

Q1) The missing country is most likely to be Canada. My hypothesis is based on the data analysis in "dataexploration.ipynb".

Q2) The recommendation in the question is on the assumption that the user is a new user (his previous search record is unknown). The implemented function is non-personalized recommendation and it is based on the similarity between cities (item-item collaborative filtering recommendation). The main computation is the similarities. The related files are:

- "dataloader" module,
- "modeltrain" module, "city\_similarity\_matrix" function
- "cityrecommend" module, "city\_recommend" function

For fast recommendation, the similarity matrix can be first dumped to a pickle file. Next time, when recommendation is needed, "city\_recommend" function will load the similarity file for recommendation.

Q3) The feature of users: user\_id, joining\_date, country is useful to predict most likely search cities. Once these information is available, it means that the website knows the user is a registered user and has his record history. It becomes a personalized recommendation problem which can be implemented based on Q2:

- Based on similarity matrix of cities, get the k-nearest neighbours of cities for this user's history records (searched cities).
- Calculate the most similar users to the current user by taking consider of the searched cities, joining year, joining month, joining date of month, and country (user features).
- A combination function that is the weighted similarities from a closet subset of users and the k-neighbours of cities is used for scoring recommended cities.
- Recommend cities based on scores.

Q4) When training a recommendation model, the date set is split into train/test(8:2) set. For the training set, a k-fold(e.g. k=10) cross validate (train/validation set) is needed. For each iteration, the model and prediction is based on training set. The final recommendation is based the average score of recommended cities. The accuracy is the average accuracy measured on the validation set.