Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

PARecommender: A Pattern-Based System for Route Recommendation

Feiyi Tang
Victoria University
Melbourne, Australia
feiyi.tang@live.vu.edu.au

Jia Zhu*
South China Normal University
Guangzhou, China
jzhu@m.scnu.edu.cn

Yang Cao
South China Normal University
Guangzhou, China
caoyang@m.scnu.edu.cn

Sanli Ma
CenNavi Technologies Co.,Ltd
Beijing, China
masanlig@cennavi.com.cn

Yulong Chen
CenNavi Technologies Co.,Ltd
Beijing, China
chenyulong@cennavi.com.cn

Jing He
Victoria University
Melbourne, Australia
jing.he@vu.edu.au

Changqin Huang
South China Normal University
Guangzhou, China
cqhuang@m.scnu.edu.cn

Gansen Zhao
South China Normal University
Guangzhou, China
gzhao@m.scnu.edu.cn

Yong Tang
South China Normal University
Guangzhou, China
ytang@m.scnu.edu.cn

Abstract

Widely adoption of GPS-enabled devices generates massive trajectory data every minute. The trajectory data can generate meaningful traffic patterns. In this demo, we present a system called PARecommender, which predicts traffic conditions and provides route recommendation based on generated traffic patterns. We first introduce the technical details of PARecommender, and then show several real cases that how PARecommender works.

1 Introduction

Widely adoption of GPS-enabled devices generates large amounts of trajectory data every day. The trajectory data not only describes the movement history of moving objects but also can produce many useful patterns [Luo *et al.*, 2013]. These patterns can be used to provide traffic prediction and route recommendation in the case that the real-time data source is unavailable, e.g., network outages. Additionally, these traffic patterns can also increase the accuracy of the route planning and estimated time arrival(ETA) calculation.

Most of existing works [Monreale *et al.*, 2009] to recommend route focus on the methods to analyze users' traveling behavior but it is not enough to discover the most efficient route between two locations due to the large number of possible routes and the difficulty in combining trajectory segments. To the best of our knowledge, there are at least two things we can solve the issues. One is to consider the traffic patterns differ significantly between holiday and working day, which will have impacts on route planning and ETA. The other is to design an algorithm that can combine trajectory segments effectively and discover the optimal route. In this demo, we present a system called PARecommender, which not only considers many factors we describe above but also provides a comphensive framework for trajectory data processing.

2 Technical Framework

The overall framework of PARecommender is shown in Figure 1. It comprises three main phases. The first phase retrieves the trajectory data from multiple source, which mainly are the GPS point on Taxi and user generated contents (UGC) from different websites, e.g., Baidu[1] and Tecent[2].

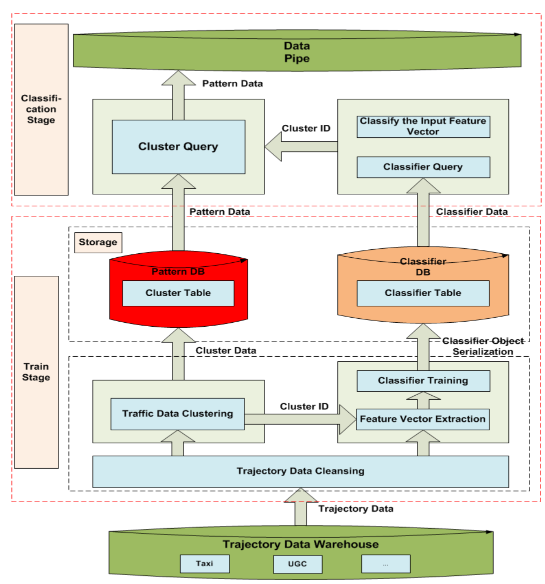

Figure 1: The PARecommender Framework

[1] http://www.baidu.com
[2] http://www.tecent.com

The second phase is the train stage that has two flows after data cleansing. One is traffic data clustering flow, which groups data based on traffic similarity, and then stores each cluster as a pattern into a cluster table so that the number of routes is reduced as the patterns are grouped based on time. The other is classifier training flow that first uses Cluster ID generated by traffic data clustering flow with original trajectory data to construct feature vector for each cluster. For each vector, we design a multiple classifiers system (MCS) to train and store it into a classifier table so that the performance of combining trajectory is improved.

MCS is a set of classifiers whose decisions are combined according to certain rules[Li *et al.*, 2013]. However, MCS may perform worse than individual classifiers without proper design. We adapted the algorithm proposed in [Zhu *et al.*, 2015], which used a weighted voting method to combine the output of each classifier to generate the final decision.

The method considers multiple factors including localized generalization error bound (LGEB), which is the generalization error of a classifier that measures the performance of a classifier generalized to unseen samples. Our key contribution is the calculation of the LGEB, we first find the greatest distance D^M between the sample X to be classified and its K neighborhoods $(Y_1, Y_2, ...Y_K)$ from training samples $D^M = max(d(X, Y_i^K))$. We then calculate the LGEB as $LGEB = \sqrt{\frac{1}{K}\sum_i^K err(f, Y_i)} + \sqrt{\frac{(D^M)^2}{K}\sum_i^K (\frac{\partial f}{\partial Y_i})^T (\frac{\partial f}{\partial Y_i})}$, where $err(f, Y_i) = f(Y_i) - F(Y_i)$ and $\frac{\partial f}{\partial Y_i} = [\frac{df}{dy_{i1}}, \frac{df}{dy_{i2}}, ..., \frac{df}{dy_{in}}]^T$. Here, $f(Y_i)$ is the function for calculating the confidence for each decision to Y_i between 0 to 1, and $F(Y_i)$ is the final decision for Y_i,which is 0 or 1. $\frac{\partial f}{\partial Y_i}$ is the sensitivity term of the classifier, and $(y_{i1}, y_{i2}, ..., y_{in})$ are the features of Y_i.

The third phase is the classification stage, for each request the system queries the classifier table with input feature vector to get the MCS and output the associated Cluster ID, and then uses the Cluster ID to retrieve the associate pattern, and pass to the data pipe for further calculation.

3 Demonstration

During the demonstration, the audience could interact with PARecommender by specifying the place they want to go, and let the system generate a route plan with ETA as shown in Figure 2 based on the traffic pattern. The pattern data used in the demo is derived from the trajectories of Taxis and UGC over six months (more than 10 million GPS points).

Our audience can also turn on the “Traffic” option in the top-right hand corner of the system interface to check the current traffic condition. As shown in Figure 3, the road with green color means the traffic on this road is great while red color means the traffic is bad. PARecommender uses this data to increase the accuracy of route planning and ETA.

Last but not least, PARecommender has many other functions for audience to find out in the demonstration. For example, audience can check the incidents in the route they search, and incidents is one of factors we consider when classifying the traffic patterns.

Acknowledgments

This work was supported by the Natural Science Foundation of Guangdong Province, China(No.2015A030310509), the National Natural Science Foundation of China(61370229,61272067), and the S&T Projects of Guangdong Province(No.2014B010103004).

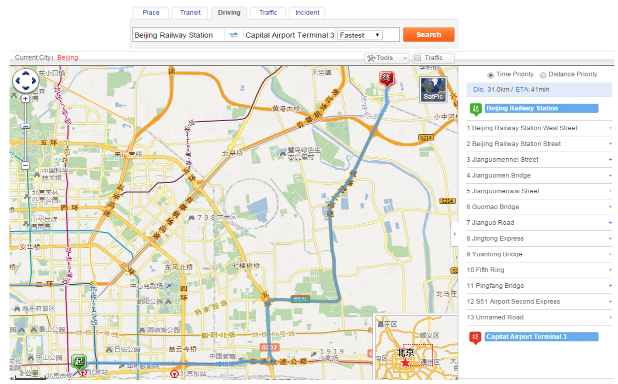

Figure 2: A Route Planning Example of PARecommender

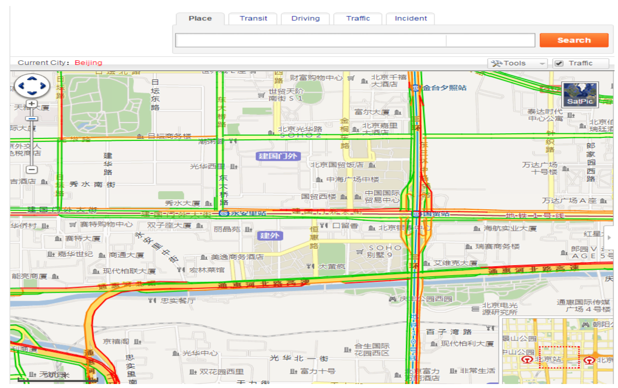

Figure 3: A Traffic Turn-on Example of PARecommender

References

[Li *et al.*, 2013] Yan Li, David M. J. Tax, Robert P. W. Duin, and Marco Loog. The link between multiple-instance learning and learning from only positive and unlabelled examples. *In: Multiple Classifier Systems*, pages 157–166, 2013.

[Luo *et al.*, 2013] Wuman Luo, Haoyu Tan, and Lei Chen. Finding time period-based most frequent path in big trajectory data. *In SIGMOD*, pages 195–203, 2013.

[Monreale *et al.*, 2009] Anna Monreale, Fabio Pinelli, Robert Trasarti, and Fosca Giannotti. Wherenext: a location predictor on trajectory pattern mining. *In SIGKDD*, pages 637–646, 2009.

[Zhu *et al.*, 2015] Jia Zhu, Qing Xie, and Kai Zheng. An improved early detection method of type-2 diabetes mellitus using multiple classifier system. *Information Sciences*, 292:1–14, 2015.