Review of the PhD thesis

Imperfect Learning of Multi-Classifier

by Mgr. Ing. Radek Hřebík

The presented dissertation is devoted to a new so-called imperfect approach to classification problems. As outlined in Chapter 2, the main goals are to propose the construction of a novel classifier, based on a principle of the critical sensitivity, to create the concept of the hidden classes, to propose an imperfect learning stratégy, and to develop a new technique for the data selforganization that could serve as an alternative to the traditional Kohonen learning.

From this point of view, the topic of the dissertation is relevant in the fields of data analysis, classification, machine learning, data structure detection.

The dissertation consists of an Introduction, 7 chapters, Conclusions, Bibliography with 80 references, Appendix A with source code examples and Appendix B with a list of 3 author-selected publications in impact journals relevant to the dissertation, along with their copies.

Background information relevant to the entire research program, including the formulation of the classification problem, leading or most commonly used methods, preprocessing methods, complexity problem, and optimization criteria, is given in Chapter 1.

The motivation of the author and the goals of the study are given in Chapter 2. The main goal is to create a new classifier. The rest of the goals are connected and subordinate to the main one.

Chapter 3 discusses the novel approach to learning algorithms of selforganising maps. The aim is to benefit from a natural process of diffusion and present an alternative approach to the traditional Kohonen learning.

The general theory of the imperfect learning are discussed in Chapter 4. This seems to be a central chapter of the whole thesis where basic methods of pattern sub-sampling, feature sub-samplings as well as the use of imperfect learning algorithms are summarized.

In Chapter 5, the results of all the previous theoretical parts are summarized and the classifier description is proposed.

Finally, the test results are presented in Chapter 6, where the known reference datasets are tested first, and then the proposed classifier processes the so-called real datasets.

Chapter 7 presents a brief summary of used Matlab functions together with the description of the related classification method. The last chapter contains the conclusions obtained in the dissertation and formulates some new research problems.

Assessment of the methods used in the dissertation

Informally, the dissertation is a free text that discusses various heuristic approaches to the problem of classification in feature spaces. In view of the vastness of the field of study and the formulated goals in the form that is complementary to the existing approaches named as "standard perfect algorithms", the author is forced to give a short overview of many methods used in this area. After that, he concluded (p. 47) that " we have no guarantee to obtain the requested properties, e.g. perfectness, accuracy, or sensitivity when the classifier learning process is driven by standard perfect algorithms". As a result, research efforts have been shifted towards the new goal indicated as „imperfect learning" which manifests itself in an improper way as „we do not insist on the algorithm perfectness", but „to work further with all the different results of imperfect classifiers providing different results on the same pattern set" (p. 47). This program is proposed to be realized in three ways: (i) selecting only a limited number of patterns and not the whole pattern set, (ii) focusing only on some features, which are selected randomly, (iii) using any known imperfect method which guarantees to find only a local extreme " (p. 48). However, in section 4.1.3. under the title "Imperfect Learning Algorithm" we find not an algorithm, but very general recommendations with the conclusion that repeated imperfect learning produces results that lead to a classifier with better properties. I consider this claim to be *non-supported*.

In the next chapter, the proposed structure of the classifier is shown in Fig. 5.1 with comments. From this image, we can see that before applying the set of imperfect classifiers, the data must be pre-processed with one of two (or both) transformations. This recommendation is *contrary* to the formulation of the classification problem as formulated in Section 1.1, where classification is applied to feature vectors. The latter differ from the objects in the raw dataset in that they are pre-processed. In addition, the role of the nonlinear transformation is *not sufficiently explained* and this type of preprocessing has not been demonstrated using the considered or proposed examples.

As a commentary on the diffusion learning proposed in Chapter 3, I can only say that this undeniably promising direction was only touched upon briefly in the dissertation. In the proposed text, diffusion processes are discussed solely in the language of partial differential equations. Except for general arguments of importance and relevance, I did not find the author's vision of how diffusion's effect, especially anomalous diffusion, can be used in the relevant classification learning.

The experimental part, which should support the author's discoveries, was finally reduced to fixed settings: $n^*=3$ and $m^*=100$ and only one RAO preprocessing method. This fact significantly reduces the impact of the proposed teaching method, creating the impression that the extended theoretical part is unnecessary.

Evaluation of the results and contributions of the dissertation

Due to the vast scope of the study and the generally stated objectives of the study, it is difficult to unambiguously assess the results achieved and the specific contribution. Unfortunately for the author, there was a lot of competition in the development and promotion of the chosen field. Nevertheless, I would like to support the proposed research (albeit incomplete and in many ways imperfect) as a courageous attempt to understand a complex area and show a new direction for development.

It is also worth noting that the author's study and the results obtained are consistent with many others obtained independently by other researchers. In addition, he has gained experience in both model-driven and data-driven approaches. His recommendations, even with narrow support, can be used by followers.

Questions for the defense

1. As stated in Section 1.4, the proposed data whitening characterization (PCA) in Section 1.4.1 is similar to the MDA method in Section 1.4.2. The formal syntactic difference lies in the weight matrices W_1 or W_3. What is the semantical difference between these matrices and, therefore, between these methods?

2. What is the difference between the proposed imperfect learning and imperfect classifier?

3. Why was diffusion learning not used in the experiments undertaken?

4. What type of learning: supervised or unsupervised does the proposed theory of parallel imperfect learning (Chapter 4) refer to?

5. What is the complexity estimate of the proposed imperfect classifier?

6. In the Experimental part (Chapter 6) you worked with the fixed settings: $n^*=3$ and $m^*=100$. Then I see that the only one preprocessing method RAO was applied. How you can explain this restricted setting?

Remarks

The language of the thesis is not precise, there are many grammatical errors. Sometimes I get the impression that the original Czech version was translated into English without carefully reading the result.

The terminology doesn't help. An imperfect classification has no definition and therefore cannot be assessed based on strict criteria. The abbreviations used for the methods differ from the generally accepted ones (PCA versus DWH, RAO versus MDA), which makes it difficult to understand what the author is writing about.

Despite the aforementioned substantive and formal shortcomings, the presented dissertation shows that Mgr. Ing. Radek Hřebík is able to perform creative and systematic research. In accordance with par. 47, letter (4) of the Law Nr. 111 /1998 (The Higher Education Act) I do recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

V Ostravě dne 15.06.2023

Irina Perfiljeva

Ostravská Universita,

ÚVAFM