# The Role of Fielded Applications in Machine Learning Education

Arno Knobbe[1], Arne Koopman[1], Joost Kok[1], Bas Obladen[2], Carlos Bosma[2], and Hessel Galenkamp[2]

[1] LIACS, Leiden University, the Netherlands
knobbe@liacs.nl
[2] Strukton, the Netherlands

## 1 Introduction

In our view, the education of Machine Learning (ML) and associated fields should not just be theoretically and technically sound, but also be motivating and realistic. In other words, courses should provide suitable examples of realistic and complex projects that are representative for the kinds of applications the students can expect after graduating and becoming active as ML or data analysis professional. In most current ML courses, there is a strong emphasis on covering a broad collection of learning techniques (e.g. decision trees, SVMs, etc.) with particular attention to the formal and algorithmic sides of these. Often, the workings of these techniques are demonstrated by means of small toy problems that indeed explain these aspects well. However, by such emphasis on small and often artificial problems, the implicit suggestion may be that all ML problems are of such nature, and that the important questions are concerned with algorithmic issues and the selection of the right learning technique. In the data analysis practice however, one is often faced with complex and large collections of data, and the ML expert needs to deal with other matters such as scale, diversity of data, disagreement about data interpretation and so forth. We believe that including one or more fielded applications of ML in the curriculum, may solve this disparity between theory and practice, by showing the (sometimes surprising) challenges of such applications. Below we give an example of such a project that deals with the analysis of high-resolution sensor-data that is being collected at a large highway bridge in the Netherlands.

Apart from being realistic, and thus a valuable addition to the course material, real and complex applications may also play a role in motivating current and potential future students. One often observes a specific preference for applied research among students and researchers early in their academic career, which is later followed by an interest in more fundamental issues. As such, the addition of said projects may help to win the students for ML by livening up the course material.

**Fig. 1.** (left) Aerial picture of the situation of the Hollandse Brug, which connects the 'island' Flevoland to the province Noord-Holland, and the adjacent railway bridge (top). (right) Some of the sensors attached to the underside of the bridge.

## 2 The Hollandse Brug

The InfraWatch project is centred around an important Dutch highway bridge that is already producing substantial quantities of data: the Hollandse Brug. This bridge is located between the Flevoland and Noord-Holland provinces, at the place where the Gooimeer joins the IJmeer (see Figure 1 on the left). It was opened in June 1969, and in April 2007 it was announced that measurements would have shown that the bridge did not meet the quality and security requirements. Repairs were launched in August 2007 and a consortium of companies has installed a monitoring configuration underneath the Hollandse Brug with the main aim to collect data for evaluating how the bridge responds. The sensor network is part of the strengthening project which was necessary to upgrade the bridges capacity by overlaying.

The monitoring system comprises 145 sensors that measure different aspects of the condition of the bridge, at several locations along the bridge (see Figure 1 on the right). The following types of sensors are employed:

- 'geo-phones' (vibration sensors) that measure the vertical movement of the bottom of the road-deck as well as the supporting columns.
- strain-gauges embedded in the concrete and attached to the outside, measuring horizontal stress in two directions.
- thermometers embedded in the concrete and attached to the outside.

Furthermore, there is a weather station, and a video-camera that provides a continuous video stream of the actual traffic on the bridge. Additionally, there are plans to monitor the adjacent railway bridge.

Prior to the start of the InfraWatch project, an initial monitoring application was developed by a team of students, that allows the visual inspection of

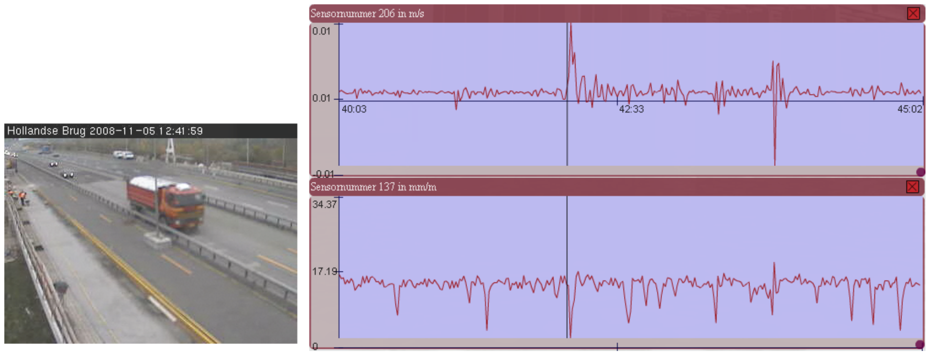


**Fig. 2.** Example of a truck passing the single camera located on the bridge. The graphs show the signal of two sensors, with a vertical bar indicating the time that corresponds to the shown video frame.

both video and sensor information. The application allows the user to navigate through a selected time-frame, and display the traffic passing over the bridge, while the data over one or more sensors is displayed in synchronised fashion 2. The user can select the nature of the sensor as well as the location of it, which does not necessarily have to correspond with the location of the camera. Using this application, it is fairly easy to already observe some patterns in the data. For example, the vertical load data nicely corresponds with heavy vehicles passing. However, more sophisticated data analysis should be developed in the course of the project, that also takes into account multivariate behaviour of the data, and spatial relationships between sensors, to name just a few options.

## 3 Education Opportunities

Besides being an excellent research challenge and a complex fielded application of Data Mining techniques, the InfraWatch project and its Hollandse Brug are also intended to serve educational purposes [6]. Because of its practical nature, the project will, and has already been an important tool in the teaching of intelligent data analysis techniques to computer science students in the first place. Rather than the traditional focus on basic analysis techniques and algorithms, we now have an opportunity to demonstrate the many complications that tend to arise in actual analysis projects [3, 4], and how these should be tackled. These complications include the measuring of data (noise, sensor-failure, ...), the continuous flow of data (data volume, versioning issues, sample rates), the range of analysis paradigms (multivariate analysis, streams, relational aspects), and the inclusion of domain knowledge (spatial aspects, feature extraction). Apart from making the existing data analysis education more attractive and realistic, the project also serves to attract potential students to analysis-related courses and computer science in general. Currently, at least two MSc students are involved with the InfraWatch project.

## Acknowledgements

The InfraWatch project is funded by the Dutch funding agency STW, under project number 10970.

## References

1. M. Dejori, H.H. Malik, F. Moerchen, N.C. Tas, and C. Neubauer, 2009 *Development of Data Infrastructure for the Long Term Bridge Performance Program*, In Proceedings of Structures '09, Austin, USA.
2. E. Doupal, R. Calderara, 2004, *Weigh-In-Motion*, In Proceedings of First International Conference on Virtual and Remote Weigh Stations, Orlando.
3. A. Knobbe, 1997, *Data Mining for Adaptive System Management*, In Proceedings of PAKDD '97, London.
4. A. Knobbe, Bart Marseille, Otto Moerbeek, Daniël M.G. van der Wallen, *Results in Adaptive System Management*, Benelearn'98
5. G. Meijer, *Smart Sensor Systems*, 2008, ISBN: 978-0-470-86691-7, Hardcover, 404 pages.
6. R. Gavaldà, 2008, *Machine Learning in Secondary Education?*, In Proceedings TML 2008, Saint Etienne, France, http://www.lsi.upc.edu/~gavalda/docencia/tml08-revised.pdf