Huawei University Challenge Competition 2021

Data Science

for

Indoor positioning

Contents

1 Introduction

This competition is an annual programming competition funded by European Research Institute (ERI) and organized by Huawei Technologies Research & Development (UK) Ltd. with challenge topics from different fields of ICT. It creates opportunities to inspire undergraduates, post-graduate, and research (PhD) students to come up with innovative ideas and experience hands on problem solving. This year's challenge is inspired by geospatial data analysis to provide the participants an experience exploring smart cities sensing data under the theme of indoor positioning & navigation.

1.1 About the challenge

Indoor Positioning Solutions (IPS) make it possible to navigate GNSS denied spaces and boost location context data availability, creating the foundation for many scientific, technical, and industrial applications. Guided navigation, virtual or augmented reality, location sharing, business intelligence, security services and other technologies already utilise indoor location contextual information. Furthermore, data science fields such as artificial intelligence, data analysis, network analysis, social media analysis and knowledge graphs all need the ability to annotate location data at scale. A very common IPS employs radio-maps of Wi-Fi signals as reference points to compute smartphone positions. However, keeping a global radio-map up to date by survey or data collection is challenging, time consuming and inefficient. Hence, it is essential to be able to perform qualitative data mining to extract radio-maps from nosier data submitted via crowd-sourcing.

The challenge is implementing algorithms that can construct geospatial graphs from unlabelled Wi-Fi data points collected in unknown locations. How could we create and utilise such graph to map the geospatial relations between 3D spaces?

This challenge encourages participants to solve the problem in the setting of network analysis using graphs as data models. Participants will be given two tasks related to prediction of the geospatial relations. In both tasks the candidate algorithm should output a prediction that is as close as possible to the target attributes in the actual geospatial relation graphs. The winning algorithms are supposed to achieve the

best precision regardless of the processing time or computational overheads.

1.2 Judging Criteria Explained

On the opening of the competition platform, you now have access to all materials for Task1, Geospatial Distance Estimation. To be qualified and gain access to the materials for the second task, your team need to submit a valid solution for the first task. The following points try to explain the main criteria for enhancing your team chance to compete:

- ✓ A valid solution should satisfy the minimum score of 0.2 for task1.
- ✓ On each valid submission scores are recalculated as a proportion of best performing algorithm precision (a score of 1.0) and the minimum accepted score (a score of 0).
- ✓ Your focus to improve the score will require improving the precision and recall of your model.
- ✓ The first task will account for 35% of the total score, and the second task for the remaining 65%.

1.3 Prizes

1st prize: £ 7,000.00 (to divide equally between the team members) **+ Huawei MatePad Pro for each member**

2nd prize: £ 5,000.00 (to divide equally between the team members) **+ Huawei Watch 3 for each member**

3rd prize: £ 3,000.00 (to divide equally between the team members) **+ Huawei band 6 for each member**

Notes & terms of prize Awarding:

- ✓ *To qualify for the prize the team need to have minimum of 2 members and maximum of 3 and all of them must be enrolled in the University.*
- ✓ *Before awarding, team members will be asked to present proof of association with a University, such as a valid student card or study certificate of the university.*
- ✓ *Before awarding, code review will be conducted to verify that the submitted answers can be recreated.*
- ✓ *Cash prizes will be divided equally between the team members and not based on any acknowledged contributions*
- ✓ *Code quality and execution time will be used as a tie breaker, only if the score match between top teams,*

2 Question & Answers

2.1 Participating

What skills I need to be able to solve the challenge? Do I have to be familiar with geospatial data?

In order to participate, you only need to be competent in one form of data science with a good stack of tools. A level of understanding of geospatial relations would help but it should not be essential to provide a valid solution

There are a lot of terms used in the description that I am not familiar with. What is trajectory, fingerprint and RSSI?

The trajectory is a set of steps sequentially connected. Each step would be relatively placed as a point with certain distance and orientation from the step before it. It is very similar to tracing a sketch of movement on a path without knowing the start point.

A fingerprint in signal processing is a term borrowed from the human fingerprints. In theory, each location point on earth would have a unique set of electromagnetic signals, such as WiFi or Bluetooth, that are different from any other location. However, it is not guaranteed that highly similar sets of signals are always in the same place or that the same place will always have the same fingerprint. In the challenge we use WiFi fingerprints only.

The RSSI is an indicator of how strong the WiFi, or other electromagnetic signal, is when received by the antennas on the smartphone. It is usually measured in dBm (a logarithmic relative power measurement with respect to 1 mW).

Do I need to form a team to participate?

Yes, you have to form a team of at least 2 and maximum of 3 to be part of this competition. We value a team work and would like to see more participants working together.

What should I do to receive the first task? Do I need to be in any online meetings and answer questions?

As you have already signed up for the competition, you only need to login and should find all the information you need for the first task in DataSet1. You are expected to be active and participating in various submission and online chat if required. We recommend that you also signup for our reddit subgroup https://www.reddit.com/r/HuaweiIPS2021/ *. Various meetings will be held to support you and will be announced via email and on reddit chat.*

2.2 Understanding the problem and analysing the datasets

How would I make sure that my understanding of the problem is correct?

The best way to ensure that you understand the problem correctly is to post on the task forum QA section. Other participants will confirm or comment on your problem statement and the mentors from Huawei will confirm or correct any misunderstanding.

How to decode the data to my preferred tool for analysis?

The data are provided in plain text files either in a tabular CSV or standard JSON. ALL IDEs and major stream data science tools should be able to decode the files correctly.

What if I found invalid data?

As the data has been carefully checked and validated, this is not likely. However, in such case, please inform us and provide a screenshot of the steps you did to uncover the invalid data.

What programing language, tech stack or libraries should I use?

We recommend using python but any language and tech stack can be used as long as you provide in the documentation enough information for the jury team to reproduce. Running of your code on a testing platform is not a condition for scoring as we will expect you to provide the predicted optimized answers directly for scoring. Only the finalists will be invited to present their solution and will be subject to results reproduction by the jury team.

How do I train and test my solution before submission?

For both task1 and tast2 we will provide a labelled dataset. This dataset can be used offline to train and test the solution before submission. It will contain labels as ground truth values for you to compute your own score or confusion matrix. However, the final score calculation on our side will use different data and might not comply with your own scoring.

Do you need to verify my code?

We don't need to check your code every time you make better score. However, the finalists will go through code review and need to make sure that their work is document and reproducible. Please make sure that if the algorithm you use is inherently random, you are using random seeds so we can reproduce your score offline during the final review. .

2.3 Code submission, scoring & leader board

How frequently can I submit my solution for scoring?

Each team can only submit answers for scoring twice per day for both tasks combined. Any team member can perform the submission but it will still count toward the limit per team.

What should I provide for submission?

The important file in the submission is the output file that contain your solution answers (prediction). This file is required to be complete in order for the scoring function to work and leader board updating. For the finalists, we also required a compressed folder that contain all the code and document used by the solution with all dependencies well explained. This will only be used to verify the scores of the winning teams before the final awarding. You may choose how best to share that with us or we can provide a submission form at the end of the competition.

Can I change the format of the submission file?

No, unfortunately as the automated scoring system will not accept different format, you will need to follow strictly the output file formatting for the submission or your score will not be updated on the leader board.

Can I get help modifying my output to be consistent with the submission platform?

We can't offer direct 1:1 support, but you can ask questions to clarify the format if there are any doubts. We don't recommend discussing your solution in the chat or forum group though.

What if I think the scoring is not fair?

If you have a strong enough case, we will be happy to investigate and run the scoring manually to review your score. However, this is not something we will do for each submission and a strong case should be presented for approving manual execution of scores.

How can I know if the submission was accepted without issues?

Accepted submissions will show on the leader board with the updated submission date. If your submission didn't show within few hours you can try again as it is likely has been rejected. We will work with the submission platform to try and provide feedback for the majority of failed submission, but this feature is not guaranteed to be feasible for all cases.