Encore Abstract: Presumably Correct Decision Sets

Gonzalo Nápoles[1], Isel Grau[2,3], Agnieszka Jastrzebska[4], and Yamisleydi Salgueiro[5]

[1] Department of Cognitive Science and Artificial Intelligence, Tilburg University, The Netherlands
[2] Information Systems, Eindhoven University of Technology, The Netherlands
[3] Eindhoven Artificial Intelligence Systems Institute, Eindhoven University of Technology, The Netherlands
[4] Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland
[5] Department of Industrial Engineering, Universidad de Talca, Campus Curicó, Chile
g.r.napoles@uvt.nl

Abstract. The paper presents the presumably correct decision sets as a tool to analyze uncertainty in the form of inconsistency in decision systems. As a first step, problem instances are gathered into three regions containing weak members, borderline members, and strong members. This is accomplished by using the membership degrees of instances to their neighborhoods while neglecting their actual labels. As a second step, we derive the presumably correct and incorrect sets by contrasting the decision classes determined by a neighborhood function with the actual decision classes. We extract these sets from either the regions containing strong members or the whole universe, which defines the strict and relaxed versions of our theoretical formalism. These sets allow isolating the instances difficult to handle by machine learning algorithms as they are responsible for inconsistent patterns. The simulations using synthetic and real-world datasets illustrate the advantages of our model compared to rough sets, which is deemed a solid state-of-the-art approach to cope with inconsistency. In particular, it is shown that we can increase the accuracy of selected classifiers up to 36% by weighting the presumably correct and incorrect instances during the training process.

Keywords: Data analysis · Granular computing · Decision sets · Rough sets.

This document is an encore abstract of the paper "Presumably correct decision sets" published at the Pattern Recognition journal [2].

1 Motivation and Results

Rough sets [3] generalize the view of classic sets using two concepts: the upper and lower approximations, which are instantiated with the help of crisp sets. The lower approximation contains instances certainly in a given set, while the upper approximation is made of instances possibly in that target set. The difference between an upper and lower approximation defines the boundary region. This formalism allows us to analyze uncertainty in the form of inconsistency. The mathematical formulation is straightforward, and the resulting models have a high degree of interpretability [1], which is of paramount importance in decision-making.

Despite the theoretical developments concerning rough sets, several sensible drawbacks remain unresolved. Two of these issues include the model's tendency to produce trivial similarity classes and the effect of outliers on the positive regions. In the first case, the similarity class of an object only contains the object itself, thus leading to empty boundary regions that do not add value to an inconsistency analysis. In the second case, the presence of outliers often causes the entire similarity classes to be placed in the boundary region. As a result, we cannot determine which boundary instances are outliers or which could have been placed in a positive region if they were not affected by the outliers.

This paper introduces the presumably correct decision sets (PCDS) to address the abovementioned drawbacks. Our model uses a neighborhood function to gather instances into three disjoint regions containing weak, borderline, and strong members. There will be three of these regions for each decision class. Instances are allocated to those regions according to their membership degree to their neighborhood (referred to as similarity class) and nothing else. Subsequently, we derive the presumably correct and incorrect decision sets from either the strong regions or the whole universe of discourse. We say that an instance is presumably correct if its decision class agrees with the most frequent class observed in its similarity class. Otherwise, we say that the instance is presumably incorrect. This simple yet highly effective approach tackles two important issues of rough sets. On the one hand, we can limit the effect of noise by increasing the number of neighbors defining the similarity class. More importantly, incorrect instances are isolated and do not affect the correct ones. On the other hand, the model will not produce trivial similarity classes, as the smallest similarity class will be composed of the instance and its closest neighbor.

The experiments reported in the paper demonstrate that when the proposed procedure is integrated into the data pre-processing pipeline, it can largely improve the prediction rates. It was also experimentally observed that the PCDS-based weighting strategy leads to larger performance increases than those obtained by weighting the instances located at the fuzzy-rough set boundary regions. The presumably incorrect regions provide a valuable analysis tool in situations where inconsistency might lead to biased decisions. This happens since our model allows extracting meaningful sets for each decision class, indicating which instances received unfair treatment.

References

1. Ji, W., Pang, Y., Jia, X., Wang, Z., Hou, F., Song, B., Liu, M., Wang, R.: Fuzzy rough sets and fuzzy rough neural networks for feature selection: A review. Wi-

ley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **11**(3), e1402 (2021)
2. Nápoles, G., Grau, I., Jastrzebska, A., Salgueiro, Y.: Presumably correct decision sets. Pattern Recognition **141**, 109640 (2023)
3. Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W.: Rough sets. Communications of the ACM **38**(11), 88–95 (1995)