity recognition [8]. However, if subjects forget to take their smartphones, the activity recognition will terminate. Meanwhile, the running of sensors in smartphones will influence their battery usage.

Under the principle that human actions between WiFi transmitters and receivers will influence WiFi signal characteristics, WiFi-based passive human activity recognition can be feasible [9]. Due to the wide availability of WiFi signals in indoor environments, human activity recognition using WiFi is a cheap solution without any additional cost. Moreover, the passive activity recognition systems based on WiFi do not require users to take any devices for recognition. Therefore, in this paper, we focus on human activity recognition using WiFi signals. The most commonly used signal for WiFi is the received signal strength (RSS) which has been widely used for indoor localization [10]. It can also be used for human activity recognition [11], but with limited performance due to the noisy and unstable RSS measurements.

Instead of RSS, a more informative characteristic of WiFi named channel state information (CSI) has attracted more and more attention due to the abundant and stable information in CSI [12]. Due to the high noise ratio, the raw CSI measure may not be representative enough for different human activities. A common practice is to manually extract discriminative features [13]. However, those hand-crafted features require expert knowledge and the generalization ability is not guaranteed because the feature extraction and recognition part are not jointly optimized. Recently, a deep learning approach, i.e., long short-term memory (LSTM),

- *Z. Chen and L. Zhang are with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Sinagpore 138632. E-mail: chen0832@e.ntu.edu.sg, zhang.le@adsc.com.sg.*
- *C. Jiang is with the School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China. E-mail: chaoyangjiang@hotmail.com.*
- *Z. Cao is with the Department of Industrial Systems Engineering and Management, National University of Singapore, 1 Engineering Drive 2, Singapore 117576. E-mail: zhiguangcao@outlook.com.*
- *W. Cui is with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China. E-mail: cuiwei@sdust.edu.cn.*

Manuscript received 4 Feb. 2018; revised 17 Sept. 2018; accepted 22 Oct. 2018. Date of publication 30 Oct. 2018; date of current version 30 Sept. 2019. (Corresponding author: Le Zhang.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2018.2878233

which can automatically learn representative features and encode the temporal information during feature learning, has achieved a new state-of-the-art for human activity recognition using CSI measurements [14].

The conventional LSTM can only process the sequential CSI measurements in one direction, i.e., the forward direction, which means that only past CSI information has been considered for the current hidden state. We argue that the future CSI information is also of great importance for human activity recognition. Besides, the learned sequential features by the conventional LSTM may have different contributions for the task of human activity recognition. In the conventional LSTM approach, however, the learned features will have equal contributions for the final identification of human activities. In this paper, we propose an attention based bidirectional long short-term memory (ABLSTM) approach for human activity recognition using WiFi CSI measurements. The BLSTM network which consists of forward and backward LSTM layers can process the sequential CSI measurements in both forward and backward directions, leading to more abundant and informative features. And the attention mechanism can assign larger weights for more important features and time steps, leading to a better generalization performance for human activity recognition. Real experiments have been conducted to verify the effectiveness of the proposed ABLSTM for human activity recognition using WiFi CSI measurements. The results are compared with some benchmark approaches in the literature.

The main contributions of this paper are summarized as follows:

- We propose a new deep learning framework for automatic feature learning and selection in the task of human activity recognition with WiFi CSI measurements. The proposed framework can perform human activity recognition from scratch, instead of manual feature extraction which requires expert knowledge and inevitably loses implicit features.
- We leverage on an advance BLSTM network which is able to process sequential CSI measurements in both forward and backward directions for automatic feature learning and sequential information encoding. The two directional operation can take both past and future information into consideration when determining the current hidden state of LSTM, resulting more abundant and informative features for feature learning.
- We present an attention mechanism to learn the importance of features and time steps for the learned features by the BLSTM network. More important features and time steps will be assigned higher weights for final human activity recognition, leading to better recognition performance.
- We apply real experiments to demonstrate the superior performance of the proposed approach for human activity recognition using WiFi CSI. We also verify the usefulness of the phase information of CSI which is not widely used due to the large interference for human activity recognition.

The remaining of the paper is organized as follows: Section 2 reviews some advanced works for human activity recognition using WiFi signals. Section 3 introduces the attention model and the BLSTM network, followed by the proposed ABLSTM approach. Section 4 describes the data for experiments and presents the experimental setup. Then, the experimental results are presented and discussed in this section. Finally, Section 5 concludes this work and shows some potential future works.

2 Related Works

Since WiFi signals are widely available, many WiFi based human activity recognition systems have been developed in the literature. Abdelnasser et al. proposed a gesture recognition system termed WiGest with WiFi RSS measurements [11]. The WiGest consists of three parts, i.e., primitives extraction, gesture identification and action mapping. Gu et al. presented a WiFi RSS based human activity recognition system [15]. They manually extracted some representative features from the raw RSS readings. Then, a fusion algorithm was proposed to identify the simple activities of empty, sitting, standing and walking.

Due to the multi-path and the fading effect, the RSS measurements are highly unstable and noisy, RSS based activity recognition systems have limited performance, even for some simple activities. The more stable and informative CSI in WiFi has gained more and more attention recently. Zhang et al. theoretically analyzed the sensing capability of WiFi signals and presented a Fresnel zone model for human activity recognition using WiFi CSI signals [16]. The proposed model achieved very high accuracy in the detection of centimeter-scale and decimeter-scale human activities, i.e., respiration and walking direction, respectively. Wang et al. presented a location-oriented activity recognition system with CSI readings in WiFi [17]. First, a moving variance thresholding approach was utilized to separate walking activities and in-place activities. Then, they proposed two profile matching classifiers for the recognition of different walking activities and in-place activities, respectively. Wang et al. proposed a fall detection and activity recognition system using WiFi CSI measurements [2]. They developed an anomaly detection algorithm based on the theoretical analysis of the radio propagation model. Then, a singular value decomposition (SVD) approach was applied to capture the key features of the CSI matrix obtained from anomaly detection. Finally, two classification algorithms, i.e., support vector machine (SVM) and random forest (RF), were employed to identify fall and other activities. In [13], Wang et al. proposed a CSI based human activity recognition and monitoring system which consists of two key models, i.e., a CSI-speed model and a CSI-activity model. The CSI-speed model is able to obtain movement features for different activities using CSI measurements. And, the CSI-activity model which was built upon HMM is able to identify a specific activity using the extracted activity features.

Hand-crafted features in previous works require export knowledge and may inevitably lose some implicit features, some other researchers intend to apply deep learning approaches to automatically learn significant features for human activity recognition using WiFi CSI. Wang et al. proposed a deep learning approach, i.e., sparse autoencoder (SAE), for localization and activity recognition using WiFi CSI signals [18]. The SAE network was applied to learn

discriminative features from CSI signals. Then, the learned features were fed into a softmax regression algorithm for final localization and activity recognition. Gao et al. proposed a CSI based localization and activity recognition system based on radio image features and deep learning [19]. First, they transferred the CSI measurements from different channels into radio images where some image features were extracted. Then, they applied a deep learning approach of SAE to learn deep features from the extracted image features. Finally, a machine learning approach of softmax regression was employed for localization and activity classification. Another prominent work was presented in [14]. The authors first performed a comprehensive review on various human activity recognition systems based on WiFi CSI. Then, they presented a deep learning approach, i.e., long short-term memory (LSTM), which can take sequential information in WiFi CSI measurements into consideration for automatic feature learning. The experimental results showed that the proposed LSTM approach outperforms the conventional machine learning approaches with hand-crafted features.

To identify some activities using WiFi CSI measurements, one may need to carefully design some specific features with domain knowledge. These features may perform poorly when applied to identify other activities. Moreover, the hand-crafted features will inevitably lose some implicit features which may be crucial for human activity recognition. Deep learning is a good tool to automatically learn discriminate features for human activity recognition. Since the CSI measurements are sequential with temporal information for different activities, the LSTM which can encode temporal information is a good candidate for automatic feature learning. Here, we consider an improved version of LSTM, i.e., BLSTM, which consists of a forward and a backward process for feature learning. Therefore, the BLSTM is able to take both past and future information into consideration when determining the current hidden state of LSTM, leading to more abundant and informative features. Meanwhile, the learned sequential features at one time instance may have different contributions for final human activity recognition. Besides, CSI measurements at different time instance may also have different importance. Therefore, in this work, we propose an attention based bi-directional LSTM approach for CSI based human activity recognition, which will assign higher weights for more important features and time steps for final recognition.

3 Attention Based Bi-Directional Long Short-Term Memory

In this section, we first introduce the BLSTM network. Then, the attention model is illustrated. Finally, we present the proposed ABLSTM approach for human activity recognition using WiFi CSI measurements.

3.1 Bi-Directional Long Short-Term Memory

Owing to the sequential modeling capability, recurrent neural network (RNN) has been successfully applied to many challenging applications, such as language understanding [20] and video processing [21]. However, conventional RNN often suffers from the problem of vanishing and exploding of the gradient when the learning sequence is long [22]. To solve this problem, Hochreiter and Schmidhuber developed a new structure for RNN, termed LSTM [23]. The LSTM network attempts to solve the problem of vanishing and exploding of the gradient by using memory cells with some gates which can preserve useful information with long-term dependencies. Since WiFi CSI signals are typical time series with temporal dependency, the LSTM has achieved a remarkable performance for WiFI CSI based human activity recognition in [14].

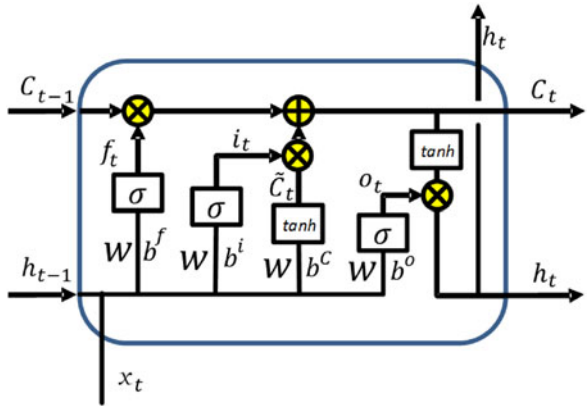


Fig. 1. LSTM network structure.

As illustrated in Fig. 1, LSTM updates itself at time t based on its input $\mathbf{x}_t$, $\mathbf{h}_{t-1}$, and $\mathbf{C}_{t-1}$ by way of:

$$\begin{aligned}
\mathbf{f}_t &= \sigma\big(\mathbf{W}^f[\mathbf{h}_{t-1}, \mathbf{x}_t] + b^f\big) \\
\mathbf{i}_t &= \sigma\big(\mathbf{W}^i[\mathbf{h}_{t-1}, \mathbf{x}_t] + b^i\big) \\
\tilde{\mathbf{C}}_t &= tanh\big(\mathbf{W}^C[\mathbf{h}_{t-1}, \mathbf{x}_t] + b^C\big) \\
\mathbf{C}_t &= \mathbf{f}_t * \mathbf{C}_{t-1} + i_t * \tilde{\mathbf{C}}_t \\
\mathbf{o}_t &= \sigma(\mathbf{W}^o[h_{t-1}, \mathbf{x}_t] + b^o) \\
\mathbf{h}_t &= \mathbf{o}_t * tanh(\mathbf{C}_t),
\end{aligned} \tag{1}$$

where $\{\mathbf{W}^f, \mathbf{W}^i, \mathbf{W}^C, \mathbf{W}^o, b^f, b^i, b^C, b^o\}$ are weights and biases. The functions of $\sigma(\cdot)$ and $tanh(\cdot)$ are *sigmoid* and *hyperbolic tangent* activation functions, respectively. $\{\mathbf{h}_t, \mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t, \tilde{\mathbf{C}}_t, \mathbf{C}_t\}$ are hidden state, input gate, forget gate, output gate, input modulation gate and memory gate, respectively. The memory cell unit $\mathbf{C}_t$ is consists of two components, i.e., previous memory cell unit $\mathbf{C}_{t-1}$ modulated by $\mathbf{f}_t$ and $\tilde{\mathbf{C}}_t$ which is modeled by the current input and previous hidden state, modulated by the input gate $\mathbf{i}_t$. The sigmoidal nature of $\mathbf{i}_t$ and $\mathbf{f}_t$ squashes themselves into a range of $[0, 1]$. They can be regarded as knobs that LSTM learns to selectively forget its previous memory or consider its current input. In the same way, the output gate $\mathbf{o}_t$ models the transfer from memory cells to hidden states. Based on these mechanisms, the LSTM learns complex as well as temporal dynamics that exist in sequential WiFi CSI measurements, resulting a remarkable performance for human activity recognition.

The conventional LSTM network can only process the WiFi CSI measurements in one direction, which means that the current hidden state only considers the past CSI information. However, the future CSI information is also meaningful for human activity recognition. To achieve this ability that is able to take both the past and future information into consideration, an advanced bi-directional long short-term

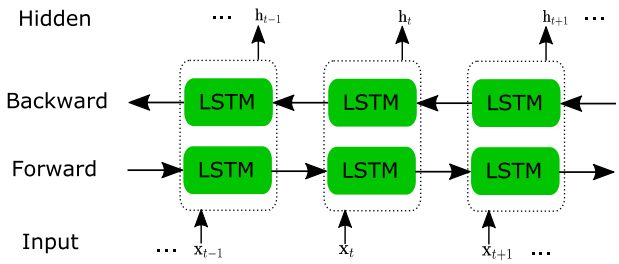


Fig. 2. Bi-directional LSTM network structure.

memory (BLSTM) can be utilized. The BLSTM consists of two layers, i.e., a forward layer and a backward layer, which are shown in Fig. 2.

According to Equation (1), the hidden states of the forward and backward layers can be represented as $\overrightarrow{\mathbf{h}}_t$ and $\overleftarrow{\mathbf{h}}_t$ respectively, where $\rightarrow$ and $\leftarrow$ denote the forward and backward processes respectively. Then, the complete hidden state for the BLSTM at time step t, i.e., $\mathbf{h}^t$, is a concatenation of the hidden states of the forward and backward layers, shown as follows:

$$\mathbf{h}_t = \overrightarrow{\mathbf{h}}_t \oplus \overleftarrow{\mathbf{h}}_t \tag{2}$$

3.2 Attention Model

The attention model was first designed for image recognition [24]. The idea is inspired by human vision systems which claim that human always pays attention to a certain region of an image during recognition, and adjust the focus over time. With the attention model, the machine is able to focus on the region of interest and obscure the rest simultaneously for a recognition task, which has been shown to be effective in image recognition [24]. Recently, the attention model has also been shown to be efficient in natural language processing [25]. For example, when using the popular encoder-decoder model without attention for machine translation, the input sentence will be encoded into a fix hidden vector for translation in the entire translation process, which means that the words in the input sentence have equal contributions for the translation at any time step. This process is ineffective and with poor performance. When applying the attention mechanism to the encoder-decoder model, the translation at different time steps will pay more attention to the words that are more related to the current translation content. For the task of WiFi CSI based human activity recognition, since the learned sequential features by the BLSTM network are of high dimensions, and different features and time steps may have different contributions for final activity recognition, we attempt to leverage on the attention model to automatically learn the importance of features and time steps, and assign larger weights to more significant features and time steps to boost the performance of WiFi CSI based human activity recognition.

For WiFi CSI based human activity recognition, no prior information can be used. Therefore, the learned sequential features by the BLSTM will be employed as the inputs of the attention model, which is also known as self-attention. Here, we demonstrate a simple case for this attention model. Given n feature vectors $\mathbf{h}_i,\ i = 1, 2, \ldots, n$ which can be derived from a feature learning network, we design a score function $\Phi(\cdot)$ which evaluates the importance of each feature vector by calculating a score s_i as follows:

$$s_i = \Phi(\mathbf{W}^\top \mathbf{h}_i + b), \tag{3}$$

where $\mathbf{W}^\top$ and b are weight vector and bias respectively. The score function can be designed as any activation function in neural networks, such as tanh, relu and linear. After obtaining the score for each feature vector, we can normalize it using the softmax function, which can be expressed as:

$$a_i = softmax(s_i) = \frac{\exp(s_i)}{\sum_i \exp(s_i)}. \tag{4}$$

The final output feature $\mathbf{O}$ of the attention model is the multiplication of the feature vectors and their normalized scores, which is shown as follows:

$$\mathbf{O} = \sum_{i=1}^{n} a_i * \mathbf{h}_i. \tag{5}$$

In this work, this attention model will be used to learn the importance of features and time steps, and assign larger weights to more important ones to boost the performance of WiFi CSI based human activity recognition. Other types of attention models can be found in [26], [27].

3.3 ABLSTM for CSI Based Human Activity Recognition

3.3.1 Channel State Information

The CSI describes the channel properties of a wireless communication link [2]. For WiFi signal propagation, it can be modeled as a MIMO (multiple inputs multiple outputs) with the orthogonal frequency division multiplexing (OFDM) technology. In frequency domain, let $\mathbf{x}_i \in \mathbb{R}^{N_{Tx}}$ and $\mathbf{y}_i \in \mathbb{R}^{N_{Rx}}$ be transmitted and received signals for subcarrier i where N_{Tx} and N_{Rx} are the number of transmitting and receiving antennas respectively, the communication system can be modeled as $\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \boldsymbol{\nu}$ for $i = 1, 2, \ldots, m$ where $\mathbf{H}_i$ is the channel state for subcarrier i, $\boldsymbol{\nu}$ is the noise term and m is the number of subcarriers. One CSI measurement will contain m CSI matrices where each has a dimension of $N_{Tx} \times N_{Rx}$. The CSI gives a fine-grained description of the communication link when compared with the widely used RSS which averages out the changes over all the channels. Thus, in this work, we adopt the fine-grained CSI measurements to detect human activities between a transmitter and a receiver. The CSI measurements consist of amplitude and phase information. The phase information is often deteriorated by some sources such as carrier frequency offset (CFO) and sampling frequency offset (SFO) [14]. The amplitude of CSI is relatively stable and has been widely used for human activity recognition [2], [14]. In this work, we also apply the amplitude information of CSI for human activity recognition and leaves the phase information for future exploration.

3.3.2 Rationale

WiFi CSI signals have been shown to be much more effective for human activity recognition when compared with WiFi RSS signals [2]. However, the relationship between WiFi CSI measurements and human activities is nontrivial. To

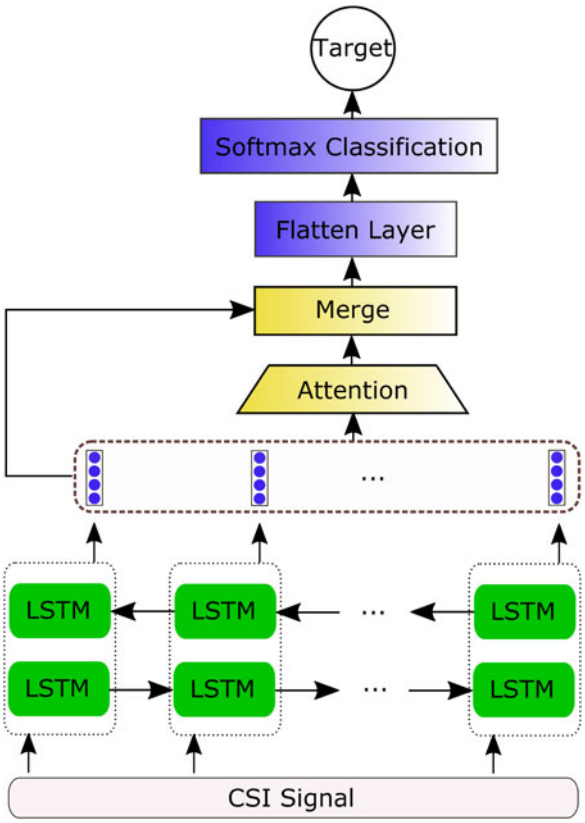


Fig. 3. The proposed ABLSTM framework for CSI based human activity recognition.

achieve better recognition performance, some researchers attempt to manually extract representative features for human activity recognition. However, manual feature extraction requires export knowledge. It is also labor intensive and time-consuming. Besides, when the activities to be recognized change, the designed features may be useless. Moreover, hand-crafted features will inevitably miss some implicit key features. Recently developed deep learning is a good tool for automatic feature learning [28]. A SAE has been employed to automatically learn useful features for human activity recognition in [18], [19]. Human activities is a sequential process which indicates that the CSI measurements will contain temporal dependencies. However, the SAE cannot take sequential information into consideration during feature learning, which degrades its performance for human activity recognition. The LSTM network which can encode temporal dependencies for feature learning has been presented in [14] for human activity recognition. It achieves a new state-of-the-art for human activity recognition with a cross-validation accuracy of around 90 percent.

The conventional LSTM can only process the sequential CSI measurements in one direction, i.e., the forward direction, which means that only past CSI information has been considered for the current hidden state. Future information is also crucial for the determination of an activity. For example, the activities of laying and sitting all require to lower the human body first, but the final positions for the two activities are different. When learning representative features for these similar activities, future information is of great importance. Thus, we leverage on a BLSTM network to learn effective features from raw CSI measurements. The BLSTM network contains two layers, i.e., a forward layer and a backward layer, which can take both past and future CSI information into consideration during feature learning. Specifically, the forward layer encodes the information of past time steps into the current hidden state, meaning to consider the past information of a CSI sequence. And the backward layer encodes the information of future time steps into the current hidden state, meaning to consider the future information of a CSI sequence. With the BLSTM network, both the past and future dependency information of the CSI sequence are considered to learn the completed context of the sequence for the identification of human activities.

Moreover, the learned sequential features from the conventional LSTM network may have different contributions for the task of human activity recognition. In the conventional LSTM network, however, the learned features will have equal weights (contributions) for the final identification of human activities. To solve this problem, we develop an attention mechanism which assigns a weight for each feature and time step. This mechanism can automatically learn the importance of each feature and time step. Then, the larger weights will be assigned to more significant features and time steps to boost the performance of human activity recognition using WiFi CSI.

3.3.3 Proposed ABLSTM Framework

The proposed ABLSTM framework is shown in Fig. 3. First, we apply a sliding window of the raw CSI signals which are fed into a BLSTM network for automatic feature learning in two directions. Here, the BLSTM network used for feature learning contains 200 hidden nodes. Since no prior information is available for the attention model, it can only use the learned features from the BLSTM as inputs to derive the attention matrix which indicates the importance of features and time steps. Here, the attention model is designed as a softmax regression layer whose outputs are normalized weights for each feature and time step. Then, we merge the learned features with the attention matrix by using element-wise multiplication, leading to the modified feature matrix with attention. After that, the feature matrix will be flattened to a feature vector for final classification by the use of a flattened layer. Finally, a softmax classification layer is used to identify different activities with the final feature vector.

3.3.4 Training of the Proposed ABLSTM

The training of the proposed ABLSTM framework is to determine all the model parameters based on the training data with true labels. At the beginning, all the parameters are randomly assigned. Then, the training data is fed into the ABLSTM to predict the labels. With the predicted labels and the given true labels, the category cross-entropy errors are calculated and back-propagated to update model parameters using gradient-based optimization methods. We adopt the ADAM [29] which can effectively compute adaptive learning rates for each parameter during optimization. In details, assume that θ_t is the parameter to be optimized, and g_t is the corresponding gradient, the updating of θ_{t+1} using ADAM is given as

$$\begin{aligned}\alpha_t &= r_1\alpha_{t-1} + (1-r_1)g_t \\ \beta_t &= r_2\beta_{t-1} + (1-r_2)g_t^2 \\ \alpha_t &= \alpha_t/(1-r_1) \\ \beta_t &= \beta_t/(1-r_2) \\ \theta_{t+1} &= \theta_t + \frac{\eta}{\sqrt{\beta_t}+\epsilon}\alpha_t.\end{aligned} \tag{6}$$

where α_t and β_t are the first and second moments of the gradient respectively, η is the learning rate which is set to be 1×10^{-4}, and the parameters r_1, r_2 and ϵ are set to be 0.9, 0.999 and 1×10^{-8} respectively. Over-fitting is a common problem in learning based systems. Here we adopt the ADAM optimizer which can compute adaptive learning rates for different parameters to reduce the risk of over-fitting. Besides, the proposed attention mechanism will only select some important features and time steps, which will also lower the probability of over-fitting.

3.3.5 Differences with Some Advanced Works

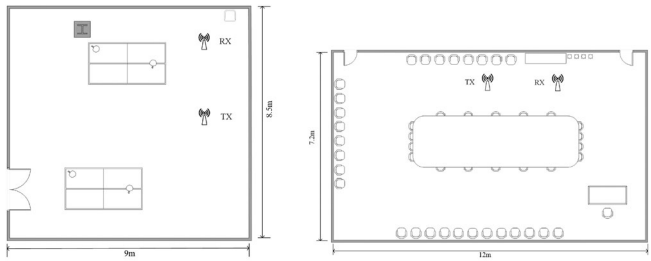
Our proposed approach is inspired by [14] which presented a LSTM based human activity recognition using WiFi CSI, but differentiates from it significantly. The main differences are: (1) we leverage on a bi-directional structure of LSTM to consider both past and future CSI information for human activity recognition; (2) an attention model is developed to assign larger weights to more important features and time steps to boost the performance of human activity recognition. Another advanced work [16] provided a theoretical analysis of RF signal propagation and developed a Fresnel zone model for the recognition of respiration and walking direction. The main differences are: (1) we focus on the recognition of human daily activities in this work, while [16] handles the specific recognition of respiration and walking direction. (2) the Fresnel zone model requires strong domain knowledge, while our proposed ABLSTM is a data-driven solution without the strong requirement of domain knowledge.

4 Experiment

In this section, we first describe the data for experiments. Then, we present the experimental setup. After that, the experimental results are shown and discussed. Finally, we evaluate the impact of one key hyperparameter and the issues of unseen activity, as well as the time complexity of the proposed ABLSTM for human activity recognition.

4.1 Data Description

The first dataset for evaluation was collected from an indoor office area by the authors in [14]. A commercial WiFi router is used as a transmitter and a laptop with Intel 5300 NIC is employed as a receiver with a sampling frequency of 1 kHz. With three antennas and 30 sub-carriers, the raw CSI data has a dimension of 90. A sliding window with a window size of 2s is used for data segmentation. The transmitter and the receiver are placed three meters apart with line-of-sight (LOS) condition. During data collection, each person performs each activity for a period of 20 seconds. Note that, at the beginning and the end of an activity, the person remains stationary. The entire data collection process is recorded by a camera to label all the data. Totally, six persons are involved for data collection with six common daily activities of "Lie down, Fall, Walk, Run, Sit down, Stand up". Each person performs each activity 20 times, yielding a dataset with the size of around 17 GB. The dataset can be found in https://github.com/ermongroup/Wifi_Activity_Recognition.



(a) Activity room (b) Meeting room

Fig. 4. The layouts of the two environments for experiments.

We also collected our own datasets with varying conditions for comprehensive evaluations. Two different environments, i.e., an activity room and a meeting room, are considered. The layouts of the two environments are shown in Fig. 4. The activity room which contains two table tennis tables has a size of $8.5\ \mathrm{m} \times 9\ \mathrm{m}$. During data collection, some subjects regularly move into or out of this room. The meeting room has a size of $7.2\ \mathrm{m} \times 12\ \mathrm{m}$. Only one subject is present during data collection in the meeting room scenario. Note that, in our experiments, we also use a commercial WiFi router as a transmitter and a laptop with Intel 5300 NIC as a receiver with a lower sampling rate of 500 Hz which is adequate to capture human activities. A sliding window with a window size of 4s is used for data segmentation. To make the category of activities more diverse, we investigate some different activities of "Empty, Jump, Pick up, Run, Sit down, Wave hand, Walk" in the experiments. Seven volunteers are involved, who are asked to perform each activity freely without any restrictions 100 times in each testing environment.

4.2 Experimental Setup

To verify the effectiveness of the proposed approach, we perform a comparison with some benchmark approaches for CSI based human activity recognition. In [14], the RF achieved a superior performance than SVM, logistic regression (LR) and decision tree (DT) for CSI based human activity recognition. The hidden Markov model (HMM) was also shown to be effective for human activity recognition in [13], [14]. Therefore, we compare our proposed approach with these two methodologies with hand-crafted features. The detailed extraction of hand-crafted features can be found in [14]. Meanwhile, we also conduct a comparison with other deep learning based methods, i.e., SAE [18], [19] and conventional LSTM [14], which are able to learn features automatically. The parameters of all the approaches are carefully tuned using a validation set from the training data. A 10 fold cross-validation is performed for evaluation. Specifically, we randomly divide all the data into 10 folds. Then, we select one fold of data for testing and the remaining for training, leading to 10 runs. The final recognition accuracy is an average of all the 10 runs.

4.3 Experimental Results on the First Dataset

The confusion matrices of all the benchmark approaches and the proposed ABLSTM approach on the first dataset are

TABLE 1
Confusion Matrices for all the Benchmark Approaches and the Proposed ABLSTM Approach

(a) RF [14]

		Predicted					
		Lie down	Fall	Walk	Run	Sit down	Stand up
Actual	Lie down	0.53	0.03	0.0	0.0	0.23	0.21
	Fall	0.15	0.60	0.03	0.07	0.1	0.05
	Walk	0.04	0.05	0.81	0.07	0.01	0.01
	Run	0.01	0.03	0.05	0.88	0.01	0.01
	Sit down	0.15	0.03	0.02	0.04	0.49	0.26
	Stand up	0.10	0.03	0.02	0.06	0.20	0.57

(b) HMM [14]

		Predicted					
		Lie down	Fall	Walk	Run	Sit down	Stand up
Actual	Lie down	0.52	0.08	0.08	0.16	0.12	0.04
	Fall	0.08	0.72	0.0	0.0	0.2	0.0
	Walk	0.0	0.04	0.92	0.04	0.0	0.0
	Run	0.0	0.0	0.04	0.96	0.0	0.0
	Sit down	0.0	0.04	0.0	0.0	0.76	0.20
	Stand up	0.16	0.04	0.0	0.0	0.28	0.52

(c) SAE [18], [19]

		Predicted					
		Lie down	Fall	Walk	Run	Sit down	Stand up
Actual	Lie down	0.84	0.01	0.03	0.03	0.04	0.05
	Fall	0.01	0.84	0.07	0.04	0.01	0.03
	Walk	0.01	0.0	0.95	0.02	0.01	0.01
	Run	0.05	0.03	0.07	0.83	0.00	0.02
	Sit down	0.05	0.01	0.03	0.03	0.84	0.04
	Stand up	0.03	0.0	0.03	0.02	0.04	0.88

(d) LSTM [14]

		Predicted					
		Lie down	Fall	Walk	Run	Sit down	Stand up
Actual	Lie down	0.95	0.01	0.01	0.01	0.00	0.02
	Fall	0.01	0.94	0.05	0.00	0.00	0.00
	Walk	0.00	0.01	0.93	0.04	0.01	0.01
	Run	0.00	0.00	0.02	0.97	0.01	0.00
	Sit down	0.03	0.01	0.05	0.02	0.81	0.07
	Stand up	0.01	0.00	0.03	0.05	0.07	0.83

(e) Proposed ABLSTM

		Predicted					
		Lie down	Fall	Walk	Run	Sit down	Stand up
Actual	Lie down	0.96	0.0	0.01	0.0	0.02	0.01
	Fall	0.0	0.99	0.0	0.01	0.0	0.0
	Walk	0.0	0.0	0.98	0.02	0.0	0.0
	Run	0.0	0.0	0.02	0.98	0.0	0.0
	Sit down	0.01	0.01	0.01	0.0	0.95	0.02
	Stand up	0.01	0.0	0.0	0.0	0.01	0.98

shown in Table 1. It can be found that the shallow learning algorithms, i.e., RF and HMM, with hand-crafted features perform the worst. The HMM model slightly outperforms the RF. The deep learning based approach of SAE has a superior performance when compared with the RF and HMM with hand-crafted features. This indicates the effectiveness of automatic feature learning using the SAE approach. Since the LSTM network also considers the temporal dependencies in sequential data for feature learning, it achieves a better performance than the SAE approach. Owing to the proposed attention mechanism and the bi-directional operation, the proposed ABLSTM approach is able to achieve the best performance for the recognition of all the six activities. The recognition accuracies for all the six activities are equal to or higher than 95 percent which is adequate for many other high-level applications.

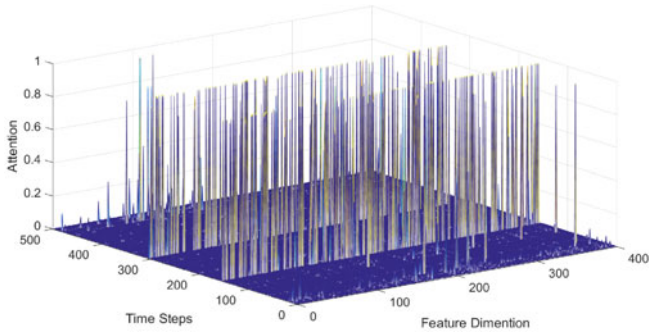


Fig. 5. An example for attention matrix.

The recognition accuracies for different activities have large difference. The activities with larger body movement, i.e., "Fall", "Walk" and "Run", have better recognition performance (see Table 1). This is because these activities will have larger influence on the characteristics of WiFi CSI signals with distinct patterns. Another observation is that the activity of "Sit down" has the lowest recognition accuracy for most of approaches including the RF, the LSTM and the proposed ABLSTM. The possible reason is that this activity has similar impact on CSI characteristics with the activities of "Lie down" and "Stand up". Note that, the recognition accuracy is even lower than 50 percent for the RF approach with hand-crafted features. For the six activities, the activity of "Fall" is of great importance, especially for elders [2], [30]. The proposed ABLSTM approach can achieve a recognition accuracy of 99 percent for the activity of "Fall", which will benefit many healthcare applications. One limitation for deep learning based approaches is the long training time. But this tedious process only needs to be done once. Note that, the online testing for the deep learning based approaches is fast enough for most of real-time applications.

To better interpret the attention mechanism, we present one attention matrix in Fig. 5. Since we set 200 hidden nodes for the BLSTM network, it will generate 400 features at each time step. Note that, a sliding window contains 500 time steps. For the BLSTM network without the attention mechanism, all these sequential features (500×400) will have equal contribution (weight) for final activity recognition. However, in most of real situations, it is not realistic. From Fig. 5, we can find that the sequential features at two time steps, i.e., 155 and 304, are dominant instead of uniformly distributed over all time steps. Meanwhile, at one time step, the 400 features have different contributions. Although these sequential features cannot be explicitly interpreted because they are high-level features learned by the BLSTM network, we can still concur that all the features have different contributions for final activity recognition, which can be achieved by using the attention mechanism. The superior performance in experiments indicates the effectiveness of the proposed ABLSTM for WiFi CSI based human activity recognition.

4.4 Additional Experiments with Different Environments

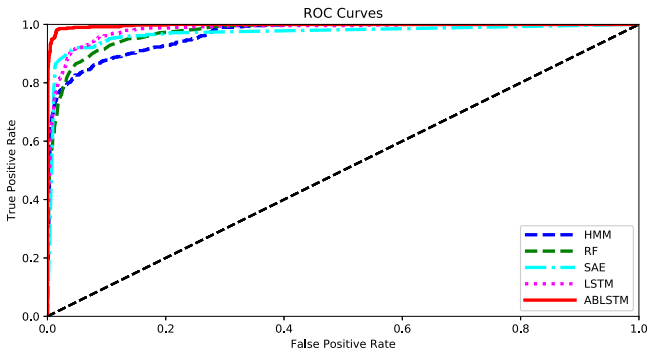
The environment for experiment is a crucial factor for WiFi CSI based applications [31], [32]. In the additional

TABLE 2
The Recognition Accuracies of All the Activities under the Two Testing Environments

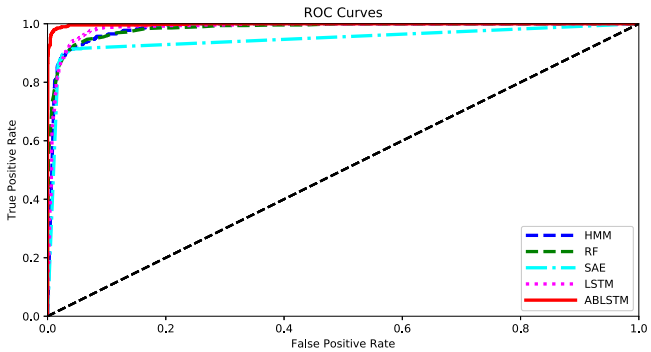
Environment	Method	Empty	Jump	Pick up	Run	Sit down	Wave hand	Walk	Overall
Activity Room	RF [14]	0.99	0.64	0.71	0.88	0.77	0.86	0.89	0.820
	HMM [14]	1.00	0.29	0.37	0.93	0.89	0.95	1.00	0.775
	SAE [18], [19]	0.87	0.75	0.88	0.87	0.86	0.92	0.86	0.859
	LSTM [14]	1.00	0.86	0.87	0.96	0.92	0.93	0.92	0.922
	Proposed ABLSTM	1.00	0.94	0.95	0.97	0.97	0.96	0.98	0.967
Meeting Room	RF [14]	0.90	0.85	0.92	0.90	0.80	0.79	0.95	0.873
	HMM [14]	0.93	0.61	0.89	0.81	0.89	0.81	1.00	0.849
	SAE [18], [19]	0.60	0.87	0.94	0.66	0.95	0.74	0.93	0.813
	LSTM [14]	1.00	0.87	0.90	0.94	0.96	0.89	0.91	0.925
	Proposed ABLSTM	1.00	0.97	0.99	0.96	0.98	0.94	0.98	0.973

experiments, we consider two distinct environments, i.e., an activity room and a meeting room. The recognition accuracies for all the seven activities under the two environments are shown in Table 2. The overall performance in the meeting room is better than that in the activity room. This is because the activity room has larger interference from additional subjects who are moving into or out of the room regularly. Among all the activities, the activity of "Empty" which means no subjects are present has the highest recognition accuracy for most of classifiers, due to the distinct patterns for this simple activity. The activities of "Run" and "Walk" which are with distinct patterns and large movements also can be easily identified.

In both environments, the LSTM approach which encodes temporal dependency for sequential WiFi CSI measurements achieves a superior performance than the approaches of RF, HMM and SAE. Owing to the efficient bidirectional structure and the proposed attention mechanism, the proposed ABLSTM significantly outperforms all these benchmark approaches. These results are consistent with the results on the public dataset. The overall accuracies of the ABLSTM in the activity room and the meeting room are 96.7 percent and 97.3 percent, respectively.



(a) Activity room



(b) Meeting room

Fig. 6. The ROC curves of all the approaches under the two environments.

We also demonstrate the receiver operating characteristic (ROC) curves of all the approaches under the two environments, which is shown in Fig. 6. The results are consistent with the conclusions from Table 2. The proposed ABLSTM achieves much better generalization performance than all the benchmark approaches, due to the merits of the bidirectional structure and the attention mechanism. The results are consistent with our previous analysis based on Table 2.

In [31], [32], the authors proposed a Doppler spectrum based method for indoor occupant counting using WiFi CSI. Cross-environment experiments were performed. Specifically, they trained their proposed algorithm with the data from one environment and tested it with the data from another two unseen environments. Due to different signal characteristics for different environments, the performance of the algorithm is limited. Inspired by their works, we attempt to test the performance of the proposed approach on the cross-environment scenario. Note that, the cross-environment scenario is even more challenging for activity recognition, because activities will have smaller and more complicated effect on WiFi CSI signals. Here, we train the proposed algorithm using the data from the activity room, and then test it using the data from the meeting room. As expected, the overall recognition accuracy greatly degrades to 0.320. This is because the CSI characteristics are quite different in these two environments (see Fig. 4) which have very distinct layouts, facilities and functionality. The cross-domain problem is very common and challenging in various machine learning and data mining applications [33]. A potential solution is to use transfer learning which can transfer the knowledge learned from one domain to another unseen domain [33]. This problem requires more efforts and is one of our future works.

For the public dataset and the self-collected datasets, some activities are the same, such as "Walk", "Run" and "Sit down", and the others are different. We intend to compare the recognition accuracies of all the activities for the proposed approach on the three datasets with different environments and data collection strategies. The results are shown in Table 3. It can be found that the performances of the three same

TABLE 3
The Recognition Accuracies of All the Activities for the Proposed Approach on the Three Datasets

Dataset	Walk	Run	Sit down	Lie down	Fall	Stand up	Empty	Jump	Pick up	Wave hand
Public	0.98	0.98	0.95	0.96	0.99	0.98	-	-	-	-
Activity room	0.98	0.97	0.97	-	-	-	1.00	0.94	0.95	0.96
Meeting room	0.98	0.96	0.98	-	-	-	1.00	0.97	0.99	0.94

activities are quite good and comparable across the three datasets. Besides, the recognition accuracies of the other activities are also quite high on different datasets. Thus, we can claim that the proposed approach performs quite well in different environments and data collection strategies. Note that, the proposed approach significantly outperforms the benchmark approaches on the three datasets with different environments and data collection strategies.

4.5 Impact of the Number of Hidden Nodes for the Proposed ABLSTM

The number of hidden nodes is an important parameter for the proposed ABLSTM. Therefore, we perform an additional experiment to investigate the impact of this parameter on the performance of human activity recognition. The experimental result is shown in Fig. 7. When the number of hidden nodes is only 50, the recognition accuracies for all the six activities are very low, especially for these activities with small body movements, i.e., "Lie down", "Sit down" and "Stand up". With the increase of the number of hidden nodes from 50 to 200, the recognition accuracies for all the activities increase beyond 95 percent. When further increasing the number of hidden nodes, we observe that the recognition accuracies of all the activities become stable. Since more hidden nodes will lead to longer training time, we choose 200 hidden nodes for the proposed ABLSTM for human activity recognition.

4.6 Impact of the Phase Information of CSI

In our experiments, we also collected the phase information of CSI for evaluations. Note that the public dataset [14] did not contain the phase information. Here, we attempt to investigate the impact of the phase information on the performance of activity recognition. The recognition accuracies of the proposed ABLSTM with and without the phased information of CSI under the two testing environments are shown in Fig. 8. It can be observed that the proposed ABLSTM with the phase information of CSI is able to improve the recognition accuracies for most of activities. Since the phase information of CSI contains large interference caused by CFO and SFO, it is difficult to manually extract some informative features. Therefore, many research works did not include this information for recognition [13], [14]. The proposed deep learning based approach can still learn informative features from the noisy phase information of CSI to further boost the performance of activity recognition. This further indicate the effectiveness of the proposed ABLSTM for human activity recognition using WiFi CSI measurements.

4.7 Unseen Activity

In some real cases, we may encounter unseen activities for human activity recognition. For all supervised learning, if one class "A" does not appear in training, the data from class "A" will be mapped to other classes during testing. Besides, if class "A" is similar to class "B" (e.g., similar patterns) which is included in training, the data from class "A" will have a high probability to be recognized to class "B". Note that, our proposed approach and the state-of-the-art approaches are all supervised learning methods.

To verify this, we perform an additional experiment where we train the proposed algorithm with the data from five activities and test the algorithm with the data from an unseen activity using the public dataset. Specifically, we train the proposed ABLSTM with the data from the five activities of "Lie down", "Fall", "Walk", "Run" and "Stand up", and test the algorithm with the data from the activity of "Sit down". The probabilities of recognizing "Sit down" as the five activities are shown in Table 4. It can be found

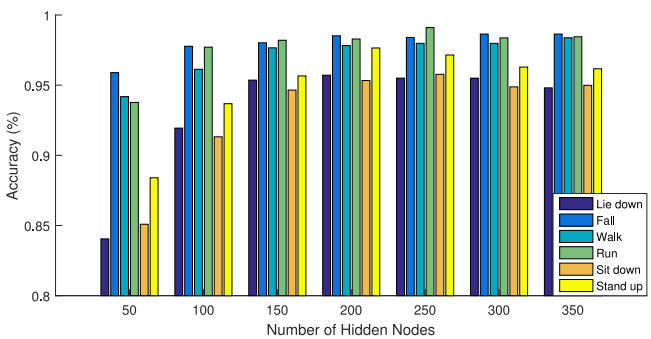
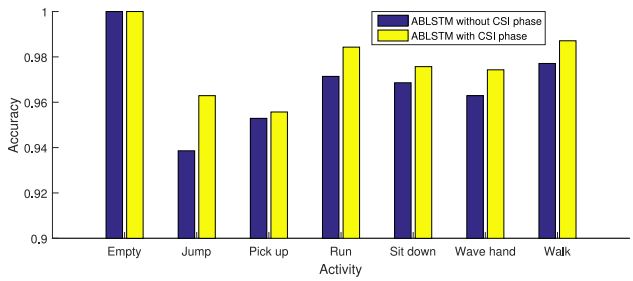
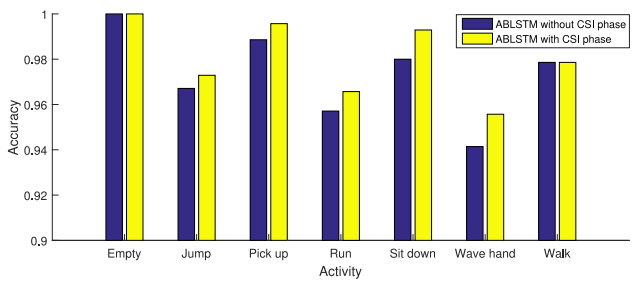


Fig. 7. The recognition accuracies for different activities with different number of hidden nodes for the proposed ABLSTM.



(a) Activity room



(b) Meeting room

Fig. 8. The proposed ABLSTM with and without the phase information of CSI under the two environments.

TABLE 4
The Probabilities of Recognizing "Sit Down" to the Other Five Activities

Activity	Lie down	Fall	Walk	Run	Stand up
Sit down	0.33	0.03	0.07	0.05	0.52

TABLE 5
The Training and Testing Time of All the Approaches

Time	RF	HMM	SAE	LSTM	ABLSTM
Training (sec)	6.09	0.029	1788.28	5168.86	13007.20
Testing (sec)	0.016	0.22	0.23	4.39	6.86

that it has higher probabilities to be classified as "Stand up" and "Lie down", which have similar movement patterns with the activity of "Sit down".

4.8 Time Complexity

It is a common concern about the time complexity of the deep learning based approaches. We have evaluated the training and testing time of all the approaches with the data from the activity room. The workstation for the experiments has twelve core CPUs of Intel i7-8700 3.20 GHz and a GPU of NVIDIA GeForce GTX1080Ti. The results are shown in Table 5. It can be found that the training time of deep learning based approaches, i.e., SAE, LSTM and ABLSTM, is much larger than that of conventional machine learning algorithms, i.e., RF and HMM. Among all the deep learning based approaches, the proposed ABLSTM has the longest training time. Although the training time of the proposed approach is large, this time-consuming training process is offline and only requires to be done once. According to Table 5, the testing time of all the approaches is quite small. For example, the testing time of the proposed ABLSTM with all the testing samples (420 testing samples) is 6.86 seconds. This means that the testing time for each sample is 0.0163 seconds. Note that each sample has a window size of 4 seconds for data segmentation. Thus, this small testing time for each sample can be neglected. We can claim that our proposed ABLSTM approach can be used for real-time WiFi CSI based human activity recognition.

5 Conclusion

In this paper, we propose an attention based bi-directional long short-term memory (ABLSTM) approach for WiFi CSI based passive human activity recognition. The BLSTM network is able to learn significant sequential features from raw WiFi CSI measurements in two directions, i.e., forward and backward. The attention mechanism will assign different weights for features and time steps based on the importance of them. We performed real experiments to verify the effectiveness of the proposed approach and compared it with some benchmark approaches, including shallow algorithms, i.e., random forest (RF) and hidden Markov model (HMM), with hand-crafted features and deep learning approaches of sparse autoencoder (SAE) and conventional LSTM. Owing to the sequential feature learning in two directions by using BLSTM and the attention mechanism to assign higher weights for more important features and time steps, the proposed ABLSTM can achieve much better performance than all the benchmark approaches. In experiments, instead of assigning the same weight for each feature and time step in conventional LSTM, we demonstrate that the attentions (weights) for different features and time steps are distinct for the proposed approach. This indicates that different features and time steps should have different importance for activity recognition. Since the number of hidden nodes is a key parameter for the proposed ABLSTM, we investigate the impact of this hyper-parameter on recognition performance. It can be concluded that when the number of hidden nodes is few, the recognition performance is limited. With the increase of the number of hidden nodes, the recognition performance improves. But when the number of hidden nodes is large enough, the recognition performance will be stable.

In our future works, we attempt to explore the two challenging issues, i.e., environment change and non-line-of-sight, for human activity recognition using WiFi CSI signals. Besides, in this work, we only focus on the recognition of single-user activity. The more realistic scenario of multi-user human activity recognition [34] will be considered in our future works. For human activity recognition, the sensory data is easy to acquire. However, the data annotation is sometimes difficult and expensive. Semi-supervised learning can be a good solution for this task [35], which will also be one of our future works.

Acknowledgments

The authors would like to thank the authors in [14] for sharing the data for our model evaluation. This work is supported by the A*STAR Industrial Internet of Things Research Program under the RIE2020 IAF-PP Grant A1788a0023, and Shandong Province Natural Science Foundation (ZR2018PF011).

References

[1] H. Ghasemzadeh and R. Jafari, "Physical movement monitoring using body sensor networks: A phonological approach to construct spatial decision trees," *IEEE Trans. Ind. Inform.*, vol. 7, no. 1, pp. 66–77, Feb. 2011.

[2] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[3] T. Wang, Z. Wang, D. Zhang, T. Gu, H. Ni, J. Jia, X. Zhou, and J. Lv, "Recognizing parkinsonian gait pattern by exploiting fine-grained movement function features," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 1, 2016, Art. no. 6.

[4] R. Rana, B. Kusy, J. Wall, and W. Hu, "Novel activity classification and occupancy estimation methods for intelligent HVAC (heating, ventilation and air conditioning) systems," *Energy*, vol. 93, pp. 245–255, 2015.

[5] X. Yang and Y. Tian, "Super normal vector for human activity recognition with depth cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1028–1039, May 2017.

[6] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Inform.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[7] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tutorials*, vol. 15, no. 3, pp. 1192–1209, Jul.-Aug. 2013.

[8] Z. Chen, Q. Zhu, Y. C. Soh, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Trans. Ind. Inform.*, vol. 13, no. 6, pp. 3070–3080, Dec. 2017.

[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.

[10] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of wifi, smartphone sensors and landmarks using the kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, 2015.
[11] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1472–1480.
[12] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, pp. 53–53, 2011.
[13] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.
[14] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
[15] Y. Gu, L. Quan, and F. Ren, "Wifi-assisted human activity recognition," in *Proc. IEEE Asia Pacific Conf. Wireless Mobile*, 2014, pp. 60–65.
[16] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Comput.*, vol. 50, no. 1, pp. 48–57, 2017.
[17] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.
[18] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.
[19] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10 346–10 356, Nov. 2017.
[20] K. Yao, G. Zweig, M.-Y. Hwang, Y. Shi, and D. Yu, "Recurrent neural networks for language understanding," in *Proc. INTERSPEECH*, 2013, pp. 2524–2528.
[21] S. Tripathi, Z. C. Lipton, S. Belongie, and T. Nguyen, "Context matters: Refining object detection in video with recurrent neural networks," 2016, https://arxiv.org/abs/1607.04648
[22] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[24] M. Denil, L. Bazzani, H. Larochelle, and N. de Freitas, "Learning where to attend with deep architectures for image tracking," *Neural Comput.*, vol. 24, no. 8, pp. 2151–2184, 2012.
[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv:1409.0473, 2014.
[26] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, "Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 335–344.
[27] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 1903–1911.
[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980
[30] C. Wang, W. Lu, M. R. Narayanan, D. C. W. Chang, S. R. Lord, S. J. Redmond, and N. H. Lovell, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Trans. Ind. Inform.*, vol. 12, no. 6, pp. 2302–2311, Dec. 2016.
[31] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential WiFi channel state information," in *Proc. 3rd Int. Workshop Phys. Analytics*, 2016, pp. 37–42.
[32] S. Di Domenico, G. Pecoraro, E. Cianca, and M. De Sanctis , "Trained-once device-free crowd counting and occupancy estimation using WiFi: A doppler spectrum based approach," in *Proc. IEEE 12th Int. Conf. Wireless Mobile Comput., Netw. Commun.*, 2016, pp. 1–8.
[33] S. J. Pan, Q. Yang, et al., "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[34] T. Gu, L. Wang, H. Chen, X. Tao, and J. Lu, "Recognizing multiuser activities using wireless body sensor networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1618–1631, Nov. 2011.
[35] L. Yao, F. Nie, Q. Z. Sheng, T. Gu, X. Li, and S. Wang, "Learning from less for better: semi-supervised activity recognition via shared structure discovery," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 13–24.

Zhenghua Chen received the BEng degree in mechatronics engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the PhD degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2017. Currently, he is a scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His research interests include data analytics in smart buildings, ubiquitous computing, internet of things, machine learning, and deep learning.

Le Zhang received the BE degree from the University of Electronic Science and Technology Of China (UESTC), in 2011, and the MSc and PhD degrees form Nanyang Technological University (NTU), in 2012 and 2016, respectively. Currently, he is a scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His current research interests include machine learning and computer vision.

Chaoyang Jiang received the BE degree in electrical engineering and automation from the China University of Mining and Technology in 2009, the ME degree in control science and engineering from Harbin Institute of Technology in 2011, and the PhD degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2017. He is currently an associate professor in the Beijing Institute of Technology. His research interests include statistical signal processing, machine learning, and sparse sensing.

Zhiguang Cao received the BS degree in automation from the Guangdong University of Technology, China, in 2009, and the MSc and PhD degrees from Nanyang Technological University, Singapore, in 2012 and 2016, respectively. He is currently a research assistant professor with the Department of Industrial Systems Engineering and Management, National University of Singapore. Prior to that, he held a position of research fellow and program leader with the BMW@NTU Future Mobility Research lab, Energy Research Institute @ NTU, Nanyang Technological University, Singapore. His research interests focus on AI and optimization for intelligent systems.

Wei Cui received the ME and PhD degrees in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2013 and 2017, respectively. From 2015 to 2016, she was a research associate with the School of Electrical Electronic and Engineering, Nanyang Technological University, Singapore. She is currently an assistant professor with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, China. Her current research interests include wireless sensor networks, localization and navigation, and machine learning.