HUMAN BEHAVIOUR RECOGNITION USING WIFI CHANNEL STATE INFORMATION

Daanish Ali Khan   *Saquib Razak*   *Bhiksha Raj*   *Rita Singh*

Carnegie Mellon University

ABSTRACT

Device-Free Human Behaviour Recognition is automatically recognizing physical activity from a series of observations, without directly attaching sensors to the subject. Behaviour Recognition has applications in security, health-care, and smart homes. The ubiquity of WiFi devices has generated recent interest in Channel State Information (CSI) that describes the propagation of RF signals for behaviour recognition, leveraging the relationship between body movement and variations in CSI streams. Existing work on CSI based behaviour recognition has established the efficacy of deep neural network classifiers, yielding performance that surpasses traditional techniques. In this paper, we propose a deep Recurrent Neural Network (RNN) model for CSI based Behaviour Recognition that utilizes a Convolutional Neural Network (CNN) feature extractor with stacked Long Short-Term Memory (LSTM) networks for sequence classification. We also examine CSI de-noising techniques that allow faster training and model convergence. Our model has yielded significant improvement in classification accuracy, compared to existing techniques.

Index Terms— Activity recognition, WiFi sensing, Channel State Information

1. INTRODUCTION

Human Behaviour and Activity recognition has always been an active research topic, largely due to the large number of applications in several areas. The ability to monitor human behaviour, without the need to directly attach physical sensors to their bodies, will have a large impact on surveillance, healthcare, and a plethora of other smart-home applications. Ideally, such monitoring must be *device-free*, i.e. requiring no special devices to be worn by the subjects being monitored. Recording devices such as cameras may not be acceptable solutions as sensors for monitoring, due to their inherent violation of privacy.

With the current ubiquity of WiFi enabled devices, the use of WiFi data for behaviour recognition has gained significant attention [1][2][3][4]. If a wireless access point and a connected device are located in the environment, analyzing the changes in the received signal can provide sufficient information for activity recognition. The signal transmitted by the access point, propagates across the room to the receiving device. As the signal propagates, it interacts with its environment as it reflects off various surfaces and objects, including human subjects. WiFi based behaviour recognition is based on the principle that different movements of body parts, will result in recognizable variations in the received signal.

WiFi sensing can broadly be split into Received Signal Strength (RSS) and Channel State Information (CSI) based methods. RSS is simply a measure of the power present in the received signal. As a human body passes between the transmitter and the receiver, the reflected signal will cause a change in the strength of the signal received. CSI provides a more detailed description of how the signal scatters, fades, and decays as it propagates. CSI provides amplitude and phase information of the signal at the subcarrier level for every pair of transmitting and receiving antennae. As a result, as the signal reflects off body parts, the change will be easy to recognize using CSI. Reflections caused by other objects will also significantly affect CSI data, making it sensitive to changes in the environment [5].

CSI captures data for 30 subcarriers for 9 pairs (three transmitting and three receiving antennae), resulting in 540 values, sampled (typically) 1000 times every second. The data are complex, and also generally degraded by environmental noise; nevertheless they have been shown to hold great promise for behavior recognition[3]. The problem remains highly challenging, though – the current best results for CSI-based behaviour recognition, which are obtained using deep-learning based systems[4] leave much room for improvement.

In this paper, we propose a novel recurrent neural network based classification system for CSI-based behaviour recognition that *greatly* enhances the state of the art. There are several significant challenges in the basic task. The data are noisy. End pointing the beginning and ends of behaviours in sensor-data streams is a serious challenge. The temporal patterns of behaviour can be quite complex. Our proposed solution addresses all of these problems. We introduce a *learned subspace-projection* de-noising algorithm to eliminate the most corrupting components of noise. We resolve the problem of end pointing through a *shift-invariant* feature extractor that naturally accounts for uncertainty in timing. The complexity of temporal patterns is addressed through an appropriately structured recurrent network model.

We compare our system to existing solutions and evaluate it on a bench-marking data-set, reducing the error by over 80% compared to the best reported result at the time of this work.

2. RELATED WORK

WiFi Received Signal Strength (RSS) data has been used for a variety of activity and behaviour recognition related tasks. As a human body passes between the transmitter and the receiver, the WiFi signal strength varies. Techniques for WiFi RSS behaviour recognition rely on this variation in signal strength for classification. RSS data has been successfully used for tasks similar to activity recognition like gesture recognition [1]. RSS based behaviour and activity recognition utilizes Software Defined Radio, and modified wireless hardware. RSS data has been shown to provide enough information for localization of activities within the environment, but does not capture fine-grained changes in the signal due to reflections off human bodies. Channel State Information captures these fine changes, and as a result, it has been shown to perform better than RSS techniques.

Techniques that use WiFi Channel State Information (CSI), utilize multiple classification techniques for behaviour recognition.

A histogram based classifier has been used on pre-processed CSI data [3], using signal modulation to classify activity. This system does not extract any high-level features, and as a result the histogram comparison method is likely to be sensitive to any minor changes in the environment. In addition, due to burst and impulse noise, the use of a low-pass filter on CSI data may not remove static multipath noise caused by the environment.

Another behaviour recognition system that uses CSI data is CARM [6]. In this system, noise caused by static-multipath is removed by performing Principal Component Analysis (PCA) [7] on the CSI data, discarding the first component and using the next five remaining vectors. It is thought that the dynamic multi-path information, caused by reflections off human body parts, will be captured in the remaining principal components. This technique has been shown to provide high classification accuracy using a Hidden Markov Model classifier, and the de-noising methods used make it far less sensitive to changes in the environment [6].

The most recent system for behaviour recognition, uses a deep learning model for classification [4]. This work utilizes a Long Short-Term Memory (LSTM) network, that is trained on the unprocessed CSI data for classification. This system was compared to the methods proposed by the CARM model [6], and the use of LSTMs was shown to outperform the HMM based approach significantly. The lack of signal pre-processing and de-noising make this approach sensitive to changes in the environment.

In this paper, we use a PCA model similar to the de-noising techniques used by CARM. We also make use of neural network based feature extractors to classify behaviour. We evaluate our system, and compare its performance to existing techniques

3. DESCRIBING CSI

The amplitude and phase information captured in CSI data are highly sensitive to any changes in the environment. Minute changes in the physical layout of the surroundings, like slight movement of furniture, results in large changes in the CSI data. These variations occur across the sub-carrier channels, and are generally hard to predict. Due to this sensitivity, CSI data is both rich in activity-related information, as well as extremely noisy. Even during periods of inactivity, where the environment is stationary, there is significant variation in the amplitudes of the CSI data. This is due to the static multi-path effect that occurs when the signal is attenuated and reflected off surfaces in the environment.
The CSI data was sampled at a rate of 1KHz, with 540 values for amplitude and phase state for each sample. Thus, for an activity that takes around 10 seconds to complete, there are 5,400,000 values in our signal, with a large amount of signal variance caused by noise. Figure 1 illustrates the appearance of a typical CSI data-stream

4. CLASSIFYING THE CSI DATA

Similar to the traditional models, our work utilizes signal de-noising techniques to improve model accuracy and generalizability. Our denoising method uses data collected while no human activity was taking place, to explicitly capture and remove the effect of environmental noise from the data. Our approach utilizes feature extraction, similar to traditional methods. There are two main categories of features in CSI data that contain significant information about behaviour: shift invariant, and temporal features. In order to extract shift invariant features, we use convolutional neural networks, a deep learning architecture that has been used successfully in signal processing, but not yet for CSI data. To extract temporal features, we use a multi-layer LSTM, a deeper model than the existing deep learning solution for behaviour recognition using CSI.

4.1. Normalization

Channel State Information is highly sensitive to changes in the environment. As the signal propagates from the transmitter to the receiver, it reflects off objects in the room (for example, walls, furniture, etc.). This static multi-path effect introduces a large amount of noise into CSI measurements. For a model to be robust under changes in environment, it will need to eliminate the noise caused by the static multi-path effect before classifying the CSI data.

In order to pre-process our data for de-noising, we perform a simple z-normalization of our data[8]. Here each component of the data in individually normalized across the recording to have zero mean and unit variance.

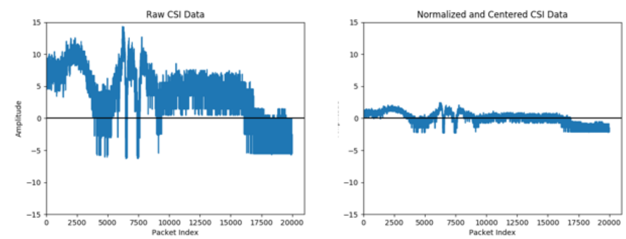

Fig. 1. Raw and z-normalized CSI data

4.2. CSI De-Noising

While the raw CSI data was shown to perform sufficiently well when used with LSTM networks for behaviour recognition, it still contains significant noise caused by the static multi-path effect caused by immobile objects in the environment (walls, furniture, etc.). In previously used systems, techniques like low-pass filters were used to filter out static multi-path noise. Due to the nature of the noise, subsequent work deemed the performance of low pass filters insufficient, and chose to use Principal Component Analysis.

We propose a learned sub-space projection approach to denoise CSI data. We make the following hypotheses, which we empirically find to be valid for CSI data.

- The overwhelming fraction of the energy of the noise lies within a low-dimensional subspace. The influence of noise on the signal is primarily within this subspace.
- This noise-dominant subspace is fixed for any environment, even if the noise itself is non-stationary.
- In the remaining space, the CSI signal retains most or all of its information.

These hypotheses lead to the following algorithm for noise reduction: we identify the *principal subspace* for the noise. Subsequently, we project all signals *out* of this subspace.

The algorithm proceeds as follows. Let $\mathbf{N} \in \mathcal{R}^{D\times T}$ represent the sequence of T D-dimensional CSI measurement vectors obtained from segments of recording where no activity is present. This represents regions of the datastream which only comprise noise. We will obtain this from labelled training data for the space. We perform the following singular value decomposition: $\mathbf{N} = \mathbf{USV}^\top$.

We compose the *denoising matrix* $\mathbf{D} = \mathbf{U}_{K+1:D}\mathbf{U}_{K+1:D}^\top$, where $\mathbf{U}_{K+1:D}$ represents the singular vectors corresponding to the

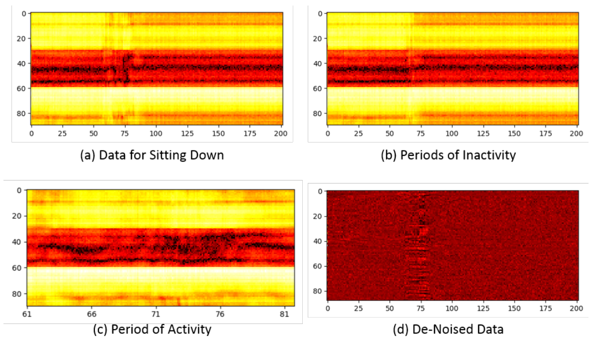

Fig. 2. Heat Map of CSI de-noising process for three antenna pairs

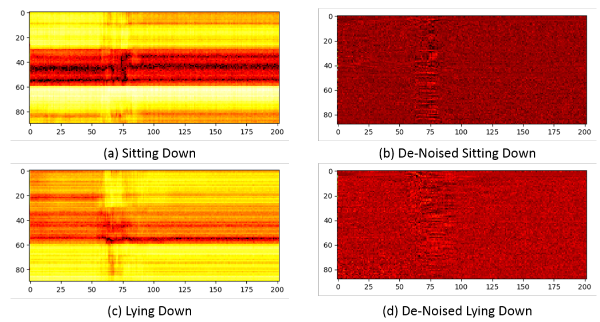

Fig. 3. Heat Map of CSI for Sitting Down and Lying Down

$D - K$ *lowest* singular vectors of $\mathbf{N}$ (assuming that the energy in the noise lies in the K-dimensional principal subspace of the noise). Empirically, we find $K = 1$ to optimal.

Subsequently we transform every CSI vector X to the denoised projection $\hat{X} = \mathbf{D}X$. All further operations are only performed on the denoised vectors $\hat{X}$. Note that we operate on the D-dimensional projections, rather than on $D - K$ dimensional principal components, since the former is more amenable to the feature extraction described in the next section.

Note that this algorithm is somewhat diametrically opposed to the more traditional sub-space projection based noise cancellation algorithms, where noise-reduction is attempted by retaining the most energetic subspaces of the *signal*.

Figure 2 contains visualizations of CSI amplitude data collected for the activity *Sit Down*. The horizontal axis represents the scaled packet index or time. This particular recording lasted a total of 20 seconds, thus every unit in the horizontal axis represents 100ms, or 100 packets. In Figures 2.a-c, the vertical axes represent sub-carrier groups. The recording contains a period of activity preceded and followed by periods of inactivity.

In Figure 2.a, we visualize the CSI data collected while a person sat down. The actual activity only occurred between time-step 61 and 82. It can clearly be seen that there is a difference in data collected in this section, with a distinct pattern most easily observed in the middle antenna pair.

Figure 2.b shows the CSI data only collected in the periods of inactivity. It can be seen that there is not much variation at the individual sub-carrier channel level. This data was produced by removing the period of activity from the recording, and concatenating the two periods of inactivity together. Figure 2.d shows the resulting data after performing our de-noising on it; the period of activity is very clearly discernible from the periods of inactivity. There is relatively little variation in the periods of inactivity, indicating that we have successfully removed a large component of the noise without losing information about the activity itself. In Figure 2.c, we visualize the data corresponding to the actual activity. Here we can see that there are distinct patterns in the CSI data.

4.3. Feature Extraction

Human behaviour tends to consist of similar sub-actions repeated in different patterns and orders. We aimed to describe all behaviours as sequences of these sub-actions. Instead of explicitly stating the sub-actions, our goal is to discover the best sub-actions from the CSI data. Each sub-action would result in a distinct signature in the CSI data, regardless of when it occurs. Figure 3 shows a visualization of data recorded during the activity of sitting down, in comparison to the activity of lying down. It can be seen that the patterns caused during the periods of activity are distinct. These patterns are caused by the movement of human body parts. Our hypothesis is that the movement of each body part in a specific direction at a certain speed, will cause the same pattern for the same sub-carrier channel. These patterns would be shift-invariant, and would be the markers for the sub-actions we want to learn.

We used a Convolutional Neural Network (CNN) to discover the signature of the sub-actions as its filters can extract shift invariant features from the signal. We propose the use of Convolutional Neural Networks as initial feature extractors, before the RNN classifier. The CNN feature extractor should, in principle, allow the model to learn to extract localized signatures from the signal. The CNN is trained in conjunction with the RNN classifier described in Section 4.4.

In the experiments reported in this paper, our CNN comprised 3 1-dimensional convolution layers, with 32, 64 and 128 filters of size 10. Convolutions are one-dimensional. Convolutions use a hop size of 2. This results in a single 128-dimensional feature at each output time. The resulting sequence of feature vectors is passed on to the RNN for classification. Figure 3 illustrates the features extracted by this model.

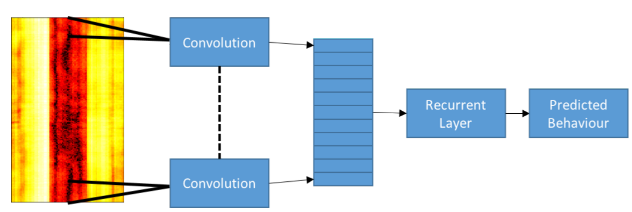

Fig. 4. Convolutional Neural Network Feature Extractor

4.4. Recurrent Neural Networks classifier

The final classification is performed using a simple multi-layer LSTM network. In our work we use 3 LSTM layers, to allow the network to maintain state information across multiple time scales. The LSTM is jointly trained with the CNN feature extractor using backpropagation through time.

5. RESULTS AND EVALUATION

We evaluate our model for behaviour recognition using CSI by using a publicly available dataset, containing 6 labelled activities, with 120 instances of each. The dataset was built using 6 subjects, performing each activity within a period of 20 seconds. The CSI data was collected using a commercial Intel IWC-5300 WiFi receiver, placed 3 meters away from the transmitter, within line of sight. The sampling rate of the data was 1kHz.

Using the LSTM RNN architecture described in [4] as our baseline system, we first try to implement this system. The LSTM network consists of 200 units and tanh activation. We train the system on raw CSI values. After 10-fold cross validation, the baseline classifier attained a baseline accuracy of 75% [4]

5.1. Evaluation of Preprocessing Techniques

Firstly, we evaluate our de-noised data effectiveness in model training. The model trained with normalized and de-noised CSI data lead to improvement in time and accuracy as compared to the model trained with raw CSI.

Secondly, we evaluate the introduction of 1-dimensional (1D) CNN feature extractor into the baseline model. Before passing the denoised CSI to the LSTM layer, features are extracted from the CNN layer. This model achieved an accuracy of 86% on the benchmark data-set. The significant improvement in classification accuracy strongly indicates that the shift-intolerant features we observed do encode significant information about the behaviour.

5.2. Evaluation of Multi-Layer RNNs

In order to evaluate our proposed LSTM classifier, we train a 3-layer LSTM on raw CSI data, resulting in an accuracy of 84%. This is a significant improvement from the baseline model. The improvement in performance validates our hypothesis that the representation of hidden states at different temporal time-scales is a better representation of human behaviour.

5.3. Evaluation of Combined Model

Building on the promising improvement in performance from preprocessing techniques, CNN feature extraction and stacked RNN architectures, we train a model combining all of these methods. This model outperforms the baseline classifier significantly, and after 10-fold cross validation, yields an accuracy of 95%, as shown in Table 1.

Model	Accuracy
Baseline LSTM	75%
De-noised LSTM	86%
CNN-LSTM	84%
Our System	**95%**

Table 1. Table of Results for CSI Classification Models: Baseline (2017), De-noised, CNN, Stacked-LSTM, and Combined Model

The confusion matrix of the combined model trained on 70% of the data-set, and evaluated on the remaining 30%, is presented in Figure 5. The model is able to classify *Lie Down* and *Fall* perfectly, but does not perform well when classifying *Walk*, *Sit Down* or *Stand Up*. The combined model is our best performing model with an accuracy of 96.7%.

6. CONCLUSIONS

Evidence from our experimental evaluation suggests that the combination of de-noising, Convolutional Neural Networks, and deep stacked Recurrent Neural Networks performs much better than traditional approaches.

De-noising techniques, such as those used in the CARM model, utilize Principal Component Analysis to eliminate noise in the CSI data. Our approach performed PCA on data collected during periods of inactivity, specifically to combat noise due to the static multi-path effect caused by the signal reflecting off objects in the environment. The use of this de-noising technique has been shown to result in an increase in convergence rate of our model, and results in a slight improvement in classification accuracy when compared to a model trained on raw Channel State Information.

The use of Convolutional Neural Networks as signal feature extractors has also been shown to perform better than using the RNN layers as the sole feature extraction mechanism. After analyzing heat-map visualizations of CSI data, we hypothesized that there existed a set of shift invariant features responsible for encoding the movement of different body parts. Our experimental analysis showed that there was a significant improvement in classification accuracy when using CNN feature extractors. This improvement came at the cost of increased training time, as the CNN model had a slower convergence than the baseline classifier.

We also see that the use of stacked multi-layer RNNs outperforms current single layer RNN solutions. We believe that this is because stacking RNN layers allows the model to learn hidden-state representation at multiple time-scales, providing a better representation of human behaviour than a single time-scale model. The combined model that utilizes de-noising, convolutional feature extraction, and stacked RNNs, outperformed all of the individual components of the system, suggesting that their individual contributions are somewhat additive.

From the confusion matrix of the combined model illustrated in Figure 5, we see that the model often mistakes sitting down for lying down, and vice versa. The activities are similar in terms of the motion of body parts, albeit in reverse order. For future research, the use of bi-directional LSTMs might perform better as they might be capable of discerning between similar activities like sitting and standing. We have seen that models are capable of performing well when evaluated in the same room they are trained in, but they do not tend to generalize well to new environments. Further improving de-noising techniques for CSI data may not only improve the classification accuracy of a model, but also the ability for a model to be generalized to a new environment.

	Lie Down	Fall	Walk	Run	Sit Down	Stand Up
Lie Down	**1**	0	0	0	0	0
Fall	0	**1**	0	0	0	0
Walk	0	0	**0.94**	0.04	0.01	0.01
Run	0	0	0.03	**0.97**	0	0
Sit down	0.02	0.01	0	0	**0.94**	0.03
Stand Up	0.01	0	0	0	0.04	**0.95**

Confusion Matrix

Fig. 5. Confusion Matrix for Combined Classifier

7. REFERENCES

[1] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 1472–1480.

[2] Stephan Sigg, Shuyu Shi, Felix Buesching, Yusheng Ji, and Lars Wolf, "Leveraging rf-channel fluctuation for activity recognition: Active and passive systems, continuous and rssi-based signal features," in *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2013, p. 43.

[3] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu, "E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 617–628.

[4] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[5] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[6] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 65–76.

[7] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[8] Erwin Kreyszig, *Advanced engineering mathematics*, Wiley-India, 2013.