Estimation of individual prediction reliability using sensitivity analysis of regression models

doctoral dissertation (extended abstract)

Zoran Bosnić

University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, Ljubljana, Slovenia
zoran.bosnic@fri.uni-lj.si
http://lkm.fri.uni-lj.si/zoranb/dissertation.htm

1 Introduction

The dissertation [1–3] discusses the reliability estimation of individual regression predictions in the field of supervised learning. In contrast to the average measures for the evaluation of model accuracy (e.g. mean squared error), the reliability estimates for individual predictions can provide additional information which is beneficial for evaluating the usefulness of the predictions. This additional information can also provide decision support to the users of the prediction systems, based on which they can decide on the corresponding consequential actions (prescribe a therapy, use the autopilot, etc.).

Measuring the expected prediction error is especially important in the risk-sensitive areas where acting upon predictions may have financial or medical consequences (e.g. medical diagnosis, stock market, navigation, control applications). In such areas, appropriate local accuracy measures may provide additional necessary information about the prediction confidence. The difference between traditional approach to the model evaluation and the reliability estimation is illustrated in Figure 1. The figure also shows an additional advantage of the reliability estimates for individual predictions: They are computed for each particular example in contrast to averaged model estimates which require a separate set of test examples.

2 Categorization of the Related Approaches

In our work, we use term reliability estimate to denote any quantity that estimates a quality of a regression prediction. In the related work, the reliability estimates have therefore appeared as either accuracy estimates or error estimates.

Depending on how the reliability estimates are implemented, the dissertation divides them into two groups: (i) the model-dependent estimates which exploit
Fig. 1. Reliability estimate for the whole regression model (above) in contrast to reliability estimates for individual predictions (below).

the properties of a particular models (e.g. the number of support vectors [4], Lagrange multipliers in the SVM optimization procedure [5, 6], splits in a regression tree, etc.) and (ii) the model-independent reliability estimates, which exploit the general properties of the supervised learning framework (e.g. changing the learning set, etc.).

Besides providing an overview of the related work for the both of the above directions, the dissertation also summarizes and defines various terms which are used in this field (e.g. reliability, sensitivity, stability, confidence, credibility, etc.). The definitions of the terms are systematically shown as a dictionary, which represents an unification of the terminology in the field.

3 Model-Independent Reliability Estimates

The main part of the dissertation focuses on developing and comparing new approaches from the group of model-independent reliability estimates. The dissertation proposes 9 new such reliability estimates [7]. Three of newly proposed estimates are developed by adapting the sensitivity analysis [8] approach for the use in the supervised learning. To apply the principles of the sensitivity analysis, we propose a framework for controlled modification of the input (learning set) and outputs (regression predictions) in the supervised learning setting. By applying minor modifications to the learning set, we exploit the instabilities in predicted values and use them to compose reliability estimates.

Five estimates are either adapted from related work based on the following approaches: analyzing variance of bagging models, local cross-validation, density estimation, and local error estimation. In the dissertation, the existing estimates are generalized to be used with other regression models. The ninth reliability estimate is proposed as the linear combination of the two individual estimates among all former.

In the experimental part, the dissertation presents an empirical evaluation of the above reliability estimates using 8 regression models (regression trees, linear regression, neural networks, bagging, support vector regression, locally weighted
regression, random forests, generalized additive model) and 28 standard benchmark test domains. The performance of the reliability estimates is measured with their correlation coefficient to the prediction error of the individual examples. The correlation coefficients are statistically evaluated to confirm whether the reliability estimates significantly estimate the prediction errors. The testing results showed usefulness of the proposed reliability estimates especially for the use with regression trees, where one of the proposed estimates correlated with the prediction error in 86% of the testing domains. On the average (across all used regression models) the estimate which is based on the bagging variance analysis and the linearly combined estimate achieved the highest performance (correlation to the prediction error).

4 Automatic Selection of the Most Appropriate Estimate

The testing results of the individual reliability estimates revealed that the estimates achieved different performance on different problem domains and using different regression models. Accordingly, in the dissertation we study the problem of the most appropriate reliability estimate selection for a given problem domain and the regression model [9]. We discuss and define two possible solutions of this problem, based on meta-learning and internal cross-validation approach.

In the context of the proposed meta-learning approach we define a meta-problem space for prediction of the best performing reliability estimate. The dissertation presents a possible attribute description of the meta-learning problem and defines it as a classification problem, where each class represents one of the 9 proposed reliability estimates. Using a collection of our 28 testing domains and in combination with 8 different regression models, we construct a meta-learning training set consisting of 224 (28 × 8) examples. We use this training set to construct a decision tree meta-classifier, which is afterward used to predict the most appropriate reliability estimates for testing domains which do not comprise the meta-training set.

Since decision tree is an interpretable model, we use the constructed meta-classifier to analyze in which cases each particular estimate perform better. The analysis results indicate that the estimates achieve better performance when used with more accurate models (models that achieve lower relative mean squared error on the testing set).

The second approach to the best performing estimate selection is based on the internal cross-validation approach. It is designed to iteratively measure the performance of the reliability estimates on different subsets of the testing domain. The best performing estimate on the average is afterwards used to estimate the reliability of the test examples which were excluded from the estimate selection process.

The testing results have shown that the dynamically selected reliability estimate using both of the above approaches outperforms any of the individual reliability estimates.
5 Implementation in a Medical Domain

The individual estimates and the both approaches for automatic selection of the optimal estimate were tested on a real domain from the area of medical prognostics. The data consisted of 1035 breast cancer patients, who had surgical treatment for cancer between 1983 and 1987 in the Clinical Center in Ljubljana, Slovenia. The goal of the research was to predict the time of possible cancer recurrence after the surgical treatment.

The analysis showed that this is a difficult prediction problem, because the possibility for recurrence is continuously present for almost 20 years after the treatment. Furthermore, the data presents a mixture of two prediction problems, which additionally hinders the learning performance: (i) yes/no classification problem, whether the illness will recur at all, and (ii) the regression problem for the prediction of the recurrence time.

In our study, the bare recurrence predictions were complemented with our reliability estimates. To implement the prediction system, the locally weighted regression was selected for the use with this problem due to its low RMSE (compared to the RMSE of the other models). The model was complemented with one of our reliability estimates which was unanimously selected by both of our approaches for the selection of the best performing estimate. A graphical representation of such predicted time of the cancer recurrence, equipped with reliability information is shown in Figure 2. The implemented prediction system helped the doctors with the additional validation of the predictions’ accuracies.

The statistical comparison of reliability estimates to prediction evaluations of the medical experts showed that our reliability estimates correlate to prediction
error with statistically equal correlation as the manual evaluations of the experts do. This results therefore showed the potential of the proposed methodology in practice.

6 Conclusion

Implementation of reliability estimates for the individual predictions can be a helpful tool when using critical decision support systems. In the dissertation, several such reliability estimates are evaluated and proposed. Additionally, two approaches for automatical estimate selection, which increase their consideration for practical use, are proposed and evaluated, as well.

The successful implementation of the proposed methodology in a medical domain indicates the importance and the potential for the use of the reliability estimation in practice. To conclude, the dissertation provides the theoretical time complexities for the computation of the estimates’ values. The ideas for the directions of the further work include work on the interpretability of the estimates’ values, analysis of the mathematical estimates’ properties and best performing reliability estimate selection for an individual example to be predicted.

References