OPTIMIZATION, VALIDATION AND APPLICABILITY OF ONE-CLASS CLASSIFICATION METHODS

Zuzanna Małyjurek

Class-modelling, known also as one-class classification, is a classification tool widely applied for authentication, quality control, origin verification, or novelty detection. Class-modelling is used for individual class-model construction based on the similarities among samples of the class studied, the target class. The model obtained is used to predict whether a new sample of unknown origin belongs to the class modelled. The class-model construction is a multistep process that includes, e.g., the construction of the training and test sets, selection of the class-modelling method, optimization of the model, and its validation. Each step has an impact on the final classification results thus they should be handled carefully.

The aim of the study presented in this thesis was to indicate the optimal strategy for class-model optimization, authentication of similar classes, limitations and scopes of applicability of selected class-modelling methods.

In the case of optimization of the class-model, i.e., the selection of the model complexity and classification rules, two scenarios can be considered: compliant and rigorous. In the compliant scenario model is optimized based on the target and nontarget class samples, whereas in the rigorous scenario only the target class samples are used. The influence of these optimization strategies on classification results was tested upon the example of SIMCA model construction (Soft Independent Modelling of Class Analogy). It was shown that the best results were obtained with class-models optimized in the compliant scenario. However, these models can be biased, thus the rigorous scenario is better suited for model optimization.

Once the optimal optimization strategy was selected, different class-modelling methods were tested considering their scope of applicability. The Support Vector Description Domain models led to the highest classification results most often, but the density-based class-modelling methods, such as Potential Functions Method, are more appropriate once data of complex structure are analysed.

Individual class-models can lead to unsatisfactory results when several similar classes are authenticated. In such situations, the discriminant model exhibit better performance than class-models, but classical discrimination cannot be applied for authentication. Thus, a two-step approach that combines class- and discriminant modelling was proposed, and its performance was compared with so-called soft discriminant methods.

The conclusions obtained from the aforementioned aspects studied allowed to propose a successful strategies for authentication of rooibos, honeybush, and three *Cyclopia* species used for honeybush production.