UNIVERSITI TEKNIKAL MALAYSIA MELAKA

# Faculty of Electronic and Computer Engineering

# WI-FI NAVIGATION USING MACHINE LEARNING WITH CONSIDERATION OF CYCLIC DYNAMIC BEHAVIOUR

**Ahmed Salih Khirbeet**

**Doctor of Philosophy**

**2020**

# WI-FI NAVIGATION USING MACHINE LEARNING WITH CONSIDERATION OF CYCLIC DYNAMIC BEHAVIOUR

**AHMED SALIH KHIRBEET**

**A thesis submitted**
**in fulfillment of the requirements for the degree of Doctor of Philosophy**

**Faculty of Electronic and Computer Engineering**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2020**

# DECLARATION

I declare that this thesis entitled "Wi-Fi Navigation Using Machine Learning with Consideration of Cyclic Dynamic Behaviour" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

| | | |
|---|---|---|
| Signature | : | ............................................... |
| Name | : | Ahmed Salih Khirbeet |
| Date | : | ............................................... |

**APPROVAL**

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

| | | |
|---|---|---|
| Signature | : | ……………….................................................... |
| Supervisor Name | : | Dr. Mohd Riduan Bin Ahmad |
| Date | : | …………………….................................... |

# DEDICATION

I dedicate this thesis to Allah Almighty my creator, my strong pillar, my source of inspiration, wisdom, knowledge, and understanding. He has been the source of my strength throughout this program and on His wings only have I soared.

My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life;

My homeland Iraq, the homeland of civilization, glories, and tournaments;

The great martyrs, the symbol of sacrifice;

My great parents, who never stop giving of themselves in countless ways,

My beloved sisters;

My friend and elder brother Mohammed Hashem, who stands by me when things look bleak,

My friends who encourage and support me, especially;

Ahmed Ali Al-Saffar, Mohammed Ahmed Jubair, Mustafa Hamid Hassan, Taj-Aldeen Naser Abdali, Fahad Taha Al-Dhief

All the people in my life who touch my heart,

I dedicate this research.

# ABSTRACT

Wi-Fi based localization using machine learning has been proven to be an attractive approach for finding the location prediction with avoidance of accumulation of errors as other sensors such as odometry and inertial sensing. Researchers have developed various models to predict locations based on trained machine learning. A site survey is typically performed to collect fingerprints and a neural network is trained on those fingerprints. The trained model is then placed into operation. However, dynamic changes in the location and navigation behavior of users make the prediction process more challenging in terms of accurate prediction of location. One common mobility behavior of navigation runs is the cyclic dynamics or re-visiting the same place more than one time. Most machine learning models, developed for location prediction, lack sufficient handling of dynamic changes or leveraging them for better predictions. To fill this gap, this study builds a new simulator with two components: one for incorporating dynamic information of navigation in given Wi-Fi dataset and using them to generate the corresponding time series of any navigation run, it is named as Wi-Fi Simulator for Cyclic Dynamic (Wi-Fi-SCD) while the other is useful for converting any dataset to time series with cyclic dynamics, it is named as Cyclic Dynamic Generator (CDG). Furthermore, in this study, two novel location prediction machine learning models were developed. The first is Knowledge Preservation Online Sequential Extreme Learning Machine (KP-OSELM) and the second is Infinite Term Memory-based Online Sequential Extreme Learning Machine (ITM-OSELM). The KP-OSELM model is distinctive from other models cited in the literature, because it preserves knowledge gained in certain areas to restore again when the person re-visits the area again. In KP-OSELM, knowledge is preserved within the neural network structure and is enabled based on feature encoding. The ITM-OSELM model is distinctive from other models cited in the literature, because it carries external memory and transfers learning to preserve old knowledge and restoration. ITM-OSELM is more efficient than KP-OSELM when the percentage of active features is low. Meanwhile, KP-OSELM does not require any external blocks to be added to the neural network (unlike ITM-OSELM), which makes it much simpler. In area based scenarios, KP-OSELM and ITM-OSELM both achieved accuracies of 68%. Moreover, when evaluating KP-OSELM and ITM-OSELM on Wi-Fi-SCD, for three navigation cycles, the highest accuracies achieved were 92.74% and 92.76%, respectively. However, the execution time of KP-OSELM was 1176 second while much less time was needed for ITM-OSELM to be executed with a value of 649 second. Furthermore, when evaluating KP-OSELM and ITM-OSELM on CDG, for three cycles, 100% accuracy was achieved for both models. As a conclusion, this study has provided the literature of machine learning in general and Wi-Fi navigation in particular with various models to support the localization without any restriction on the type of Wi-Fi that is used and with consideration of the practical and dynamic behaviors that can be leveraged to improve the localization performance.

# *ABSTRAK*

*Pengenalpastian lokasi berasaskan Wi-Fi menggunakan pembelajaran mesin terbukti menjadi pendekatan yang menarik untuk mencari ramalan lokasi dengan mengelakkan kesilapan yang dilakukan oleh sensor lain seperti odometer dan penginderaan inersia. Penyelidik telah membangunkan pelbagai model untuk meramalkan lokasi berdasarkan pembelajaran mesin terlatih. Tinjauan tapak biasanya dilakukan untuk mengumpul cap jari dan rangkaian saraf yang telah dilatih menggunakan cap jari tersebut. Model yang telah dilatih kemudian diletakkan dalam sistem operasi. Walau bagaimanapun, perubahan dinamik di lokasi dan tindak balas navigasi pengguna membuatkan proses ramalan menjadi lebih mencabar dari segi ramalan ketepatan lokasi. Tingkah laku pergerakan umum navigasi adalah kitaran dinamik atau lawatan semula ke tempat yang sama lebih daripada satu masa. Kebanyakan model pembelajaran mesin, dibangunkan untuk melakukan ramalan lokasi, kekurangan pengendalian terhadap perubahan dinamik atau memanfaatkannya akan menjejaskan ramalan. Melalui kajian ini, simulator baharu dibina dengan dua komponen iaitu satu, untuk memasukkan maklumat dinamik navigasi dalam data set Wi-Fi yang diberikan dan menggunakannya untuk menjana siri masa yang sesuai bagi mana-mana jangka masa navigasi yang dikenali sebagai Wi-Fi Simulator for Cyclic Dynamic (Wi-Fi-SCD) manakala yang lain adalah berguna untuk menukar mana-mana data set kepada siri masa dengan dinamik kitaran yang dinamakan sebagai Cyclic Dynamic Generator (CDG). Dalam kajian ini, dua model pembelajaran ramalan lokasi telah dibangunkan. Yang pertama adalah Mesin Pembelajaran Terperinci Pengekalan Pengetahuan Dalam Talian (KP-OSELM) dan yang kedua ialah Mesin Pembelajaran Ekstrem Talian Terperinci Memori Tanpa Batasan (ITM-OSELM). KP-OSELM, memelihara pengetahuan dalam struktur rangkaian saraf dan berfungsi berdasarkan pengekodan ciri. Model ITM-OSELM adalah tersendiri berbanding model-model kajian lepas yang lain kerana membawa memori luaran dan melakukan pemindahan pembelajaran untuk memelihara serta memulihkan pengetahuan terdahulu. ITM-OSELM adalah lebih cekap daripada KP-OSELM apabila peratusan ciri aktif berada pada tahap yang rendah. Sementara itu, KP-OSELM tidak memerlukan sebarang blok luaran untuk ditambah ke rangkaian saraf (tidak seperti ITM-OSELM) yang menjadikan operasinya adalah lebih mudah. Dalam senario kawasan, KP-OSELM dan ITM-OSELM kedua-duanya mencapai ketepatan 68%. Penilaian ke atas KP-OSELM dan ITM-OSELM di Wi-Fi-SCD, bagi tiga kitaran navigasi, menunjukkan ketepatan tertinggi iaitu 92.74% dan 92.76%. Walau bagaimanapun, masa pelaksanaan KP-OSELM adalah 1176 saat manakala lebih sedikit masa diperlukan untuk ITM-OSELM dilaksanakan dengan nilaian 649 saat. Selain itu, ketika menilai KP-OSELM dan ITM-OSELM menggunakan CDG, selama tiga kitaran, kedua-dua model telah mencapai ketepatan 100%. Kesimpulannya, kajian ini telah menyediakan kajian lepas mengenai pembelajaran mesin secara umum dan navigasi Wi-Fi khususnya dengan pelbagai model untuk mengenal pasti lokasi tanpa sebarang sekatan ke atas jenis Wi-Fi yang boleh dimanfaatkan untuk meningkatkan prestasi ketepatan pengenalpastian lokasi.*

# ACKNOWLEDGEMENTS

I am grateful to the Almighty for His guidance and blessing throughout my pursuit of the degree of Doctor of Philosophy. He made this journey possible by granting me the patient, the courage and the strength. With His permission, an aspiration has become a reality.

I would like to express my gratitude to my principal supervisor, Dr. Mohd Riduan Ahmad for his advice, guidance, support, encouragement and understanding throughout this academic journey. He has always been inspiring role model and the source of motivation. It has been my privilege to have him as my supervisor. My sincere appreciation goes to my co. supervisor Associate Professor Dr. Azmi Awang Md Isa and Dr. Mona Riza Mohd Esa for the knowledge and wisdom that he willing to share with me. This thesis may have not been possible without their guidance and encouragement.

Also, I would like to thank all the lecturers, administrators, and staff of Universiti Teknikal Malaysia Melaka (UTeM) and all academic and non-academic staffs of the Graduate school for their help and support.

Most importantly, I would like to thank my parents for everything. Thank you for the prayer, love and confidence in me. Thank you for teaching me to always do what is right and believes in myself. To all my dearest friends, a big thank you for their support and encouragement and for putting colors in my life, may Allah bless you all.

## TABLE OF CONTENTS

**LIST OF TABLES**

## LIST OF FIGURES

## LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| AcMu | - | Automatic and continuous radio map self-updating service |
| AFKF | - | Adaptive fingerprint kalman filter |
| AMKL | - | Adaptive multiple kernel learning |
| ANNs | - | Artificial neural networks |
| AOA | - | Angle of arrival |
| APs | - | Access points |
| BLE | - | Bluetooth low energy |
| BS | - | Base station |
| CDG | - | Cyclic dynamic generator |
| COSELM | - | Constraint online sequential extreme learning machine |
| CSI | - | Channel state information |
| DA | - | Domain adaptation |
| DAELM | - | Domain adaptation extreme learning machine |
| DANN | - | Discriminant adaptive neural network |
| DBSCAN | - | Density-based spatial clustering of applications with noise |
| DCs | - | Discriminative components |
| DNN | - | Deep neural networks |
| DOS | - | Denial of service |
| DTMKL | - | Domain transfer multiple kernel learning |
| EDA | - | Extreme learning machine based domain adaptation |
| ELM | - | Extreme learning machine |
| EM | - | External memory |
| F1 | - | F-measure |
| FA-OSELM | - | Feature adaptive online sequential extreme learning machine |
| FCR | - | Feature change ratio. |
| FN | - | False negative |
| FNR | - | False negative rate |

| | | |
|---|---|---|
| FP | - | False positive |
| FPR | - | False positive rate |
| GPS | - | Global positioning system |
| GSM | - | Global system for mobile communication |
| HMM | - | Hidden markov model |
| ILBSs | - | Indoor location-based services |
| IM | - | Improvement in measure |
| IMU | - | Inertial measurement units |
| IoT | - | Internet of thing |
| IPS | - | Indoor positioning systems |
| IR | - | Infrared |
| ISM | - | Industrial, scientific and medical |
| ITM-OSELM | - | Infinite term memory online sequential extreme learning machine |
| JSD | - | Jensen-shannon divergence |
| KDD | - | Knowledge discovery and data mining |
| KDDA | - | Kernel direct discriminant analysis |
| k-NN | - | k-nearest neighbor |
| KP-OSELM | - | Knowledge preserving online sequential extreme learning machine |
| kWNN | - | k-weighted nearest neighbor |
| LEDs | - | Light emitted by diodes |
| LOS | - | Line of sight |
| LTE | - | Long-term evolution |
| LuMA | - | Localization using manifolds alignment |
| MAC | - | Maximum accuracy change |
| MAC | - | Medium access control |
| MDA | - | Multiple discriminant analysis |
| ML | - | Machine learning |
| MLP | - | Multi-layer perceptron |
| MN | - | Measure in new model |
| MO | - | Measure in old model |
| MU | - | Mobile unit |

| | | |
|---|---|---|
| NA | - | Number of features in location A |
| NB | - | Number of features in location B |
| NC | - | Number of features in location C |
| NCF-AB | - | Number of common features between A and B |
| NCF-BC | - | Number of common features between B and C |
| NIC | - | Network Interface card |
| NLOS | - | Non line of sight |
| NN | - | Neural networks |
| OSELM | - | Online sequential extreme learning machine |
| PCA | - | Principle component analysis |
| PDR | - | Pedestrian dead reckoning |
| PGFE | - | Parameterized geometrical feature extraction |
| PNAF | - | Percentage of new active features |
| PNN | - | Probabilistic neural network |
| POAF | - | Percentage of old active features. |
| PoI | - | Point of interest |
| PSD | - | Power spectral density |
| R2L | - | Root to local |
| RBF | - | Radial basis function |
| RBF | - | Rank based fingerprinting |
| RELMs | - | Robust extreme learning machines |
| RF | - | Radio frequency |
| RFID | - | Radio frequency identification |
| RMM | - | Ratio-based map matching |
| ROWA | - | RandOm k-sample sets feature for the weighting approach |
| RPs | - | Reference points |
| RSS | - | Received signal strength |
| RSSI | - | Received signal strength indication |
| RTT | - | Round trip time |
| RVR | - | Relevance vector regression |
| SD | - | Standard deviation |
| SDA | - | Stacked denoising autoencoder |
| SK | - | Segmental k-means |

| | | |
|---|---|---|
| SLAM | - | Simultaneous localisation and mapping |
| SLFN | - | Feedforward neural network |
| STI | - | Similarity metric, termed signal tendency index |
| SVM | - | Support vector machine |
| TDOA | - | Time difference of arrival |
| TKL | - | Transfer kernel learning |
| TL | - | Transfer learning |
| TN | - | True negative |
| TNR | - | True negative rate |
| TOA | - | Time of arrival |
| TP | - | True positive |
| TPE | - | Target position estimation |
| TPR | - | True positive rate |
| TTP | - | Target tracking process |
| U2R | - | User to root |
| UWB | - | Ultra-wideband |
| VLC | - | Visible light communication |
| VLP | - | Visible light positioning |
| WAPs | - | Wireless access points |
| WELM | - | Weighted extreme learning machine |
| Wi-Fi-SCD | - | Wi-Fi simulator for cyclic dynamic |
| WinSMS | - | Wi-Fi-based non-intrusive sensing and monitoring system |
| WLAN | - | Wireless local area network |
| WSN | - | Wireless sensor network |

# LIST OF SYMBOLS

| | | |
|---|---|---|
| $PT$ | - | Period of time |
| $mp$ | - | Marker points |
| $v$ | - | Velocities between each two stop points |
| $N_1$ | - | Number of repeating the point of the pause in the time series |
| $N_2$ | - | Number of repeating point $p_t$ of the trajectory in the time series |
| $FR$ | - | Framerate of the sensors |
| $T$ | - | Pauses times at the stop points |
| $p_t$ | - | Points in the path |
| $Res.x$ | - | The x axis of grid granularity |
| $Res.y$ | - | The y axis of grid granularity |
| TS | - | Time Series |
| $D$ | - | Original dataset |
| $D_t$ | - | Time series dataset |
| $FCR$ | - | Feature change rate |
| $NF$ | - | Number of features |
| $\langle B_t \rangle$ | - | Number of ones in the vector |
| R | - | That represents the number of records |
| $C$ | - | Number of classes |
| $y_t$ | - | The class that is extracted from D at moment t |