Advances in Knowledge-Based Technologies

Proceedings of the
Master and PhD Seminar
Winter term 2015/16, part 1

Softwarepark Hagenberg
SCCH, Room 0/2
11 November 2015

Software Competence Center Hagenberg
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 800
Fax +43 7236 3343 888
www.scch.at

Fuzzy Logic Laboratorium Linz
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 431
www.flll.jku.at

Program

Chair: Susanne Saminger-Platz

9:00 Reinhard Stumptner:
Adaptive Flood Forecasting for Small Catchment Areas

9:30 Thomas Hoch:
Predict the total travel time of taxi trips based on their initial partial trajectories

Adaptive Flood Forecasting for Small Catchment Areas

Reinhard Stumptner and Bernhard Freudenthaler

Software Competence Center Hagenberg, Data Analysis Systems,
Softwarepark 21, 4232 Hagenberg, Austria
{reinhard.stumptner, bernhard.freudenthaler}@scch.at

1 Introduction and Motivation

There are several sophisticated flood forecasting systems for large rivers like Danube, Rhine or Main. Thereby, hydrological models are built by experts defined for a particular catchment area. But, large rivers generally have a larger warning time than small ones [1], which has to be considered when designing a flood forecasting system.

The water gauge of small rivers can rise rapidly and dams as well as retention basins are sparsely available. Flood forecasting models for such small rivers are normally not available because the development of a flood forecasting system for each small river is too expensive.

Therefore, within the project INDYCO[1] a prototypical flood forecasting system has been developed for small catchment areas. The INDYCO project provides an innovative approach based on dynamic workflows and integration as well as interpretation of sensor data in the frame of disaster management (for more information, see [2]).

2 Flood Forecasting System for Small Catchment Areas

The objective was to develop a model template which can be adapted to small river environments as easy and fast as possible.

First of all, different sensor data has to be rapidly integratable. Therefore, a generic data warehousing structure was developed, into which any available sensor data can be integrated in a short period of time [3]. Such environmental sensor data could be actual and predicted rainfall, water gauge and soil humidity (see Figure 1) for instance. In operation, the system permanently analyses these sensor data and provides warnings to decision makers. A warning could be for instance “A possible 30-year flood threats”. A decision maker can take such warning including background information like weather forecast and geographical overview (GIS) into account and can then make a decision – the warning is valid or not. If a warning is approved by the decision maker, a dynamic workflow system, which provides possible measures, is triggered [4, 5]. The model for analyzing the sensor data is flexibly replaceable.

[1] “Integrated Dynamic Decision Support System Component for Disaster Management Systems”, ERA-NET EraSME program under the Austrian grant agreement No. 836684 (FFG)

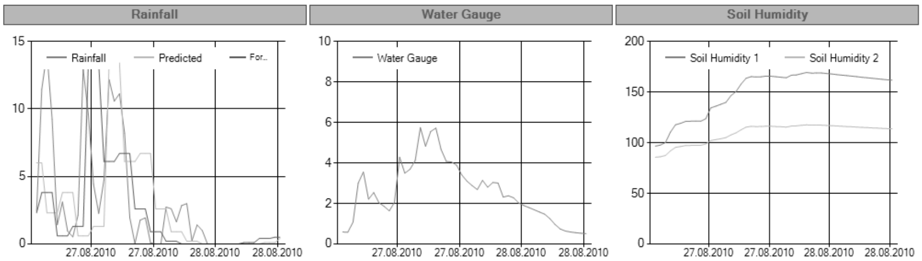

Fig. 1. Environmental Sensors.

3 Experimental Results and Conclusions

The developed models could be trained and verified based on a data set (real data provided by a German city), which contained sensor data (rainfall, water gauge, soil humidity and weather prediction; 5 min. intervals) starting from 1970. After splitting these data into training and test set, the performance of the learned model has the following figures: overall accuracy: 95.21% (best class: 96.71%, worst class: 58.82%). As the model performs quite well in that environment, the next step was to apply it to further environments to receive a more reliable evaluation of the adaptive flood forecasting system.

Acknowledgments. The research leading to these results has received funding from the ERA-NET EraSME program under the Austrian grant agreement No. 836684 (FFG), project "INDYCO - Integrated Dynamic Decision Support System Component for Disaster Management Systems".

References

1. Mahmoodi, O: Hochwasser: Kleine Flüsse gefährlicher als große. online: http://www.pressetext.com/news/20120123011 (2012)
2. Ziebermayr, T., Huber, J., Kollarits, S., Ortner, M.: A Proposal for the Application of Dynamic Workflows in Disaster Management: A Process Model Language Customized for Disaster Management. In: Morvan, F., Tjoa, A.M., Wagner, R. (eds.): DEXA Workshops 2011. IEEE Computer Society, (2011) 284-288
3. Stumptner, R., Freudenthaler, B., Krenn, M.: BIAccelerator – A Template-based Approach for Rapid ETL Development. In: Chen, L., Felfernig, A., Liu, J., Ras, Z.W. (eds.): Foundations of Intelligent Systems, ISMIS 2012. Springer, (2012) 435-444
4. Tropmann-Frick, M., Ziebermayr, T.: Generic Approach for Dynamic Disaster Management System Component. In: DEXA Workshops 2014. IEEE Computer Society (2014)
5. Thalheim, B., Tropmann-Frick, M, Ziebermayr, T.: Application of Generic Workflows for Disaster Management. In: Tokuda, T., Kiyoki, Y., Jaakkola, H., Yoshida, N. (eds.): Information Modelling and Knowledge Bases XXV, EJC 2013. IOS Press, (2013) 64-81

An Ensemble Learning Approach for the Kaggle Taxi Travel Time Prediction Challenge

Thomas Hoch

Software Competence Center Hagenberg GmbH
Softwarepark 21, 4232 Hagenberg, Austria
Tel.: +43-7236-3343-831
thomas.hoch@scch.at

Abstract. This paper describes the winning solution to the Taxi Trip Time Prediction Challenge run by Kaggle.com. The goal of the competition was to build a predictive framework that is able to predict the final destination and the total traveling time of taxi rides based on their (initial) partial trajectories. The available data consists of all taxi trips of 442 taxis running in the city of Porto within one year. The presented solution consists of an ensemble of expert models combined with a spatial clustering approach. The base classifiers consist of Random Forest Regressors where as the expert models for each test trip where based on a combination of gradient boosting and random forest. The paper shows how these models can be combined in order to generate accurate predictions of the remaining traveling time of a taxi.

Keywords: taxi-passenger demand, GPS data, ensemble learning, random forest regressor, gradient boosting, spatial clustering, machine learning

⋆ The research reported in this talk has been partly supported by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry of Science, Research and Economy, and the Province of Upper Austria in the frame of the COMET center SCCH