Fuzzy Rough Sets: The Forgotten Step

Martine De Cock, Chris Cornelis, and Etienne E. Kerre

***Abstract*—Traditional rough set theory uses equivalence relations to compute lower and upper approximations of sets. The corresponding equivalence classes either coincide or are disjoint. This behaviour is lost when moving on to a fuzzy T-equivalence relation. However, none of the existing studies on fuzzy rough set theory tries to exploit the fact that an element can belong to some degree to several "soft similarity classes" at the same time. In this paper we show that taking this truly fuzzy characteristic into account may lead to new and interesting definitions of lower and upper approximations. We explore two of them in detail and we investigate under which conditions they differ from the commonly used definitions. Finally we show the possible practical relevance of the newly introduced approximations for query refinement.**

***Index Terms*—Fuzzy rough set, lower and upper approximation, query refinement, transitivity.**

I. Introduction

Since its introduction in the 1960s, fuzzy set theory has had a significant impact on the way we represent and compute with vague information. More recently it has become part of the larger paradigm of soft computing, a collection of techniques that are tolerant of typical characteristics of imperfect data and knowledge—such as vagueness, imprecision, uncertainty, and partial truth—and hence adhere closer to the human mind than conventional hard computing techniques. During the last decades new approaches have been developed that generalize the original fuzzy set theory (which is also called type-1 fuzzy set theory in this context). Type-2 fuzzy sets, intuitionistic fuzzy sets, interval-valued fuzzy sets and fuzzy rough sets have in common that they can all be formally characterized by membership functions taking values in a partially ordered set P, which is no longer the same (but an extension of) the set of membership degrees $[0,1]$ used in fuzzy set theory. The introduction of such new, generalizing theories is often accompanied by lengthy discussions on issues such as the choice of terminology and the added value of the generalization.

In this paper we focus on fuzzy rough set theory. Pawlak [23] launched rough set theory as a framework for the construction of approximations of concepts when only incomplete information is available. The available information consists of a set A of examples (a subset of a universe X, X being a nonempty set of objects we want to say something about) of a concept C, and a relation R in X. R models "indiscernibility" or "indistinguishability" and therefore generally is a tolerance relation (i.e., a reflexive and symmetrical relation) and in most cases even an equivalence relation (i.e., a transitive tolerance relation).

Manuscript received February 15, 2005; revised December 18, 2005, March 20, 2006, and June 16, 2006.

The authors are with the Department of Applied Mathematics and Computer Science, Ghent University, B-9000 Gent, Belgium (e-mail: Martine.DeCock@UGent.be; Chris.Cornelis@UGent.be; Etienne.Kerre@UGent.be).

Digital Object Identifier 10.1109/TFUZZ.2006.889762

After a public debate reflecting rivalry between rough set theory and the slightly older fuzzy set theory, many researchers started working towards a hybrid theory (e.g., [11], [18], [20], [25], [26], [28], and [30]). In doing so, the central focus moved from elements' indistinguishability (objects are indistinguishable or not) to their similarity (objects are similar to a certain degree), represented by a fuzzy relation R. As a result, objects are categorized into classes with "soft" boundaries based on their similarity to one another; abrupt transitions between classes are replaced by gradual ones, allowing that an element can belong (to varying degrees) to more than one class.

Soon, researchers started exploring possible applications of the new paradigm of fuzzy-rough hybridization. For a comprehensive literature review up to 1999, we refer to [15]. Among the more recent work is that of Drwal [10] and Jensen and Shen [13], [14], who studied extensions of the well-known rough set approaches to data reduction and classification.

In Section II, we recall the necessary background leading to the definition of a fuzzy rough set as presented in [26]. This definition is an elegant fuzzification of the concept of a rough set and at the same time absorbs earlier suggestions in the same direction. The most striking aspect of all the studies on fuzzy rough set theory mentioned above however is that none of them tries to exploit the fact that an element y of X can belong to some degree to several "soft similarity classes" at the same time. This property does not only lie at the heart of fuzzy set theory but is also crucial in the decision on how to define lower and upper approximations. For instance, in traditional rough set theory, y belongs to the lower approximation of A if the equivalence class to which y belongs is included in A. But what happens if y belongs to several "soft similarity classes" at the same time? Do we then require that all of them are included in A? Most of them? Or just one? And then, which one? In Sections III and IV, we continue this discussion touched upon for the first time in [5].

As such it becomes clear that there is still significant room for improvement and generalization of the definition of a fuzzy rough set, beyond the most "obvious" fuzzification established so far. Furthermore Section V reveals that this generalization is not just of theoretical interest but becomes crucial in a topical application such as query refinement for searching on the WWW.

II. From Rough Sets to Fuzzy Rough Sets

Rough set analysis makes statements about the membership of some element y of X to the concept of which A is a set of examples, based on the indistinguishability between y and the elements of A. To arrive at such statements, A is approximated in two ways. An element y of X belongs to the lower approximation of A if the equivalence class to which y belongs is included in A. On the other hand y belongs to the upper approximation