<u>Title</u>

Fuzzy-Rough set hybridation and its applications in Business Informatics

<u>Introduction</u>

The resemblance between the terms “fuzzy sets” and “rough sets” tends to create confusion. But, while fuzzy set theory characterizes a concept through a membership function ranging between zero and one, rough set theory describes the concept by means of relation-dependant lower and upper approximations. They are not rival theories but two different mathematical tools aimed for two different purposes. They can be combined in a fruitful way by defining rough-fuzzy sets and fuzzy-rough sets. The concept of fuzzy-rough set was proposed by replacing crisp binary relations with fuzzy relations over the universe.

The main components of the Rough Set Theory (RST) are a decision system and an Indiscernibility relation. A Decision System is a pair DS = (U, A $\cup$ {d}), where U is a non-empty finite set of objects called the Universe, A is a non-empty finite set of features, and d$\notin$A is the decision feature. The basic concepts of RST are the lower and upper approximations of a subset X $\subseteq$ U. These were originally introduced with reference to an indiscernibility relation IND (B), where objects x and y belong to IND (B) if and only if x and y are indiscernible from each other by features in B.

Let B $\subseteq$ A and X $\subseteq$ U. It can be proved that B defines an equivalence relation. The set X can be approximated using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted by $\underline{B}X$ and $\overline{B}X$ respectively, where
$\underline{B}X = \{x \mid [x]_B \subseteq X\}$, and $\overline{B}X = \{x \mid [x]_B \cap X \neq \phi\}$. and $[x]_B$ denotes the class of x according to B-indiscernible relation

The objects in $\overline{B}X$ are sure members of X, while the objects in $\underline{B}X$ are possible members of X. If $BN_B(X) = \overline{B}X - \underline{B}X$, is not empty, then X is a rough set.

In order to use the subset B $\subseteq$ A to define an equivalence relation, the features in B must be discrete, that is, the domain of the features are a set of finite elements. As in real applications many continuous features can exist, this forces to discrete these continuous features. The results of the discretization process depend on the discretization method, being usually the results very sensitive to the method.

The fuzzy sets can be used to carry out the discretization process. In this case, each continuous feature should become in a linguistic variable, and the values of the features are substituted in the decision system by the linguistic terms. For example, if we have the feature Age, with values between 0 and 150, this it can become into the linguistic variable Age with the linguistic terms {baby, young, adult, old}, in this case the value 17 maybe could be substituted using the maximum membership principle by Young/0.92, where the value 0.92 indicates the membership grade to the fuzzy set Young.

Once the decision system has been transformed, the indiscernibility relations can be defined based on the linguistic terms. Given these indiscernibility relations, the lower and upper approximations are built. The effects of this approach could be studied by means of the following tasks:

Research task

1. Define indiscernibility relations and study their mathematics properties (they are equivalence relations or tolerance relations?). To consider both the operators of the classic fuzzy logic and the operators of the compensatory fuzzy logic.
2. Study of the performance of this approach in some machine learning method based on rough sets, such as editing training sets, discovering causal rules, and Discovering association rules, etc. In order to evaluate the performance, international data sets, such UCI Repository, could be used.
3. Develop an application of the methods resulting in task 2 in a real application of Business informatics. The new methods must be compared with the original ones.