International Journal of Innovative Computing, Information and Control
Volume **2**, Number **1**, February **2006**

ICIC International ©2006 ISSN 1349-4198
pp. 1–12

MEASURES BASED ON UPPER APPROXIMATIONS OF ROUGH SETS FOR ANALYSIS OF ATTRIBUTE IMPORTANCE AND INTERACTION

Masahiro Inuiguchi and Masayo Tsurumi

Graduate School of Engineering Science
Osaka University
1-3, Machikaneyama, Toyonaka, Osaka 560-8531, Japan
{ inuiguti, tsurumi }@sys.es.osaka-u.ac.jp

Received February 2005; revised September 2005

Abstract. *In rough set analysis, lower approximations have been utilized more frequently than upper approximations. In this paper, we emphasize that a family of upper approximations can retain more information than that of lower approximations. After describing the importance of reducts preserving upper approximations, we propose two measures based on upper approximations: one is a measure of specificity and the other is a classification power index. Finally, we give a numerical example. It shows that the attribute importance calculated by those measures are very different from those calculated by the quality of approximation based on lower approximations.*
Keywords: Rough sets, Reduct, Specificity, Classification power

1. **Introduction.** The literature [6, 12, 16, 17, 18] has demonstrated the usefulness and the effectiveness of the rough sets [10, 11] in analysis of data presented by decision tables. In a decision table, the columns are labeled by attributes which are partitioned into two groups: condition attributes and decision attributes, the rows are labeled by objects of interest and the entities show attribute values. Given a decision table, objects are partitioned into a number of decision classes based on decision attribute values. Then rough sets are defined by pairs of lower and upper approximations of decision classes.

Lower approximations are much more frequently used for the analysis of decision tables than upper approximations. Lower approximations indicate the certain members of decision classes while upper approximations indicate the possible members. Since scientific analysis attaches the importance to the certainty, we can understand why lower approximations are used more frequently.

However, as upper approximations are used in extraction of possible rules, lower approximations are not always sufficient. The information by the lower approximations of decision classes of a partition does not include the information about the boundary regions. On the other hand, based on the upper approximations of decision classes the partition can include the information about the boundary regions as well as the information about the lower approximations. In this sense, knowledge of upper approximations can be more informative than that of lower approximations.

In this paper, emphasizing the usefulness of upper approximation information, we discuss the analysis of attribute importance and interaction among attributes. To analyze