University of Glasgow | School of Computing Science

Assessed Coursework

Course Name	**CSC3105 Data Analytics and AI**			
Coursework Number	**1**			
Deadline	**Time:**	**23:59**	**Date:**	**1 April 2026**
% Contribution to final course mark	**35%**			
Solo or Group	**Solo** ✓		**Group**	✓
Anticipated Hours	**10 per group member**			
Submission Instructions	**1. Submit via SIT_Xsite a Zip or compressed tar file containing your source code directory (UoG-DA.tgz or UoG-DA.zip), one DAReportGroupxx.pdf file, Youtube video plus group presentation video(if needed)** **2. Declaration Form**			
Please Note: This Coursework cannot be Re-Done				

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below.

The primary grade and marks awarded for coursework which is submitted after the published deadline will be calculated as follows:

(i) in respect of work submitted not more than four working days after the deadline

- the work will be assessed in the usual way;
- the primary grade and mark so determined will then be reduced by 15% for each working day (or part of a working day) the work was submitted late.

(ii) work submitted more than four working days after the deadline will be awarded Grade F.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause to Admin-In-Charge

Penalty for non-adherence to Submission Instructions is 2 bands
You must complete an "Own Work" form

Data Analytics (UoG-DA) 2025-26

Assessed Exercise: Data Analytics in Line-of-Sight (LOS) & Non-Line-Sight(NLOS) Wireless Signal Predictor in Indoor Environment for Precise Positioning

Introduction

The goal of this mini-project exercise is to familiarize yourselves with the design, implementation, and performance testing of the Line-of-Sight (LOS) & Non-Line-Sight(NLOS) Ultra-WideBand (UWB) wireless signal classification prediction. Given the problem statement with the real empirical dataset, you will be required to go through the whole cycle of problem definition, problem analysis, algorithm and pseudocode design. You will need to demonstrate the three stages of Data Analytics, namely **D**ata preparation (Data cleaning and preprocessing), **D**ata Mining and **D**ata Visualization with result analysis.

Assessed task

You will be working in group of 5-6 members. Your group are expected to work on the given real-life problem statement with substantial dataset to perform Data Analytics and AI to design a proposed data analytics/machine learning solution. The solution to the problem statement is open ended and not constrained if it can be implemented with your own novelty using any data mining algorithms be it classification, probabilistic, regression, association or clustering baselined on data analytics design concept and evaluation. The focus is not about getting the full correct answer but strong familiarization of the 3D process of Data Analytics/Data Science. As such, teamwork is cardinal requirement.

Every member of your group is expected work cohesively to search, brainstorm to derive the team design that is unique and holistic in solution and implementation. You can search such design online or any research database as following but not limited to. (You can access the IEEE Xplore and ACM using your Glasgow GUID and password).

IEEE Transactions on Big Data

https://ieeexplore-ieee-org.ezproxy.lib.gla.ac.uk/search/searchresult.jsp?newsearch=true&queryText=IEEE%20Transactions%20on%20Big%20Data

IEEE Big Data Mining and Analytics

https://ieeexplore-ieee-

org.ezproxy.lib.gla.ac.uk/xpl/issues?punumber=8254253&isnumber=9430128

IEEE Access

https://ieeexplore-ieee-org.ezproxy.lib.gla.ac.uk/xpl/tocresult.jsp?isnumber=9312710&punumber=6287639

ACM SIGKDD International Conference on knowledge discovery and data mining

https://dl-acm-org.ezproxy.lib.gla.ac.uk/conference/kdd/proceedings

IEEE International Conference on Data Mining

https://ieeexplore-ieee-org.ezproxy.lib.gla.ac.uk/xpl/conhome/1000179/all-proceedings

ACM Conference on Recommender Systems

https://dl-acm-org.ezproxy.lib.gla.ac.uk/conference/recsys/proceedings

IEEE Conference on Data Science and Advanced Analytics

https://ieeexplore-ieee-org.ezproxy.lib.gla.ac.uk/search/searchresult.jsp?newsearch=true&queryText=IEEE%20International%20Conference%20on%20Data%20Science%20and%20Advanced%20Analytics

IEEE Conference on Big Data and Analytics

https://ieeexplore-ieee-org.ezproxy.lib.gla.ac.uk/xpl/conhome/1823704/all-proceedings

or any academic reference that will address below mentioned problem statement such as IEEE transactions on wireless communication.

In general, the given problem statement must allow the following implementation

1) Data preparation (Data Cleaning and Data Preprocessing)
2) Data Mining Algorithm
3) Data Visualization
4) Data Analysis

LOS/NLOS UWB signal estimation for precise localization

Precise localization has been one of the key pillars in many various commercial and government applications, for example autonomous applications such as autonomous vehicle, drone, robot, keep tracking.

https://www.youtube.com/watch?v=krez6-QvgAw
https://www.youtube.com/watch?v=k3hpgBzD3xk
https://www.youtube.com/watch?v=C7xOZknlkEE

To provide precise localization, Global Positioning System (GPS) is prevalent for outdoor environment such as your daily Google Map Navigation or Grab/Uber services. However, for indoor environment, GPS cannot penetrate indoor and as such there are many means such as using Radio Frequency approach such as Wi-Fi, UWB, Bluetooth, 5G to locate the target, the most popular being Samsung Smart Tag and Apple Air tag

https://www.youtube.com/watch?v=rJGl6_crZmw

https://www.youtube.com/watch?v=Zqu3Pz22TZ0

https://www.youtube.com/watch?v=C7xOZknlkEE

https://www.youtube.com/watch?v=ZdkmPmtxz5U

https://www.youtube.com/watch?v=DzZeOm1f_zk

Ultra-Wideband (UWB) is a high bandwidth wireless signal operating at 3.1GHz to 10 GHz, providing high resolution position data. However, like all RF based approach, the anchor (such as phone transmitter) and the tag (receiver) must be in visual LOS allow correct ranging or position localization as shown in Fig. 1

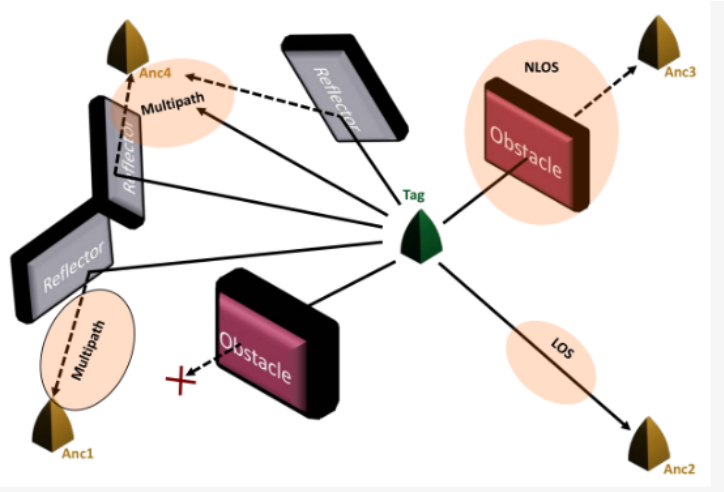

Fig. 1: Illustration of LOS, NLOS and multi-path (MP) scenarios in a UWB-based ranging system [1]

As shown, in indoor environment, it is not always possible to have visual contact between

anchors and tag due to obstacle such as walls, doors etc. We would have many anchors deployment in visual LOS with the tag which result in cost (man effort to deploy and unit cost), efficiency and computation that impedes the business implementation. One of the current research projects is to explore understanding the indoor environment through the channel impulse response (CIR) which is the signal propagation between the anchor and tag as shown in Fig. 2.

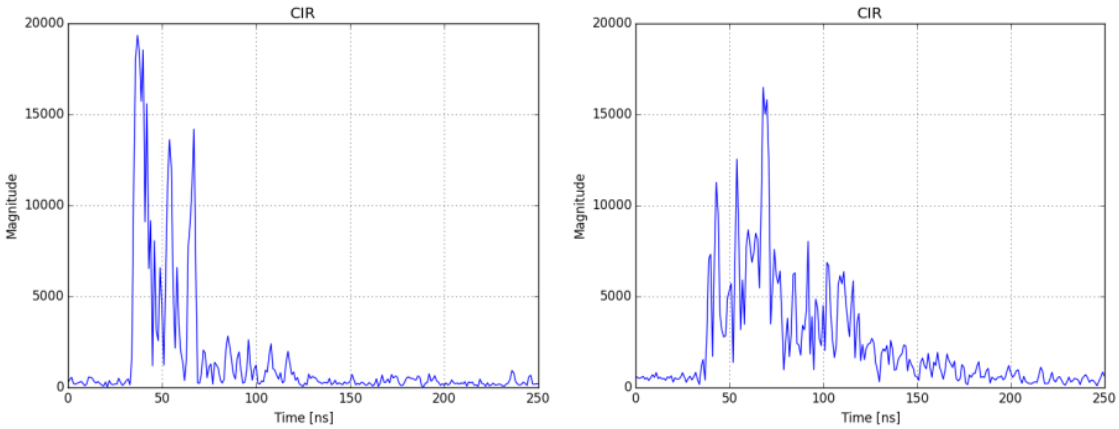

Fig. 2: UWB LOS CIR (left) and UWB NLOS CIR[2].

The objective is to recognize the CIR pattern in various indoor environment and ability to map the CIR to LOS condition or NLOS condition. In the case of NLOS condition, the pair of measurement between the anchor and tag will be discard. Only LOS pair will be used for ranging and localization. Below is some reference

[1] https://www.mdpi.com/2076-3417/10/11/3980

[2] https://www.researchgate.net/publication/308986067_NLOS_Channel_Detection_with_Multilayer_Perceptron_in_Low-Rate_Personal_Area_Networks_for_Indoor_Localization_Accuracy_Improvement

You are given a dataset, UWB-LOS-NLOS-Dataset. This is the dataset that is based on Decawave (DWM1000) measurement. The company, Decawave is acquired by Apple UWB vendor (Qorvo). The dataset contains real empirical CIR information for 7 indoor environments mainly

1. Office 1
2. Office 2
3. Small Apartment
4. Small Workshop
5. Kitchen with a living room
6. Bedroom
7. Boiler room

In every indoor location 3000 LOS samples and 3000 NLOS samples were taken. Different locations were chosen to prevent building of location-specific LOS and NLOS models. All together 42000 samples were taken: 21000 for LOS and 21000 for NLOS

channel condition. To make data set ready for building LOS and NLOS models, samples are randomized to prevent overfitting of a model to places.

There are 15 features and 1 Class namely

Class : NLOS(1 if NLOS, if LOS)
Feature:

1. Measured range (time of flight)
2. FP_IDX (index of detected first path element in channel impulse response (CIR) accumulator: in data set it can be accessed by **first_path_index+15**)
3. FP_AMP1 (first path amplitude - part1) [look in user manual](http://thetoolchain.com/mirror/dw1000/dw1000_user_manual_v2.05.pdf)
4. FP_AMP2 (first path amplitude - part2) [look in user manual](http://thetoolchain.com/mirror/dw1000/dw1000_user_manual_v2.05.pdf)
5. FP_AMP3 (first path amplitude - part3) [look in user manual](http://thetoolchain.com/mirror/dw1000/dw1000_user_manual_v2.05.pdf)
6. STDEV_NOISE (standard deviation of noise)
7. CIR_PWR (total channel impulse response power)
8. MAX_NOISE (maximum value of noise)
9. RXPACC (received RX preamble symbols)
10. CH (channel number)
11. FRAME_LEN (length of frame)
12. PREAM_LEN (preamble length)
13. BITRATE
14. PRFR (pulse repetition frequency rate in MHz)
15. CIR (absolute value of channel impulse response: 1016 samples with 1 nanosecond resolution)

A understanding of the Decawave can be found at https://www.decawave.com

The objective is to train using the given dataset is **twofold**. Firstly, you should recognize whether inside a CIR measurement, the two shortest dominant paths is either a pair of LOS and NLOS or both are NLOS. Secondly, you should also predict the measured range for these two paths so that accurate localization using these two paths can proceed. Your team should design your ML training to obtain such LOS/NLOS classifier. The expectation is on the 3D process rather than accuracy. The team is not restricted to the given dataset for training and testing after the LOS/NLOS classifier model and distance estimation model have been developed and tested. The team can go beyond to obtain new dataset online to show the competency of the team. However, this requirement is not compulsory.

For ML, each group is free to use any ML algorithm that deem fit such as supervised (classification or regression) or unsupervised (clustering) learning for

1) prediction of the LOS/NLOS classifier for **two dominant shortest paths**
   (Note: If the first path is LOS, next path will be NLOS. If first path is NLOS, next path will be NLOS too. LOS is always the shortest path if it exists)
2) distance estimator for **two dominant shortest paths.**
   (Hint: Use FP_IDX and measured range to correlate to next second dominant path

if possible)

In terms of implementation, each project group should clearly work out but not limited to

a. Data Preparation/Preprocessing
   I. Decide is there a need for data reduction?
   II. Decide is there a need for data feature extraction or transformation?
   III. Decide do you need to do class labelling since there are two paths classification and estimation?
   IV. Decide the need for data cleaning?
   V. Is there feature importance (rank the feature importance, namely which feature is the most important and rank accordingly)
   VI. Is there a need to create synthetic data for performance robustness
b. Data Mining
   I. Decide whether are you using supervised or unsupervised learning? What algorithm should be preferred for your team?
   II. Decide the split ratio of the training/test dataset such as 70:30 or 80:20?
   III. Determine the classification, regression performance accuracy such as RMSE, confusion matrix, ROC etc?
c. Data Visualization
   I. Plot the various performance indicators to illustrate the team choice and various result in part (a) and (b)
d. Result Analysis
   I. Justify the results with theoretical understanding.

What to hand in

According to your Group number, use

- SIT-Xsite Dropbox to submit a single zip or compressed tar file with the contents of the UoG-DA plus a separate DAReportGroupxx.pdf file and the scanned version of the signed declaration forms of all members in the group

To aid in testing and assessing of your code, please make sure that:

1) Your submission file is named UoG-DA.tgz or UoG-DA.zip.
2) When uncompressed, your files will be in a folder named UoG-DA_Groupxx. The folder should contain all the following
   a) your dataset and also provide the link to your GitHub that store the data
   b) all python source codes that are necessary to take in the dataset and generate all the plots and the printed results. Also provide the link to your Github that store the code.
   c) DAReportGroupxx.pdf where x and xx indicate your lab session group and

group number respectively according to the grouping list.

d) An overleaf document in latex that summarize your part C) report in 6 pages in double column using IEEE format. Use Glasgow student account. https://www.overleaf.com/latex/templates/ieee-conference-template/grfzhhncsfqn

e) A group power point presentation video of about 25-30mins (if situation don't allow physical or online presentation on 31st March that week) where all group members must give their speech on their respective portion of work. Confirmation will be known on week 9.

f) A short YouTube video of less than 5mins to demonstrate the problem and the implementation with results.

3) Your DAReportGroupxx.pdf file should outline and contain

a) The name of the group member and the **individual contribution**.

b) your design solution in terms of

- problem definition,
  - What are your group trying to solve and why is it important to solve this problem?
- problem analysis,
  - What is the novel solution and its originality that your group is going to propose?
  - How does your group break down the problem and give data analytics approach in providing the solution?
- algorithm
  - What are the algorithms in the solution from data preparation, data mining to the data analysis?
  - What are the evaluation criteria such as dissimilarity/similarity measure and test accuracy methodology such as confusion matrix, RMSE, ROC etc?
- Pseudocode for the algorithm

c) all the source codes

d) clear mathematical and logical explanation on all the plots and results

e) any interesting aspects of your solution (e.g., assumptions you've made, optimization that you thought of, etc.).

4) Your overleaf document should be summarized version of your report and provide the link and pdf version. It should outline as usual conference format such as

a) Abstract

b) Introduction (Literature Review) highlighting what is the problem statement and state-of-the-art solution out there. What is the weakness of the state-of-art-solution and what is your proposed solution.

c) Your proposed solution description

d) Experimental/Simulation Result and Discussion

e) Conclusion

5) Your submission will be tested for plagiarism. Plagiarism cases will be dealt on a

case-by-case basis but suffice to say there will be little tolerance.

How this exercise will be marked

Following timely submission, the exercise will be given a numerical mark between 0 (no submission) and 100 (perfect in every way). The numerical marks will then be converted to a grade. The marking scheme as a group is as follows:

- 10 marks: Approach

  For the approach that clearly aligns and articulates the data analytics concept and methodology, namely the 3D of Data Analytics and AI (Data Preparation, Data Mining and Data Visualization). Marks will be awarded for solution in terms of originality and partial or full two dominant path classification and estimation.

- 40 marks: Implementation.

  A clear implementation and articulation on the practice of Data Analytics and AI for partial or full two paths classification and estimation in terms of

  a. Data Preparation
  b. Data Mining
  c. Data Visualization
  d. Result Analysis

- 20 marks: Demonstration and Presentation

  A concise and clear demonstration by the group (and individual) of the proposal and solution for partial or full two paths classification and estimation in terms

  a. Problem statement and its importance
  b. Proposed solution (with the "3D") and its originality
  c. Proposed Implementation/Methodology and Result Analysis
  d. Demonstration
  e. Conclusion and Future Work

- 30 marks for Report/Code/Overleaf document/GitHub

  a. As stated in the report requirement in earlier section on submission on DAReportGroupxx.pdf file and overleaf document in pdf and link.
  b. quality and structure of the code; make sure that you use appropriate names for variables/function, library, plot and that your source code is properly structured.
  c. comments and documentation in the code and your DAReportGroupxx.pdf file; make sure that you comment and document in your source files, at the very least, the basic steps taken to comment on the function, library and various variables used. Do not make an essay of your code; use your pdf file to discuss further details.

The final marking scheme for individual will be as follows:

Each group member grade = (weighted peer review score by group member with his/her presentation during group presentation) *Group score