Improving Indoor Localization of Mobile Devices by Generating Training Samples Using a Modified Transformer

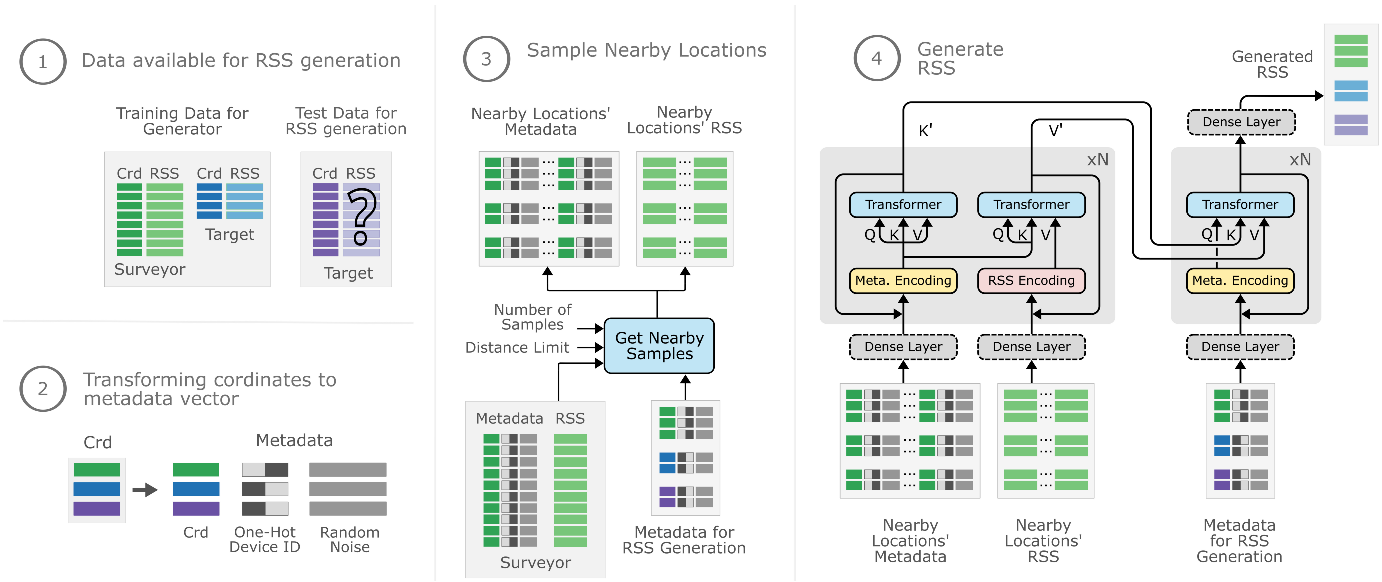


Fig. 1: Detailed overview of the proposed method for RSS generation. Crd denotes the smart device's location coordinates.

Technical Overview

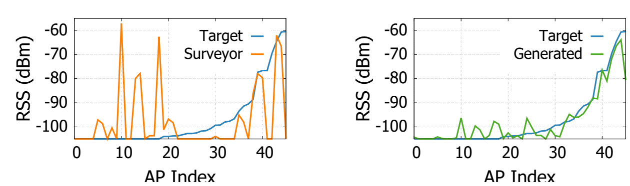
- **Libraries** – Tensorflow, Keras, scikit-learn
- **Technologies** – Deep Learning, Transformer Network, Device Localization
- **Programming Languages** – Python

Project Summary

The failure of GPS based location estimation of mobile devices in indoor settings demands alternate solutions. One such solution is the use of WiFi Access Points" (APs) received signal strength (RSS) information to locate a mobile device in indoor settings, as the RSS information from multiple APs serves as a reliable fingerprint of that location. Unfortunately, the addition, removal or movement of APs cause this fingerprint to change resulting in environmental variations. Also, the RSS values recorded for a given location vary from device to device, as RSS values are subjected to the properties of the receiving antenna used in each device. Hence every time the APs change or a new mobile device is introduce the laborious task of training set creation ensues. To alleviate some of the effort of collecting new training set samples, we propose a RSS generator by using a modified Transformer network. With this Transformer old samples from a reference device/environment, that we term as a *surveyor*, are used to generate synthetic samples for the new device/environment, that we term as a *target*. In addition to the large surveyor training set, the Transformer makes use of limited target samples to estimate the RSS properties of the new target device/environment.

Key Findings

To train the network to generate target RSS values, we use few target RSS samples as labeled output and their corresponding coordinates (metadata) as input to the network. In addition to the coordinates, we use 6 RSS samples from the surveyor and their coordinates as input as well. Think of this as using English sentence words to predict the next work in the French translation of that sentence. After training, the self-attention layer is able to use the surveyor RSS input and surveyor & target metadata values to compute attention weights and eventually predict the target RSS output.



(a) Surveyor vs Target RSS (b) Generated vs Target RSS

Fig. 2: Comparison of generated, surveyor, and target RSS samples for an unseen test location.

As seen in Fig. 2(a), even though surveyor and target RSS values differ drastically for the same location, our generator framework accurately generates RSS values similar to the target (see Fig 2(b)). This allows us to generate synthetic target RSS samples for training from previously unseen locations, helping improve the training set.

For complete details about this study refer to the publication. URL https://doi.org/10.1109/ICCWorkshops50388.2021.9473553

SELE: RSS Based Siamese Embedding Location Estimator for a Dynamic IoT Environment

TECHNICAL OVERVIEW

- **Libraries** – Tensorflow, Keras, scikit-learn
- **Technologies** – Deep Learning, Siamese Networks, Device Localization, Representation Learning
- **Programming Languages** – Python

PROJECT SUMMARY

While GPS is a reliable way to estimate the location of a mobile device in an outdoor setting, the quality of this estimation degrades drastically in an indoor environment like car parking, office space, or warehouse. Considering the ubiquity of WiFi Access Points (APs) in indoor environments, received signal strength (RSS) information from multiple APs can be used as a reliable fingerprint of a location. ML approaches have been proven to be good estimators of device location in indoor spaces by using RSS fingerprints. But as the environment changes, i.e, due to the addition/removal of APs or AP movement, the RSS fingerprint of that location changes with it. To make things more complicated, the RSS values recorded in each mobile device vary based on the physical properties of the receiving antennas. Hence the effort taken to create a training set by collect RSS information and its corresponding coordinates is wasted when a new device is introduced or the environment changes. To counter this problem we propose a Siamese Network based solution.

Our framework tries to handle temporal variations and device heterogeneity by introducing an intermediate feature space. Instead of directly using RSS as input features to train a ML classifier, we first use them to learn location-specific embeddings using a Siamese Network. This initial device/environment in which the training set is created is called the *surveyor device/environment*. When a new device is introduced or the environment changes, we call it the *target device/environment*. By making use of the initial training samples from the surveyors as reference and limited samples from the target, we can map target RSS values to the same location specific embeddings.

KEY FINDINGS

This transfer learning approach significantly reduces the amount of new training data required from the target device/environment to handle variations in RSS. But most importantly, by using these location-specific invariant embeddings as input features to any localization framework, the prediction pipeline need not be retrained or updated every time the RSS values change. The only thing that needs to be done is to learn a mapping from the new RSS values to the embedding vector of that location.

As expected, the Siamese Network trained on the surveyor training set generates invariant location-specific embeddings for the surveyor (see Fig. 2) but fails to do so for the target device (see Fig. 2(a)). To fix this we use transfer learning, where the target RSS values are passed through 2 feed-forward layers. During tuning, the weights of the entire network are frozen, and only the tuning layer parameters are learned. The result of this can be seen in Fig. 2(b), where invariant embeddings are generated irrespective of the device/environment.

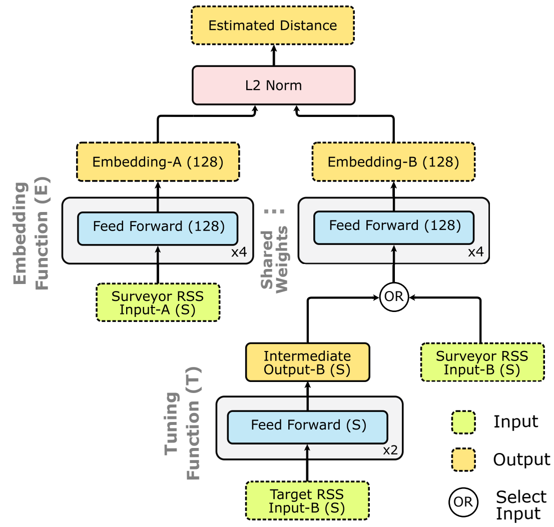
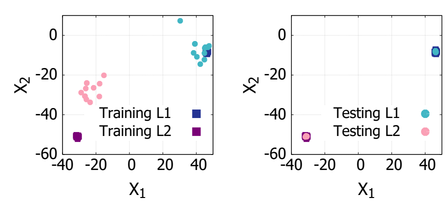
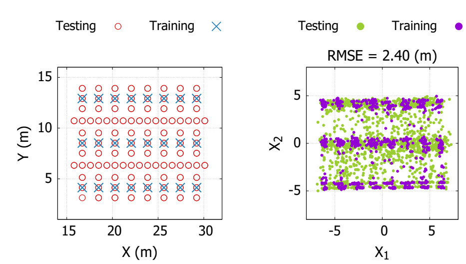


Fig. 1: Illustration of Siamese-Network architecture



(a) (b)

Fig. 2: 2D PCA plot of embeddings for two locations generated for target RSS (lighter shade) (a) without tuning layers, (b) with tuning layers.



(a) (b)

Fig. 3: (a) Coordinates corresponding to RSS samples, (b) embeddings of the same RSS samples. The 2D PCA plot of the embeddings resembles actual coordinates even for test set.

For complete details of this study refer to the publication. URL: https://ieeexplore.ieee.org/document/9490634

Study of the Twitter Follower Network to Characterize the Spread of Prescription Drug Abuse Tweets

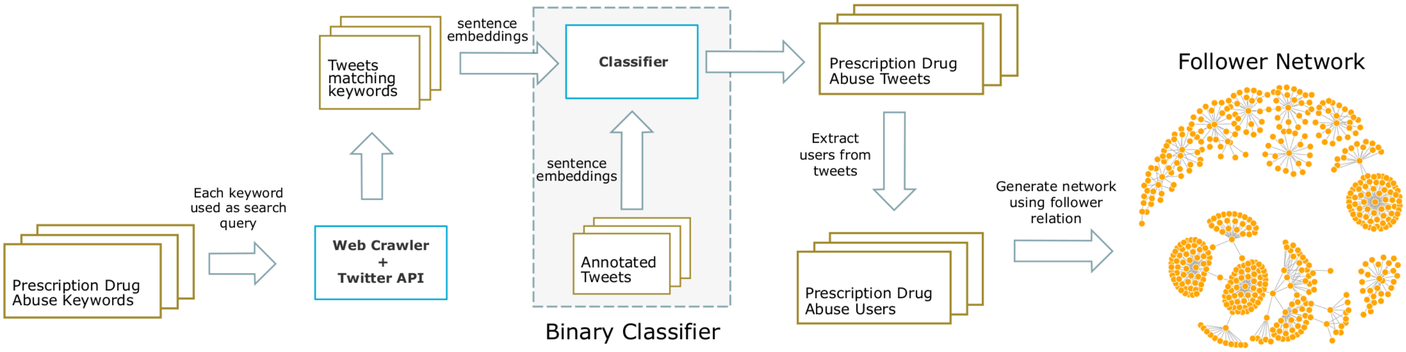


Fig. 1: Steps of follower network formation. This network and the content of its users was used to study prescription drug abuse related behavior on Twitter.

Technical Overview

- **Libraries** – NetworkX, Tensorflow, Keras, scikit-learn, Twitter API
- **Technologies** – Deep Learning, Natural Language Processing, Network Science
- **Programming Languages** – Python

Project Summary

Previous literature on this subject has pointed out that there is substantial evidence to indicate that Twitter is used to promote drug abuse (DA) [1], and in some cases sell illicit drugs [2]. The goal of our study was to investigate the network effect of Twitter that may amplify the spread of DA tweets among users. For this, we manually annotated $8,400$ tweets. In addition to some hand-crafted features, sentence embeddings (Sent2Vec) of these tweets were used to train a SVM classifier. Out of the 2.2 million tweets collected from the Twitter API using keyword search using drug names, the classifier identified around 0.77 million tweets (36% of total tweets) of $420,502$ unique users as prescription DA tweets. We performed our analysis on this data.

Key Findings

By collecting the follower information of these users we observed that there existed a large network consisting of approximately 0.42 million unique users, with 17 million links between them. 99.85% of the total nodes in the network were part of the giant component, i.e., there existed a path that connected any two nodes in this component. This high local connectivity hinted at multiple alternate communication channels among these users. Combining this information with user activity revealed that fringe users (i.e. users with low followers) were primarily responsible for spreading DA content (refer Fig. 2(a))

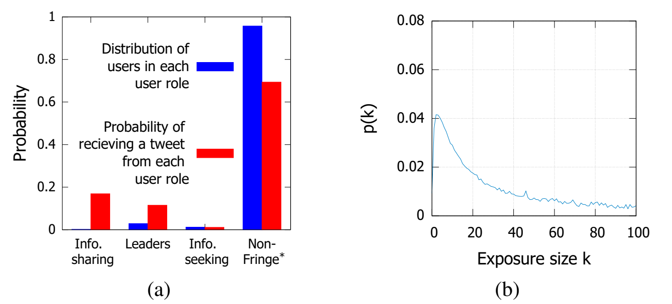


Fig. 2: (a) Probability distribution of users in each category and the probability distribution of receiving a DA tweet from each user category. (b) # exposures vs probability of posting of DA tweet.

Investigations on the cascades of DA tweets helped us discover that these tweets spread over long paths across the Twitter follower network through groups of closely connected users in the network. We also observed a significant percentage of cascades were initiated and driven by fringe users with low positional importance in the network. Finally, our observations suggested that drug abuse promoters on Twitter have a much higher risk of adopting newer drugs as increasing exposure of tweets promoting these new drugs enhances the probability of their adoption (refer Fig. 2(b)).

For complete details of this study and our findings refer to the publication. URL: https://doi.org/10.1109/TCSS.2019.2943238

References

[1] H.-W. Meng, S. Kath, D. Li, and Q. C. Nguyen, "National substance use patterns on twitter," *PLOS ONE*, vol. 12, no. 11, pp. 1–15, 11 2017.
[2] T. K. Mackey, J. Kalyanam, T. Katsuki, and G. Lanckriet, "Twitter-based detection of illegal online sale of prescription opioid," *American journal of public health*, vol. 107, no. 12, pp. 1910–1915, 2017.