

Opposition report for degree project

Version 2.1 – March 13, 2019

DV2572: Masters thesis course in Computer Science

January 26, 2020

Opponent	Name	Hangdong Chen
	e-Mail	hach17@student.bth.se
	Social security nr	9611221297
Thesis	Title	Implementation of BLE Beacons in a Healthcare Centre for Indoor Localization using Machine Learning
	Author(s)	K. Shiv Sidhartha

1 Introduction

This study addresses the movement and scheduling problems caused by frequent movement of patients and devices, using BLE (Bluetooth Low Energy) beacon and fingerprinting technology, the performance of four algorithms, random forest regression, ridge regression, linear regression, and K-NN regression, are compared. A dataset was established by collecting RSSI, a prediction experiment was performed using five-fold cross validation and calculated the Euclidean distances. In the analysis, the average absolute error and the root mean square error were compared. It is concluded that random forest regression performs better than other three.

This report describes suggestions for K. Shiv Sidhartha's Master Thesis work under the guidance of Prof. Yulia Sidorov from the department of Computer Science & Engineering. The purpose of this report is to provide some suggestions for changes to opponents. I mainly proposed some changes to the structure of the article, recommend that the experimental design details and algorithm implementation methods should be described, and the results be discussed in more depth.

2 Critical review

This section will describe the strength and weaknesses of the paper.

First I make a simple summary of the content and structure of the article.

- In Chapter 1 introduction, briefly introduced the development and role of indoor positioning services, and raised the problem that staff wasted a lot of time in finding medical equipment. To address the challenges, two research questions were raised.
- In Chapter 2 background, the basic concepts of indoor positioning system, iBeacon, RSSI and fingerprinting are briefly introduced, the advantages of BLE beacon are described, and four machine learning models are briefly introduced.
- In Chapter 3 related work, a brief description of some comparative research and research results on indoor positioning algorithms is given.

- In Chapter 4 method, literature review is used for to determine suitable machine learning methods for indoor positioning using BLE. And uses experiment method to measure the performance of learning algorithms. The Euclidean distance errors, MAE and RMSE of four algorithms are compared.
- In Chapter 5 result, experiments were performed for each algorithm and Euclidean distance errors were collected. Cumulative Distribution Function was plotted
- In Chapter 6 analysis, the differences between MAE and RMSE were compared. Box plots of RMSE and MAE are drawn for a clearer comparison. The proposed RQs were answered separately. The threat of effectiveness and the limitations of experiments are briefly described. It is believed that a larger data set is needed to get more reliable results.
- Chapter 7 briefly summarizes the results of the experiment.

Through reading the paper, I found the following strength:

- Content for tables and figures help readers easy to review.
- A clear description of the motivation for the research question, and the research question also fits the logic of the objective.
- The basic technology and algorithm are described.

Through reading the paper, I also found some points strange or can be improved:

1. The section 4.1 literature review describes search keywords but does not describe the results.
2. The description of chapter 3 seems described the result of Literature Review.
3. The experimental method should establish and prove research hypotheses, but I have not found relevant hypotheses and hypothesis verification in the article.
4. The description of the indicator comparison is incomplete. Three performance indicators are described in 4.8, but only the average absolute error and the root mean square error are compared in the analysis table.
5. The expression is not clear enough. As a reader, I will encounter difficulties in reading and understanding.
6. Analysis of conclusions and future work is too simple.

3 Required changes

In this section, detailed required problem description and related suggestions will be mentioned.

Formulate questions:

This research focuses on the location of medical facilities and hopes to provide solutions. In the field of research, the algorithm analysis and comparison of indoor positioning are mainly carried out, but in the research and discussion, I did not find a description related to the actual situation. For example, will Bluetooth technology be greatly affected if it is complex and changeable? If the impact is great, is the BEL technology suitable for use in located medical facilities? In the medical indoor positioning, people walk around and move equipment very frequently, and the number of equipment is also large, which poses a great challenge to the accuracy of positioning. If the positioning accuracy is low or the error distance is large, it is difficult or even impossible to improve the efficiency of finding medical facilities. I recommend analyzing the actual situation of the medical center to see which technology is more advantageous to determine the significance of the research

Method – Literature review

This paper uses literature review to determine experimental machine learning algorithms. There are few suggestions for literature review method.

1. In section 4.1, the key strings for searching literature review are explained, and inclusion criteria are also mentioned. But the search result and literature review are missing. The result of literature review provides the algorithm for experiment, so as a reader, I am confused that why you compared these four algorithms.
2. The choice of algorithm should be explained in more detail in this method. For example, in chapter 3, it was mentioned that the performance of a new model KNN-RF is good, but why was only KNN regression algorithm selected in this research?
3. The related work in Chapter 3 seems to describe related research algorithms and research results. I personally think that this is the result of literature review.

Method – Experiment

This paper uses experiment method to compare algorithm differences. There are suggestions for experiment method.

1. The experimental method specifies the dependent and independent variables, but does not make experimental hypotheses. Differences in algorithms should be explained using hypothesis testing.
2. The machine learning model was briefly introduced in 2.8, but the design of the machine learning algorithm was not described in detail. For example, in the K-NN algorithm, the choice of K value will greatly affect the results of the prediction, but there is no relevant description in this paper.
3. The collection description of the data set is not detailed enough. For example, what is the distribution of the samples among 1916 data points? Does the points picked randomly or have a regular pattern?

4. Explain why use five-fold cross-validation, rather than ten-fold cross-validation. In general, the more training sets that can be trained to fit the model, will there be better results if the ten-fold cross-validation is used?

Analysis

In the analysis, the attributes of the three parameters are not described. For example, in the Euclidean distance error index, the smaller the distance error, the closer the predicted point and the actual position are. The data itself is not meaningful. The information the data represents is what we need to discuss.

three performance indicators are described in 4.8, but in analysis part, only MAE and RMSE are compared in the performance indicator comparison (Table 6.1).

Although the CDF chart was drawn in the results, there was no description of this item in the discussion. I don't understand the motivation and the meaning of the CDF chart. Please explain.

4 Recommended changes

1. In addition to the required changes mentioned above, there are some recommended changes.
2. The expression is not concise. As a reader, I will encounter difficulties in reading and understanding. Reduce particularly long and verbose sentences, describe in more concise way.
3. Some abbreviations are used in the paper, such as BLE. If add an abbreviation list, it will make the reading more clear.
4. The replacement of the null value with the minimum possible value is described in Section 4.4. Please explain why and explain the minimum possible value.
5. In Chapter 3, some relevant research are mentioned, which are quite different from the conclusions of this study. For example, in the study of Alexander and Kusuma [1], BLE beacon and fingerprint technology were also used to collect data in this study. Random forest and linear regression models were compared. The conclusion is that linear regression is better than random forest regression, which is quite different from this study. I suggest it is necessary to verify the effectiveness of the algorithm implementation and make some analysis of the results difference in conclusions or future work.
6. Compared to some of the studies mentioned in Chapter 3, the Euclidean distance error obtained in this paper is much higher than the results under the same algorithm. The reasons for this can be discussed.
7. In Section 6.6, if the validity threat analysis uses “what kind of threats, why it is threatened, and how it is resolved” format description will help reader to understand.
8. The description of future work is too brief. In fact, there are many factors not considered in this research, such as how to perform in complex situations in actual situations, and that the

experimental results and the research results are very different from research mentioned in Chapter 3. These can be used as future work and discussed.

References

- [1] Ivan Alexander and Gede Putra Kusuma. Predicting indoor position using bluetooth low energy and machine learning