Bachelor's thesis

Machine Learning-based User Movement Prediction in Layer 2 Networks

Vorhersage von Benutzerbewegungen in Layer 2 Netzwerken basierend auf Maschinellem Lernen

Lina Wilske

Hasso Plattner Institute at University of Potsdam

July 23, 2023

Bachelor's thesis

Machine Learning-based User Movement Prediction in Layer 2 Networks

Vorhersage von Benutzerbewegungen in Layer 2 Netzwerken basierend auf Maschinellem Lernen

by
Lina Wilske

Supervisors
Prof. Dr. Holger Karl
Chair for Internet Technologies and Softwarization
Leonard Paeleke
Chair for Internet Technologies and Softwarization

Hasso Plattner Institute at University of Potsdam

July 23, 2023

Abstract

...

Contents

1 Introduction **1**
- 1.1 Roaming in Wireless Networks . . . 1
- 1.2 Machine Learning for AP Prediction . . . 1
- 1.3 Data Set for Machine Learning . . . 1
- 1.4 Machine Learning Model . . . 1
- 1.5 Implementation and Improvements . . . 2
- 1.6 Evaluation . . . 2

2 Data for machine learning model **3**
- 2.1 Getting data . . . 3

3 Discussion **5**
- 3.1 Suitable machine learning algorithm . . . 5

4 Implementation **7**
- 4.1 Preprocessing . . . 7
- 4.2 Machine Learning Model . . . 7

5 Evaluation **9**
- 5.1 Adapting parameters . . . 9

6 Conclusion **11**
- 6.1 Conclusion . . . 11

References **13**

1 Introduction

1.1 Roaming in Wireless Networks

In big Wireless Fidelity (Wi-Fi) scenarios with multiple Access Points (APs) in office buildings, shopping malls, and airports people are moving around indoors with their mobile phones and need to be connected to an AP. This is called roaming and is a very important part of Wi-Fi, since 802.11k[2]. Roaming was improved with AP-initiated roaming, which is a feature of 802.11r[1]. But this roaming process does not include human movement, because if the AP1 knows, that the station is passing it and moves towards AP2, the AP2 should initiate the roaming process. This is not possible with the current roaming process, because the AP1 does not know, that the station is moving towards AP2.

1.2 Machine Learning for AP Prediction

Machine learning is used to predict human movement in other scenarios like digital health or telecommunications. So what if we could predict the next AP and initiate the roaming process before the station is moving towards the next AP? In this thesis, the prediction will be based on the Received Signal Strength Indications (RSSIs) of the Basic Service Set Identifiers (BSSIDs) and waypoints of the station.

1.3 Data Set for Machine Learning

We need to decide, if we want to generate the data on our own or use existing data. We want to use machine learning to predict the next AP with data generated by users devices. Generating data on our own is not an option, because of the time and effort it would take to collect data. So we decided to use existing data. We will use a data set from kaggle[4] of a competition of Microsoft Research in 2021[3]. In this chapter, we will analyze the data. Furthermore, we will identify which parts of the data is needed for the machine learning model.

1.4 Machine Learning Model

There are several machine learning models which could be used. We will take a look into some pre-selected models and discuss and propose a model, which could be the best for our data. This thesis will not concept a new machine learning model and also will not use combined models.

1.5 Implementation and Improvements

We will implement the decisions made in the previous chapters. As mentioned in 1.3, we need to preprocess the data for our model. We will prefilter specific parts of the data, because as we found out in 1.3, not all data is needed for the machine learning model. In this chapter, we will preprocess the data and implement the machine learning model as well as discuss the chosen encryption. A big part of machine learning models is tweaking the hyperparameters.

1.6 Evaluation

At the end we will evaluate the model and discuss the results. We will also discuss, if the model could be used in a real-world scenario.

2 Data for machine learning model

2.1 Getting data

- ...

3 Discussion

3.1 Suitable machine learning algorithm

- ...

4 Implementation

4.1 Preprocessing

- Preprocessing important step of ML
- Data is not consistent
  - Wi-Fi and waypoint Data are not measured at the same time (some could be event triggered, but just speculation)
  - First: Interpolate waypoints to the timestamps of Wi-Fi data
  - Second: Merge interpolated waypoints and Wi-Fi Data
  - Detected jumps in time and in position
  - Present solutions: Split data into several parts, where the position more than 10 meters from the last position or time difference more than 60 minutes
  - time difference of more than 60 minutes could also lead to a jump in position: Therefore, position more than 10 meters away
  -

4.2 Machine Learning Model

5 Evaluation

5.1 Adapting parameters

- ...

6 Conclusion

6.1 Conclusion

- ...

Bibliography

[1] "IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Fast Basic Service Set (BSS) Transition". In: *IEEE Std 802.11r-2008 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008)* (July 2008). Conference Name: IEEE Std 802.11r-2008 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008), pages 1–126. DOI: 10.1109/IEEESTD.2008.4573292.

[2] "IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs". In: *IEEE Std 802.11k-2008 (Amendment to IEEE Std 802.11-2007)* (June 2008). Conference Name: IEEE Std 802.11k-2008 (Amendment to IEEE Std 802.11-2007), pages 1–244. DOI: 10.1109/IEEESTD.2008.4544755.

[3] *Indoor Location & Navigation | Kaggle*. https://www.kaggle.com/competitions/indoor-location-navigation. (Visited on July 11, 2023).

[4] *Kaggle: Your Home for Data Science*. https://www.kaggle.com/. (Visited on July 23, 2023).

Acronyms

AP	Access Point
BSSID	Basic Service Set Identifier
RSSI	Received Signal Strength Indication
Wi-Fi	Wireless Fidelity

Zusammenfassung

...

Eidesstattliche Erklärung

Hiermit versichere ich, dass meine Bachelor's thesis "Machine Learning-based User Movement Prediction in Layer 2 Networks" ("Vorhersage von Benutzerbewegungen in Layer 2 Netzwerken basierend auf Maschinellem Lernen") selbstständig verfasst wurde und dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt wurden. Diese Aussage trifft auch für alle Implementierungen und Dokumentationen im Rahmen dieses Projektes zu.

Potsdam, den 23. Juli 2023,

_____

(Lina Wilske)