



To whom it may concern

SUBJECT: Report on the PhD thesis of Zuzanna Matyjurek

The thesis titled “Optimization, validation and applicability of one-class classification methods”, as the title clearly summarizes, focuses on evaluating the optimal strategy for building and optimizing one-class classifiers, also addressing issues such as how to proceed when there are classes with very similar characteristics, and on identifying the possible limitations and ranges of applicability of the various class-modeling techniques.

The thesis itself is organized as a collection of papers published on high quality peer-reviewed journals, all well reputed within the chemometrics community, where the candidate is first author and her fundamental contribution is easily enucleable. Indeed, also the last paper, which was, at the time of the submission of the Thesis, still under review, has in the meanwhile been accepted and published.

Going in detail into the scientific content of the work, it is possible at first to affirm that, as a general consideration, the research work presented in the PhD thesis is highly timely. Indeed, there are different domains (e.g., food authenticity, process analysis, metabolomics or omic sciences in general, and forensics, just to cite a few) where the need of suitable class-modeling tools is becoming more and more stringent, since the classification problems they involve are highly asymmetric and sometimes only the class of interest can be sampled with sufficient representativeness. At the same time, class-modeling tools haven't developed as extensively as discriminant ones and quite often attention is focused mostly on the SIMCA algorithm, which remains the most commonly used technique in this context. In this respect, the critical comparative work presented in the thesis and, in particular, in one of the Chapters (“The scope of applicability of the selected class-modelling methods”) is of high value, since several datasets of various structures are analysed with different class-modelling methods to test their scope of applicability and the pros and cons of the techniques from the practical point of view. Such guidelines can be a valid help for people approaching the use of one-class classifiers in deciding which tool to use, and also provide clear indication on which aspect of the different techniques could represent weak points that could be improved.

Similarly, when considering the already mentioned SIMCA algorithm, the candidate has examined and compared some of the most commonly used implementations of the

technique and the two alternative approaches for model selection and optimization, i.e., rigorous vs compliant. Also in this case, the possible strategies for model building and model selection have been compared on a solid basis, by means of real and simulated data set with specific characteristics, so to be able to draw some general conclusions about what the best way of proceeding could be.

Another important contribution that stemmed from the research work presented in the Thesis is the definition of a strategy which couples discriminant and modeling approaches for authentication problems which involve classes with very similar characteristics. In the suggested strategy, the strong point of discriminant and modeling techniques are exploited to define a hierarchical data processing pipeline where at first class-modeling allows to identifying a sample as belonging to any of the similar categories or not (thus, allowing to highlight possible outliers or samples from other origins) and successively discriminant analysis on the filtered data provides the classification into one of the restricted possible categories. As demonstrated by the practical application reported in the Thesis, such a strategy is extremely useful in the cases of food authentication, where many categories can share similar features with one another: notwithstanding this, the proposed approach provided a very high classification accuracy.

Altogether, the research work is well placed in the context of the existing literature, which the candidate has demonstrated to know extensively: indeed, the introduction is quite exhaustive and describes critically the different possible strategies for class-modeling, their main characteristics and their strenght points and weaknesses.

Furthermore, it is worth noticing that the candidate, in addition to oral and poster contributions in national context, has been presenting authors of oral communication in five international conferences and for one of those (17th Scandinavian Symposium in Chemometrics) has also received the prize for the best oral talk.

All the considerations reported above clearly indicate that the research work presented in the Thesis constitutes a relevant contribution to chemometrics in general and, in particular, to the field of classification. The work was conducted with high scientific rigor, the results have been presented clearly and the Thesis itself is well-organized and easy to read. Moreover, as already discussed above, the maturity with which the pros and cons of each technique are illustrated and discussed provide an added-value to the reported research.

Altogether, the quality of the work, the wealth of results obtained and the Thesis itself place the candidate well above the average level of a PhD student.



To conclude, I can affirm that the Thesis meets the highest quality standards for a graduate research and therefore it is fully adequate to be presented in the PhD defense.

Yours truly,

Prof. Federico Marini

A handwritten signature in black ink that reads 'Federico Marini'.

Rome, September 11th 2022