FYP 2016 Evaluation and response

Presentation Date:

Project # 8:

Project Name: SmartGuide: A smart campus guide using BLE based indoor localization

Participants:

2016-CE-54	2016-CE-72	2016-CE-81	2016-CS-159
Rida Mehmood	Tooba Naseer	Rabeya Hamood	Ayesha Jabbar

Supervisor: Dr. Sheikh Faisal

Comments:

Explain proposed algorithm. Explain random forest algorithm and how you improve it. Why you used random forest.

Response and steps: Date: 14-02-2020

Proposed Algorithm:
Our selected algorithm is ANN whose explanation is provided in our thesis but as per dept instructions, we decided to make a try and proposed our own algorithm. In our proposed algorithm we combine clustering and classification for better accuracy, firstly we use clustering to cluster the similar observations and then we train the classifier per cluster. Here, we choose random forest as our classifier per cluster. Suppose if we decided to make 4 clusters. Then we built 4 random forest classifier each for one cluster. When we want to predict an unseen instance, it will first assign to the closest cluster then the respective classifier gives its prediction.

Random Forest Explanation:
Random Forest is the ensemble learning classification algorithm. It consists of a large number of individual decision trees. Final output predicted by the random forest is based on the number of individual outputs predicted by each decision tree. The class with the most votes becomes the final output of random forest model. The data set of each decision tree is made up of random selection of features and samples from the original dataset. Decision tree built the classification model in the form of tree structure. It splits the dataset at each node on the basis of selected feature at that node. Feature selection at each node is carried out by calculating gini index/entropy and information gain of each feature and then choose the feature whose information gain value is high at that node. Each branch in the tree represents the decision rule. Each leaf node is assigned to a class. All or majority number of instances at leaf node belongs to that assigned class.

Why we use random forest?
When we do the literature analysis related to our project. We find out that k-NN, random forest and ANN works best on this kind of data. We apply k-NN and random forest on our data and after comparing accuracy, precision and recall, we find out that random forest is better than k-NN. Also, random forest works better than decision tree because single decision tree over fits the data. But when we do further analysis and after applying ANN we find out that ANN outperforms the random forest because ANN has better accuracy, precision and recall on test data as compared to random forest. Then finally, we select ANN algorithm.

How we improve it?
We can improve it by applying our proposed algorithm. One classifier trained per cluster will better learn the subset of data. It will also improve the accuracy of overall predictions on test data. If we train all classifiers for each cluster in parallel then training time might be reduced. Also clustering would be beneficial on the data set that we used.

Supervisor Signature: _____