Channel Charting-Based Localization Algorithm Competition

Task Description

Prepared by	Sweden Research Center, Algorithm Lab	Date	2024-10

华为技术有限公司
Huawei Technologies Co., Ltd.

修订记录 Revision record

Date	Revision version	change Description
2024-05-31	V1.0	First version
2024-06-06	V1.1	Modifying the background information
2024-06-12	V1.2	Modifying description to the data file
2024-10-28	V1.3	Modifting the task description

1. Introduction

This competition focuses on developing a Channel Charting (CC)-based algorithm for efficient and accurate wireless localization. For additional information on wireless localization and channel charting, a supplementary document named as ***Hackathon2024_TaskBackground.doc*** is provided. In the following, we outline the organization of the competition and participant expectations in Section 2. In Section 3, we provide a summary of the competition's objectives, along with guidance on how to read the provided data and details on the evaluation criteria.

2. Organization of the competition

The competition consists of two rounds**:** **online qualification** and **on-site hackathon**. The first round is held online, and all participants are given ***5*** days to develop/submit their solutions. The top ***15*** teams will then be invited to an on-site competition in Stockholm, Sweden.

2.1 Submission requirements

In the online phase, all teams are only required to submit the output (i.e., the predicted positions) of their algorithms and our competition platform will evaluate the performance of each teams' solutions based on the predicted coordinates and the ground truth data. In the second round, all teams are required to submit:

- **Algorithm output** to our evaluation platform
- **Source codes** (with comments) to the organization team
- **Algorithm documentation** that provides a detailed explanation and discussion of their algorithm.

During the review, multiple senior experts from various departments will evaluate the algorithms, taking into consideration their potential for practical application and integration into real-world products.

2.2 Important Notes

- The allowed programming language for algorithm development is **Python.** Note that algorithm complexity is important, and the solutions that you have developed must finish execution within 20 mins.
- To improve the data validation speed, all stages of the competition require teams to generate and submit their own results in files. The organizer will conduct random checks of the submitted codes (only for the second round). If the submitted output result does not match the result generated by the organizers' execution, it will be considered as cheating and the submission will be disqualified.
- **Each team has 300 submission opportunities.** The ranking will be based on the best submission. Ensure proper local code version management to make sure that the best-performing code is saved and available for submission to the organizer.
- The competition organizer has final authority over all aspects of the challenge. Any deficiencies in this document will be corrected in future updates.

3. Task Description

3.1 Task Summary

The competition aims to enhance the accuracy of localization based on channel charting**. The objective is to design a learning algorithm that predicts the positions of users from their corresponding channel data measured at the BS, given that a small set of channel data is labelled with the true UE positions.** We provide multiple training and evaluation datasets (see Section 3.2 for details). Each dataset includes channel data collected from various locations within a cell served by a single base station (e.g., 20,000 data points). For a subset of these locations (e.g., 2,000 for dataset1 or 500 for dataset2), we also provide the corresponding true 2D coordinates (i.e., positions where the channels are measured). The task is to predict the positions for the remaining channel data points, where true positions are unknown.

Each team is required to design ***an algorithm*** that maps:

- the **input** (the measured channel data) to
- the **output** (the positions associated with the provided channel data)

by using limited anchor points (i.e., channels with known positions).

3.2 Data Description

We will provide the following 3 sets of data:

- Warming up dataset: use this to explore ideas before the actual competition. Filenames start with **Dataset0.**
- ***Online qualification data***: Filenames start with **Dataset1.**
- ***On-site Hackathon data***: Filenames start with **Dataset2.**

[Note] Both **Dataset1** and **Dataset2** contain data measured from 3 different scenes (i.e., 3 different cells of a base station), resulting in a total of 6 scenes. The data across these 6 scenes are uncorrelated, and your algorithm should be evaluated on each of these 6 scenes separately. Note that **Dataset1** contains clean channel data where no impairments are presented. **Dataset2,** on the other hand, is corrupted by both AWGN and and timing advance. For the impairmnets, more information is provided in a complementary document named ***Compelmentary_Data_Description.docx***

In each dataset we include the following (.txt) files:

- Config data: Filenames include ***CfgDataX***, where X stands for scence ID
- Input Channel data: Filenames include ***InputDataX***, where X stands for scence ID
- Provided Position data: Filenames include ***InputPosX,*** where X stands for scence ID. This will be used as the anchor points.

Note that only the warming up dataset (i.e., dataset0) also contains the ground truth positions named ***GroundTruth***, you can evaluate the effectiveness of your proposed solution using this warming up dataset.

To mitigate the complexity, we will provide a sample data reading file in Python. You are welcome to customize your own.

Config Data

The config data file ***CfgData*** contains the following information:

- Line #1: Base station position, example $[0,0,30]$, where 30 is the height of the base station,
- Line #2: Number of collected channel data points, e.g., $20,000$,
- Line #3: Number of points with provided positions, e.g., $2,000$,
- Line #4: Number of UE antennas (SRS ports), e.g., 2,
- Line #5: Number of base station antennas, e.g., 64,
- Line #6: Number of subcarriers in the channel data, e.g., 408.

Channel Data

The input channel data file ***InputData*** is large (20~40GB). Each line contains the actual channel information of each collected point. We provide the format in the following.

- The first line contains the information of the collected channel data corresponding to the first position and the second line to the 2nd position and so on.
- The channel data in every line is provided as a sequence of *N* real numbers separated by spaces, where ***N* = number of base station antennas * number of UE antennas * number of subcarriers * 2** (see the values in the ***CfgData***). The factor *2* accounts for the real and imaginary parts in the channel data. After reading data from each line, the following reshaping should be considered:
  - ❖ (Numpy reshape): (2, number of subcarriers, number of base station antennas, number of **UE** antennas).
  - ❖ You can freely change the shape of the data after this.
- In total, you should expect, for example, 20000 such lines.

Sampled Position Data

The sample position data **file *InputPos*** has the following structure. Each line contains the values:

- UE ***index***, its ***x position***, its ***y position***.
- The UE index is aligned with the line number in the channel data file ***InputData***.
- The postion values are in ***meter***. It has 4 precision places.

3.3 Submission Format

In each stage of the competition (i.e., online qualification and on-site hackathon), we expect that teams would submit their results according to the following description:

- Filename: **DatasetXOutputY.txt,** where X stands for the dataset ID and Y stands for the scene ID.
- The file should be compressed and sent in the .zip format, for which the file name should be given as **Team xxx + Round xxx.zip.**
- Each txt file in the zip file should contain, for example, $20,000$ lines.
- Each line in the submit file should contains the estimated position values for each channel data point indexed in the ***InputData***:
- x position, y position
- Each value has 4 precision places. Longer formats would be truncated in the result evaluation.

The submission documents should be named as follows:

Competition Stage	Dataset	Result Name	Upload File Name
Round 1 & Round 2	Dataset 0	NA	NA

Round 1	Online Qualification	Dataset 1	Dataset1Output1.txt Dataset1Output2.txt Dataset1Output3.txt	Team xxx_Round 1.zip
Round 2	Onsite Hackathon	Dataset 2	Dataset2Output1.txt Dataset2Output2.txt Dataset2Output3.txt	Team xxx_Round 2.zip
Note: Dataset 0 is for you to test the performance of your algorithm				

Note that you should replace the algorithm output corresponding to the anchor points with their true position (i.e., provided by InputPos) to improve accuracy.

3.4 Scoring Metrics

In each round we evaluate the accuracy of the submitted results based on the following guidelines:

- **Localization Error Calculation**: The localization error is computed as the Euclidean distance between the team's estimated position and the actual position coordinates, defined as $d_i = \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}$。
- **Tie in Point Error**: If multiple teams have the same localization error, for example, if Team A has the smallest error and ranks 1st, while Teams B and C have the second smallest error, they will be ranked jointly at 2nd place.

3.5 Additional Channel Information

In this competition, the base station uses a dual polarized 64 antenna array to collect the channel information. In particular, there are 32 atennas in each polarization. The antennas in each polarization are organized into 8 columns, with 4 antenna elements in each column (note that this is important if you want to process your channel into beam-delay domain). The indexing of the antennas is shown in Figure 1.

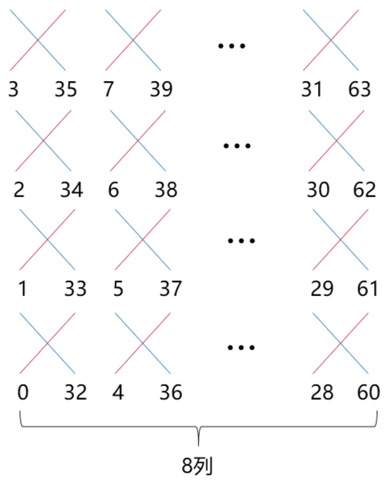

Figure 1: Arrangement and Numbering of the 64 Antennas on the BS Side. Red colored (forward slash shape) antennas belong to the first polarization. Blue colored (back slash shape) antennas belong to the second polarization.

The horizontal antenna spacing is $d_h = 0.5\lambda$ and the vertical antenna spacing is $d_v = 2.0\lambda$.

The data is collected at carrier frequency of 3.5 GHz, with a subcarrier spacing of 30 kHz. Since SRS are transmitted at every second subcarriers, the subcarrier spacing between each SRS resource element (RE) is

60 kHz. With a 100 MHz bandwidth, the number of SRS REs is 1632. Considering the data volume, a 1-out-of-4 sampling method is used, resulting in a final SRS RE (Resource Element) count of 408, with an effective subcarrier spacing of 240 kHz.