Predicting functional outcome after acute ischemic stroke: A comparison of learning machines

Andreas Ziegler[1], Inke R. König[1], Sinisa Pajevic[2], Christian Weimar[3], Hans-Christoph Diener[3] & James D. Malley[2], on behalf of the German Stroke Study Collaborators

[1] Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany

[2] Center for Information Technology, National Institutes of Health, Bethesda (MD), USA

3 Klinik und Poliklinik für Neurologie, Universität Essen, Germany

Andreas Ziegler
Institut für Medizinische Biometrie und Statistik
Universitätsklinikum-Schleswig-Holstein, Campus Lübeck
Universität zu Lübeck
Ratzeburger Allee 160, Haus 4
23538 Lübeck
Germany
Email: ziegler@imbs.uni-luebeck.de

This study aimed to develop and externally validate two models for predicting functional outcome and 100 day survival after acute ischemic stroke. A systematic literature search was performed to identify all possible risk factors. Only variables that could be measured within the first 72 hours of admission were selected to allow for the models to be used for prediction at this time. The models were first estimated and internally validated using data from the German Stroke Foundation's stroke data bank and the classical logistic regression model (Weimar et al. 2002 J Neurol 249: 888-95). The resulting models were based on 1,768 patients who had been prospectively collected in seven neurology departments between 1998 and 1999. The first model predicted incomplete functional recovery (Barthel Index < 95) versus complete functional recovery with eleven variables, while the second model predicted mortality versus survival with three variables.

Finally, data on 1,802 patients were used to externally validate the two models to prospectively from February 2001 to March 2002 (The German Stroke Study

Collaboration 2003 Neurol, in press; Weimar et al. 2003 Stroke, Epub ahead of print). On admission to a participating hospital, patients were registered prospectively and selected into the study based on defined inclusion/exclusion criteria (for details of the study protocol see König et al. 2003 Z Ärztl Fortbild Qualitätssich 97:717-722). At 72 hours the predictive variables were collected, the models were run and predictions of the respective outcomes were made. Follow up was performed 100 days after the original event. Based on the models developed in the initial sample, the first correctly predicted 73.5% of the patients from the validation study who were incompletely recovered or had died and 83.0% of the completely recovered patients. The second model predicted 35.1% of the patients who had died and 97.9% of the surviving patients. Both models performed better than the treating physicians´ predictions made within 72 hours after admission.

Meanwhile, we applied several methods that have been developed over the last years by the machine learning community for classification and regression. In the talk, we shall restrict ourselves to support vector machines and boosting stumps. We sketch the main ideas of these learning machines. Specifically, we observe that there are important differences in how these machines might be applied to such data, quite distinct from how they have been used in the computer sciences. These differences include much smaller sample sizes and a focus on prediction error estimates. Other differences include unbalanced data and questions of interpretability of the proposed prediction scheme. We compare the results from the different learning machines and discuss similarities and differences.