

# CSI 5325 Project Final Report

Denton Wood

4/24/2021

## 1 Introduction

Location-based online services have become an increasing part of our lives in the advent of smartphones. The initial development of these services centered around navigation and were designed to provide real-time directions to the user from point A to point B. However, developers quickly realized that these same services could be used to connect individuals with each other and with advertisers. Recent phenomenons such as the summer of Pokémon GO [1] and social media location filters [2] have demonstrated the success of corporation-driven location-based technologies.

These services are traditionally powered by Global Positioning System (GPS) technology. While GPS technology works well outdoors, it is less precise indoors. Current location technology has difficulty accurately placing users as they wander through large buildings like shopping malls, especially if the building has multiple levels. Another option is to use nearby signals such as Wi-Fi, Bluetooth, beacons, etc. Wi-Fi in particular is a strong contender for measurement due to its near-ubiquity in public indoor areas. Many establishments such as coffee shops and restaurants allow consumers to connect to a public Wi-Fi access point for convenience; others, such as administrative offices, use Wi-Fi for internal access. This technology may be of interest to companies who can leverage it for advertising purposes [3].

This project is an attempted solution to the Kaggle problem "Indoor Location & Navigation" [4]. The goal of the project is to detect location with some accuracy in large indoor facilities such as shopping malls or marketplaces using indoor signals. The company, XYZ10, in partnership with Microsoft Research, wants to use this method for a variety of reasons, including location-based marketing offers and autonomous devices. We attempt to solve this problem using a Long Short-Term Memory (LSTM) neural network and looking only at strong Wi-Fi signals from the data. We discuss our proposed approach and the status of our work.

## 2 Related Work

Recurrent neural networks (RNNs) were first proposed as "dynamic error propagation networks" in 1987, just a year after backpropagation made multi-layer neural networks feasible [5]. The authors proposed maintaining a local state for each member of the network in addition to the input in order to analyze the output of previous inputs alongside new inputs. This facilitated solutions for more advanced problems for neural networks, such as speech processing.

The problem with traditional RNNs is the problem of all multi-layer neural networks: vanishing and exploding gradients. Using backpropagation repeatedly with no regularization or error control exponentially shifts values away from the usable spectrum, defeating the purpose of learning. To solve this problem, Hochreiter and Schmidhuber introduced LSTM models [6]. These networks are composed of individual units called cells which filter irrelevant inputs at both the input and output cells. Like RNNs, each unit of an LSTM model has the concept of both external input and internal state input; however, LSTM models additionally introduce the idea of both internal state and memory inputs calculated by different portions of the neurons in each layer of the network. [7]. LSTM models also introduce additional hyperparameters for tuning these inputs. While additional variants such as bidirectional LSTM models have been introduced, we use a mainline LSTM model for this paper.

LSTM models are a popular solution for indoor location detection using Wi-Fi signals [8, 9, 10]. This is likely because location data is highly temporally and spatially correlated, making it feasible for sequential

analysis using recurrent networks. We hope to replicate or even exceed the success of these works on the XYZ10 data by fine-tuning the LSTM using regularization techniques. While other machine learning techniques have been tried for this problem, including k-means clustering [11, 12, 13], regression [11], support vector machines [12], and random forest classifiers [13], we believe we can achieve better results using networks which are specifically suited for sequential analysis.

### 3 Methodology

The original data for the project consists of two parts: sensor readings and locations. The sensors used include Bluetooth, Wi-Fi, and other indoor sensors. The locations are series of (x,y) points. The data is organized by building floors, and each file corresponds to a path walked on a single floor. All data comes from building sensors from hundreds of buildings in China. For this project, we will use a preprocessed version of the data which eliminates all readings other than Wi-Fi readings with a strong signal (BSSID  $\geq 1000$ ) [14].

To begin, we will need to do some preprocessing of the data in order to feed it to the LSTM model. Currently, the data consists of timestamped Wi-Fi signal measurements with an occasional location "waypoint" reading. We will first attach each Wi-Fi reading to the temporally closest waypoint. Since a waypoint may have a variable number of measurements attached to it, we will take the average number of measurements per waypoint and divide the measurements among copies of the waypoint value for any waypoint with more than that number of measurements. This ensures that the data matrix X has a static number of columns.

We will then feed the data to an LSTM model with an embedding layer to convert the data into a processable format. The output will have two dimensions: an x-coordinate and a y-coordinate which corresponds to the nearest waypoint. The number of neurons required for each layer will depend on the number of features present in the data. We will implement the model using the Keras API, backed by TensorFlow.

To train the model, we will run cross-validation using an 80/20 split of training & validation data. The data has already been partitioned into training and test sets, so we simply need to partition the training data as well.

### References

- [1] Zsila, G. Orosz, B. Bőthe, I. Tóth-Király, O. Király, M. Griffiths, and Z. Demetrovics, "An empirical study on the motivations underlying augmented reality games: The case of Pokémon Go during and after Pokémon fever," *Personality and Individual Differences*, vol. 133, pp. 56–66, Oct. 2018.
- [2] V. Lubbers, "The Stories of Our Lives: Online identity performance and construction on Snapchat," 2018.
- [3] M. Bues, M. Steiner, M. Stafflage, and M. Krafft, "How Mobile In-Store Advertising Influences Purchase Intention: Value Drivers and Mediating Effects from a Consumer Perspective," *Psychology & Marketing*, vol. 34, pp. 157–174, Feb. 2017. Publisher: John Wiley & Sons, Inc.
- [4] Microsoft Research, "Indoor Location & Navigation," 2021. <https://kaggle.com/c/indoor-location-navigation>.
- [5] A. J. Robinson and F. Fallside, "The Utility Driven Dynamic Error Propagation Network," Cambridge University Engineering Department, Nov. 1987.
- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997.
- [7] P. Rivas, *Deep Learning for Beginners: A beginner's guide to getting set up and running with deep learning from scratch using Python*. Birmingham, UK: Packt Publishing, 2020.

- [8] M. Elbes, E. Almaita, T. Alrawashdeh, T. Kanan, S. AlZu'bi, and B. Hawashin, "An Indoor Localization Approach Based on Deep Learning for Indoor Location-Based Services," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 437–441, Apr. 2019.
- [9] Z. Chen, H. Zou, J. Yang, H. Jiang, and L. Xie, "WiFi Fingerprinting Indoor Localization Using Local Feature-Based Deep LSTM," *IEEE Systems Journal*, vol. 14, pp. 3001–3010, June 2020. Conference Name: IEEE Systems Journal.
- [10] G. Hussain, M. S. Jabbar, J.-D. Cho, and S. Bae, "Indoor Positioning System: A New Approach Based on LSTM and Two Stage Activity Classification," *Electronics*, vol. 8, p. 375, Apr. 2019. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- [11] S. Pasricha, V. Ugave, C. W. Anderson, and Q. Han, "LearnLoc: a framework for smart indoor localization with embedded mobile devices," in *Proceedings of the 10th International Conference on Hardware/Software Codesign and System Synthesis, CODES '15*, (Amsterdam, The Netherlands), pp. 37–44, IEEE Press, Oct. 2015.
- [12] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, "An enhanced WiFi indoor localization system based on machine learning," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, Oct. 2016. ISSN: 2471-917X.
- [13] E. Jedari, Zheng Wu, R. Rashidzadeh, and M. Saif, "Wi-Fi based indoor location positioning employing random forest classifier," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–5, Oct. 2015.
- [14] D. Anzelmo, "Indoor Navigation and Location Wifi Features," 2021. <https://kaggle.com/devinanzelmo/indoor-navigation-and-location-wifi-features>.