HANDLING CONCEPT DRIFT IN WIFI-BASED INDOOR LOCALIZATION USING REPRESENTATION LEARNING

Anonymous authors

Paper under double-blind review

Abstract

We outline the problem of concept drifts for time series data. In this work we analyze the temporal inconsistency of streaming wireless signals in the context of device-free passive indoor localization. We show that data obtained from WiFi channel state information (CSI) can be used to train a robust system capable of performing room level localization. One of the most challenging issues for such system is the movement of input data distribution to an unexplored space over time, which leads to an unwanted shift in the learned boundaries of the output space. In this work we propose a phase and magnitude augmented feature space along with a standardization technique that is little affected by drifts. We show that this robust representation of the data yields better learning accuracy and requires less number of retraining.

1 INTRODUCTION

Concept drift is one of the most common problems that degrades the predictive performance of passive WiFi-based localization systems. In most of the predictive models it is assumed that a static relationship between input and output exits. Thus in the context of machine learning, there is a mapping function f(x) = y, where the algorithm tries to estimate the underlying relationship between the input x and the output y. The presence of concept drift means that the accuracy of the predictive models that is trained from historical data degrades over time due to evolving nature of the data. Hence, predictive models often needs to be retrained frequently with a new set of labelled data, which might be expensive to obtain. These pattern changes can be categorized based on their transition speed from one state to another into abrupt, or gradual drifts (Gama et al., 2014). In either case, the deployed solution is expected to diagnose unintended changes automatically and adapt accordingly.

The problem of concept drift in WiFi-based localization systems, was first mentioned in (Ghourchian et al., 2017), which presents a technology that utilizes only off-the-shelf WiFi-enabled devices such as access points, laptops, smart TV for passive sensing in the environment of interest. The authors have applied an online semi-supervised approach to automatically detect gradual shifts in the feature space and propose an adaptive learning strategy to regain the prediction accuracy. We aim to address the same problem without making any assumption about the drift type. In this work, we illustrate that from time to time, both sudden and gradual drifts, can occur to the streaming WiFi data, which often hinder the performance of the trained models when tested on the measurements.

Majority of the existing WiFi-based indoor localization systems are device-based, where the user's location is determined by a WiFi-enabled target device that needs to be carried by the subject all the time (Wang et al., 2017). Practical challenges of using device-based approaches, impose some restrictions and therefore, a device-free and passive solution is a promising line of research both for academia and industry. For example, (Wang et al., 2015a;b; Sen et al., 2012), are some of the existing research where device free passive WiFi localization is used along with deep learning. In (Chen et al., 2017), the authors address drifts and the inconsistency of WiFi fingerprints for stationary subjects. However, most of these researches and their experiments were performed in a very controlled environment and within a limited time frames. On the other hand, the effect of concept drift mostly appears over time due to real-world conditions such as natural WiFi channel or bandwidth switches, or when certain exogenous factor such as temperature and humidity changes. Therefore, the existing methods do not address them explicitly and the experimental results does

not reflect the performance of the model taken from measurements that are a few days apart. In this paper, we use the idea of feature augmentation in order to include both phase and magnitude of the CSI data. To the best of our knowledge this is the first work that exploits both the phase and magnitude of the CSI in order to construct a feature space that is less affected by drifts. We show that once such a feature space has been constructed, we can use classical machine learning algorithms in order to create a more robust model. In the next sections, we discuss nature of the WiFi CSI data being obtained and how drifts cause a shift in the feature space. In Section 3 we discuss our methods including the phase and the magnitude sanitization procedure. In Section ?? we present the training strategy for off line training and online prediction. Finally in Section 5, we conclude our paper and present discussions on future work.

2 DATA DESCRIPTION

In wireless communication, channel state information (CSI) contains potential information that describes the propagation of a signal from transmitter to receiver. The CSI contains vital information that describes the combined effect of scattering, fading and decay with distance. In other words CSI reflects the variation in the channel that is experienced during propagation. Transmitted from a source, a wireless signal can experience various forms of distortion including fading, shadowing and multipath effects. For our application, we considered a WiFi channel at the 5GHz band which can be considered as a flat fading channel. Our network interface card (NIC) implements an OFDM system with 56 subcarriers, all of which can be read from CSI measurement. The receiver (Rx) and the transmitter (Tx) have 4 antennas each and in total our NIC establishes 16 links or streams altogether (one per Rx-Tx pair). The channel frequency response $CSI_{i,k}$ for subcarrier *i* and stream *k* is a complex number which is defined as:

$$CSI_{ik} = |CSI_{ik}| \exp\{j \sin \angle CSI_{ik}\}$$
(1)

where $|CSI_{ik}|$ and $\angle CSI_{ik}$ denote the magnitude and the phase response respectively. Let \mathcal{I} be the total number of subcarriers, \mathcal{M} be the total number of packets and \mathcal{S} be the total number of streams. Therefore, our system produces 3 dimensional complex gain matrix with cardinality $|\mathcal{I}| \times |\mathcal{M}| \times |\mathcal{S}|$. In the sequel, we show WiFi mesh variation for an empty capture of an indoor area and a capture containing walking inside a particular room of that area, obtained on different time stamps.

Figure 1 shows the CSI amplitudes captured over 60 seconds (with sample rate of 20packets/second), Figure 1(a) shows the empty capture i.e. a capture with no motion and Figure 1(b) shows a capture with walking activity in Room 1. Figure 2 shows a similar capture taken after 9 hours. These figures illustrate the effect of drift on the WiFi mesh, both for empty and walking captures. This change in distribution of the data along the feature space is what we refer to as concept drift. In the next section we formally discuss the methodology that we adapt in order to construct a new robust feature space that is less affected by drift.

3 Methodology

In this section, we discuss the methods that we followed starting from the sanitization of the raw CSI data. Then We proceed on discussing the ways in which we incorporated both the phase and magnitude of the CSI and justify why the combined feature space is little affected by drifts.

3.1 INITIAL PREPROCESSING

We start our data processing with Received signal strength indicator (RSSI) drops filter. This filter looks at the successive packets and measures sudden peaks for RSSI values, these peaks can be results of constructive interference from neighboring devices, multipath fading and temporal dynamics. The filter then discards the corresponding packets from the CSI. Since we have multiple subcarriers and streams which correspond to different links between the transmitter and the receiver, at each point in time, they can take values which are scaled to a wide range. Hence after discarding the packets based on RSSI corrections, we perform normalization of the CSI amplitudes to a predefined range. The L_2 norm of the CSI vector is then calculated for each of the CSI vectors in order to re-scale their values to the predefined range.



Figure 1: Variation of CSI mesh with no motion and walking in Room 1



Figure 2: Variation of CSI mesh with no motion and walking in room 1 after 9 hours

3.2 PHASE INFORMATION

In this section we outline the process by which we extract the phase information from the CSI and use it for feature augmentation. Prior research conducted with extraction of phase information from CSI reported an extensive amount of preprocessing being involved (Yang et al., 2013). In (Wang et al., 2017), the authors discuss the stability of phase for consecutive antennas for 5Ghz OFDM channel. Since, our NIC implements a 5GHz OFDM channel, we utilize the fact that phase difference between successive antennas are stable. We consider the phase difference between stream 1-2, stream 2-3 and stream 3-4 as they correspond to the links from a single transmitter to all 4 of the receiving antennas. A phase correction is then performed for the phases such that their values lie with in the range $(-\pi, \pi)$. We then use a Hampel filter in order to remove the DC component of the phase information and to detrend the phase data. For our Hampel filter we use a large sliding window of 300 samples and with a small threshold of 0.01 in order to get the general trend of the data. Once the trend has been computed, it is then removed from the the phase difference information. Then, we further leverage Hampel filter with a smaller sliding window of 15 samples and a threshold of 0.01 in order to remove the high frequency noise from the streaming phase data. In Figure 3 we can see that the raw phase information has a wider spread and hence is more unstable compared to the phase difference information between successive antennas. In Figure 3(a) we show the plot for the raw phase information that corresponds to subcarrier 1 and stream 1 and in Figure 3(b) we show the plots for the phase difference between stream 1 and stream 2 which corresponds to the links from a single transmitter and two adjacent receiving antennas. In the next section we discuss about the feature augmentation and the standardization technique that we use in order to create a feature space that is robust to drifts.

3.3 FEATURE AUGMENTATION

We propose an augmented feature space, comprising of both phase and magnitude of the WiFi CSI. For our feature space we consider CSI magnitude for 8 streams and phase difference data from the



(a) Polar plot for stream 1

(b) Polar plot of difference between stream 1 and 2

Figure 3: Comparison for stable and unstable CSI phase for an Empty Capture



Figure 4: Variation of CSI mesh in the presence of drift



(a) Augmented feature space for empty capture





first four streams i,e we take the phase difference between stream 1, stream 2; stream 2, stream3 and stream 3,stream 4. We consider the first 800 packets for our data, therefore for 8 streams and all 56 subcarriers the cardinality of the magnitude \mathcal{M} of the CSI is of the order $|56| \times |800| \times |4|$ whereas for the phase information the data matrix has a cardinality of $|56| \times |800| \times |3|$. Our combined feature space \mathcal{F} suitable for learning is a 2-D matrix comprising of both phase and magnitude with cardinality $|800| \times |392|$ for each location class (Room). Once the augmented feature space is obtained, we perform a standardization that standardizes the features by removing the mean and scaling to unit variance. Figure 4 shows the change is feature space observed when there is a drift, Fig 5 shows that the augmented feature space is almost resistant to drifts. In the next section we discuss broadly the result of different learning algorithms for this augmented feature space.We discuss the training strategy that we have adapted for the experiments and present detailed results and the effect of each algorithm on this enriched feature space.

4 TRAINING STRATEGY

We present an offline training and an online prediction strategy for our system. We use classical machine learning algorithms to train on the un-drifted dataset using the augmented features. During testing, the algorithm in tested on the drifted data, which when projected to the combined feature space is least resistant to drifts. Thus trained models are used for the online prediction of the data that has drifts. We compare the performance of different learning algorithms for our training and classification for, a) the case of training only on magnitude data, b) training only on phase data and c) training on the combined data, which represents our most stable feature space.

4.1 OFFLINE TRAINING

In this scenario, we train the data offline and perform classification using different learning algorithms. We mainly use Support Vector Machines (SVM) (Hearst et al., 1998) and Random Forest algorithms (Liaw et al., 2002). We then provide an incremental learning framework, that is popular with streaming data, specially for dealing with datasets associated with concept drifts. In the sequel we provide a detailed analysis of the performance of learning under these different frameworks and present the suitability of a learning framework that will be used for the real time localization application.

4.1.1 DATA ACQUISITION

We start our data collection procedure from apartments of three different sizes with different layouts (Apt 1, Apt 2 and Apt 3). We use two different devices for the experiments. Namely, the Tx and the Rx which corresponds to the routers for transmission and the reception, respectively. Both of the devices are placed further apart in the apartment and the experiment begins by taking an empty capture at the first instance. This empty capture corresponds to no motion at any of the rooms in the apartment. We next proceed towards capturing 1 minute data by walking in each of the rooms of the apartment, respectively, in order to obtain annotated data. The data is then collected and processed and converted to the augmented feature space as described in Section 3.1, 3.2 and 3.3, respectively. For Apt 1 we captured 5 rounds of data each of which is roughly 30 minutes apart. Although drift is more apparent for measurements taken over longer intervals, for measurements associated with Apt 1 we force a channel switched (abrupt drift) before collecting the 5th round by switching the devices off. This ensures that a drift has occurred since drifts are expected during a channel change. For Apt 2 we perform a more rigorous measurement and hence take 6 rounds of data, where we captured 3 rounds which are 12 hours apart and the last 2 rounds which are 2 days apart. For Apt 2 we take the measurements in such a diverse manner so that the effect of drift can be thoroughly studied. Finally for Apt 3, we capture 3 rounds of data where round 1 and round 2 are data which are 6 hours apart and round 3 is captured at an interval of 12 hours from round 2. Figure 6 shows the layout of the three apartments in which the experiments were conducted. Through all of the experiments we ensured that the position of Tx and Rx remains fixed. Although our proposed augmented feature space results in a dataset with high dimension, we found through repetitive experiments that using Principal Component Analysis (PCA) for dimensionality reduction yields very poor classification accuracy even when appropriate components are chosen based on explained cumulative variance analysis. Hence, we do not perform any dimensionality reduction on our dataset.

For our experiments, all the learning algorithms are trained on rounds showing no drift and tested on rounds that has both gradual and abrupt drifts.

4.1.2 TRAINING AND VALIDATION

In order to validate whether walking in different rooms of an indoor space actually correspond to distinguishable clusters from WiFi propagation perspective, we do an unsupervised clustering analysis over the dataset to evaluate our location partitioning. From the elbow analysis of the unsupervised clustering we found that WiFi mesh distortions can also be categorized in an unsupervised manner, where the number of clusters correspond to walking or physical activities in number of areas in the apartment.

Figure 7 presents the elbow analysis done in Apt 1 which consists of labelled data for 6 locations. For the elbow analysis described in Figure 7 we can see that there is not much reduction in distortion



Figure 6: Layout of the apartments where experiments are conducted



Figure 7: Elbow analysis of K-means clustering algorithm

when increasing the number of clusters from 6 to 8 thus we can conclude that the way in which we label the data in fact represents the different distribution that arises due to motion in different positions.

In order to do offline training, we use Support Vector Machines (SVM) and Random Forest (RF) classifiers as the base learners. For each of these learners, we consider a K class classification problem where K is the number of positions / rooms where walking is performed. We chose SVM since it is effective in high dimensional spaces and because of its memory efficiency since it only uses a subset of the dataset in order to calculate the support vectors. We implement the SVM for performing non-linear classification using RBF kernels.

We chose the Random Forest classifier, since it is a meta estimator that creates decision trees for different sub-samples of the training set and averaging the performance over them in order to find a better predictive accuracy. The performance for both the classifiers are compared when trained on the rounds with no drift and tested on the rounds with drift, which is set aside as a held out set. We also compare the time of training of these two base learners in order to justify the suitability of the corresponding learner for real-time indoor localization.

4.1.3 INCREMENTAL LEARNING

In this section we present an incremental learning algorithm based on the proposed feature space. We chose an incremental learning framework since this allows the input data to continuously extend the existing knowledge of the model. Although the proposed feature space is almost resistant to concept drifts, we incorporate incremental learning so that the model can adapt to new data without forgetting its existing knowledge, in such a way it will adapt quickly to a very slow change in distribution of the data. For our learner, we use a SGD classifier with hinge loss and L2 regularizer, this results

Datasets	Feature	es (Magn	itude Only)	Feature	es (Phase	Only)	Augmented Features			
	SVM	RF	Inc. SVM	SVM	RF	Incr SVM	SVM	RF	Inc SVM	
Apt 1	77.94	59.84	54.85	55.09	46.04	38.92	80.52	60.31	70.67	
Apt 2	90.2	87.8	67.26	78.8	88.67	74.33	92.0	89.8	75.12	
Apt 3	78.24	60.5	35.6	32.95	36.12	38.3	74.53	63.95	83.35	

Table 1	1:	Accuracy	in	%	of	the	learning	g alg	orithms	for	different	feature	space
							· · · · · · · · · · · · · · · · · · ·	··· (-)					

in an SVM that can be updated incrementally. For the incremental learning framework, we keep on updating the model parameters for each round with the augmented feature space, where minimal drift is present. We then test the models on the rounds, where the feature space corresponding to only phase or the magnitude would perform poorly, and use our proposed stable feature space for the multi-class classification problem.

Table 1 shows the performance of different learning algorithms for CSI features incorporating magnitude only, phase only and the proposed augmented feature space. The table shows that for all the learning algorithms the augmented feature space performs better and is more resistant to drifts. Also we note that in case of streaming data, the performance of incremental SVM described is consistently better for the augmented feature space. Thus we show that in case of large incoming data, the augmented feature space presents more robust features in terms of representing the WiFi CSI for localization.

5 CONCLUSION AND FUTURE WORK

We have presented a comprehensive study in order to handle drifts for WiFi CSI data. We focused on the challenges presented by drifts for the application of indoor localization and proposed a combined feature space that is robust to drifts. We then incorporate this augmented feature space and provided a detailed analysis of the performance of different learning algorithms. Although we mainly focus on off line training, our work also focuses on robust online prediction in the presence of drifts. Such a stable feature space will will mean that we do not have to learn the abrupt and gradual drifts and retrain our models each time when there one. Our proposed feature space will also allow for applying deep convolution neural network, that has been only applied to either the phase or the magnitude information, but not both. The proposed feature space can be projected into an RGB image where, vital information can captured using a convolution layer which we keep for future work.

REFERENCES

- Xi Chen, Chen Ma, Michel Allegue, and Xue Liu. Taming the inconsistency of wi-fi fingerprints for device-free passive indoor localization. In *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*, pp. 1–9. IEEE, 2017.
- João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):44, 2014.
- Negar Ghourchian, Michel Allegue-Martinez, and Doina Precup. Real-time indoor localization in smart homes using semi-supervised learning. In AAAI, pp. 4670–4677, 2017.
- Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.
- Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3): 18–22, 2002.
- Souvik Sen, Božidar Radunovic, Romit Roy Choudhury, and Tom Minka. You are facing the mona lisa: spot localization using phy layer information. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 183–196. ACM, 2012.

- Xuyu Wang, Lingjun Gao, and Shiwen Mao. Phasefi: Phase fingerprinting for indoor localization with a deep learning approach. In *Global Communications Conference (GLOBECOM)*, 2015 *IEEE*, pp. 1–6. IEEE, 2015a.
- Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Deepfi: Deep learning for indoor fingerprinting using channel state information. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pp. 1666–1671. IEEE, 2015b.
- Xuyu Wang, Xiangyu Wang, and Shiwen Mao. Cifi: Deep convolutional neural networks for indoor localization with 5 ghz wi-fi. In *Communications (ICC)*, 2017 IEEE International Conference on, pp. 1–6. IEEE, 2017.
- Zheng Yang, Zimu Zhou, and Yunhao Liu. From rssi to csi: Indoor localization via channel response. *ACM Computing Surveys (CSUR)*, 46(2):25, 2013.