

Reframing Bike Challenge Problem using Model Selection

Chowdhury Farhan Ahmed

University of Strasbourg, France



<http://www.reframe-d2k.org/>

MoReBikeS Challenge (ECML-PKDD 2015)

Chowdhury Farhan Ahmed
University of Strasbourg
France

Introduction

- ▶ It is natural in real-life scenarios that we may have several data in some places and few data in other places.
- ▶ For example, there may be a new bike station beside my house and an old bike station within one kilometer. If I want to predict the available bikes of the new station for the next day, I might not have enough train/previous data. But, I can get an idea from the behavior of the old station nearby.
- ▶ However, its a challenging problem to use the existing knowledge effectively to predict for the behavior of the new station. That means, if I have three nearby old stations described for the above example, knowledge from which station should I use?
- ▶ In this paper, we propose an effective model selection method to solve this problem.

The MoReBikeS Challenge

- ▶ Different models of 200 training stations (numbered as 1 to 200) are given where these models have been trained with data for a long period of time (more than two years).
- ▶ Unlabelled test data for 75 stations (different from the 200 training stations) are given for 3 months (Nov. 2014 to Jan 2015). These test stations are numbered as 201 to 275.
- ▶ For learning the similarities between stations, data of one month (October-2014) are given for all the 275 stations.
- ▶ The challenge is to perform the prediction for the unlabelled test data of 75 stations.

The Proposed Approach

1. Prepare the input and output data in the proper format. Here, we have used ARFF format for using Java and Weka.
2. Select the best training station among the 200 models (short_full_temp) according to their MAE value for a particular test station on the given October-2014 data.
3. For the unlabelled test data (Nov. 2014 to Jan 2015), use the best training model for a particular test station to perform the prediction.

Conclusions

- ▶ To solve the bike challenge problem, we have proposed a simple but effective method.
- ▶ It selects the best similar model according to performance (MAE).
- ▶ Experimental results on the small test data show that it can achieve a good performance.
- ▶ The proposed method can be improved by selecting combination of models rather than the best one.

Thank You