APPROACHES TO THE UTILIZATION OF RULES INDUCED FROM DATASETS

MOTOYUKI OKI

MARCH 2018

APPROACHES TO THE UTILIZATION OF RULES INDUCED FROM DATASETS

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

MOTOYUKI OKI

MARCH 2018

Abstract

With the development of computer, social networking, and sensor networking systems, an enormous amount of data is being stored and made available. Various approaches have been proposed to machine learning and knowledge discovery using those available data. Amongst them, we focus on rough set approaches, which provide useful tools for data analysis. In the rough set approaches, the dataset is arranged into a data table called a decision table where objects are described by condition attributes and a decision attribute. Using the decision attribute, the objects are classified into decision classes. Attribute reduction and rule induction are two successful methods based on rough set approaches. Attribute reduction is a minimum set of condition attributes maintaining the consistency of the decision table. Rule induction is to induce accurate if-then rules inferring memberships of decision classes from the decision table. Such if-then rules are called decision rules.

This thesis describes three approaches to the utilization of rules induced from datasets through rough set approaches. First, we propose robustness measures as a new type of index of rules for evaluating the goodness of rules. The measure evaluates to what extent the interestingness of a rule is preserved against partial matched objects. Various robustness measures are defined depending on the evaluation attitude against uncertainty. Numerical experiments are conducted to examine the usefulness of robustness measures. The rules selected according to the robustness measures are compared to those selected by recall, which is the measure frequently used for selecting good rules. Our results reveal that a different aspect of the goodness of a rule is evaluated by robustness measures. A robustness measure is meaningful as an independent and complementary index of recall.

Second, we propose a rule clustering approach to achieving k-anonymity for privacy protection of rules. Induced rules are not always supported by a sufficient number of objects. If a rule is supported only by a few objects, some condition or decision attribute values of some particular objects can be revealed. If some of the revealed attribute values are sensitive or identify an object, it may invade the privacy of the object. To protect

the privacy, achieving k-anonymity, i.e., using only rules supported by at least k objects is useful. However, erasing induced rules whose supports are less than k deteriorates the accuracy of the rule-based classifier. Therefore, we propose the merging of rules with low support based on the similarity and consistency of rules. In this way, we try to achieve the k-anonymity without any significant deterioration of the accuracy. Two numerical experiments are conducted. One is to verify the accuracy of the rule-based classifier with the k-anonymized rules, and the other is to evaluate the risk of privacy invasion of decision attribute values. The experimental results demonstrate the usefulness of the proposed method.

Third, we propose a mixture model approach to constructing an effective rule-based classifier by utilizing the published rules. When published rules are available, we may utilize them to construct an effective and useful rule-based classifier by incorporating them into the rules induced from the dataset at hand. However, the published rules may include some special characters caused by territorial, cultural, ethnic or religious characters, different age strata bias and so on. Then, we apply the idea of the mixture model in order to utilize both the published rules and the rules induced from the dataset at hand. By applying LERS algorithm to each set of rules, we obtain two score distributions over decision classes. To combine those two score distributions, we apply the mixture model. EM algorithm is then used for estimating the appropriate mixture ratios for constructing a rule-based classifier. Numerical experiments are conducted to examine the performance of the proposed approach. Our approach is compared with four alternative methods using several datasets obtained from public domain. The effectiveness of the proposed approach is demonstrated by the numerical experiments.

Finally, the results obtained in this thesis are summarized. The proposed methods will be useful for the utilization of induced rules in real-world problems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of This Study

With the development of computer, social networking, and sensor networking systems, an enormous amount of data is being stored and made available. Owing to the availability of large amounts of data, there has been a considerable increase in the need for the effective utilization of data. The development of techniques for data mining and knowledge discovery from data is strongly required. Various approaches have been proposed to machine learning and knowledge discovery using those available data for real-world problems. Amongst them, we focus on rough set approaches, which provide useful tools for data analysis. In the rough set approaches proposed by Pawlak [59], the dataset is arranged into a data table called a decision table where objects are described by condition attributes and a decision attribute. Using the decision attribute, the objects are classified into decision classes. Attribute reduction and rule induction are two successful methods for data analysis based on rough set approaches. Attribute reduction is a minimum set of condition attributes maintaining the consistency of the decision table. Rule induction is to induce accurate if-then rules inferring memberships of decision classes from the decision table. In rough set approaches, such if-then rules are called decision rules. A decision rule is composed of a premise described by condition attribute values and a conclusion describing the attribute attribute values or decision classes. For convenience, the decision rule is simply called 'rule' throughout this thesis. The rough set approaches are applied to various fields such as medicine, Kansei engineering, economics and finance, decision analysis, environmental issues, and web mining [6, 15, 16, 42, 62, 69, 73, 74]. In this thesis, we tackle three problems for the utilization of rules induced from datasets : selection of good rules, privacy preservation

for publishing rules, and utilization of published rules. The backgrounds of these topics are described in the following three sections.

1.2 Selection of Useful Rule by Interestingness Measure

Although induced rules are typically used to construct a rule-based classifier which predicts the decision classes of unseen objects, the rules are also useful for design knowledge. For example, rules that indicate the relations between the product design and the successful selling of products can be deemed design knowledge. New products that satisfy the conditions of acquired rules may attract customers more than those that do not [42]. Although we are interested in a good set of rules for constructing a rule-based classifier, we are also interested in good individual rules that are useful as design knowledge. In this sense, we consider an index to evaluate the goodness of individual rules. Selecting useful rules based on the appropriate evaluation is an essential task for constructing a rule-based classifier and supporting the decision criteria for designers.

To evaluate the goodness of a rule, various measures of rule interestingness called interestingness measures have been proposed in previous studies [20, 39]. Interestingness measures have been developed to find meaningful patterns and construct an effective rule-based classification system composed of several association rules. These measures are different from those with the primary characteristics of generality, conciseness, reliability, novelty, surprisingness, utility, actionability, and so on. (see reviews [11, 37, 38, 61, 67, 72].) Support and confidence [3], gain [17], and conviction [13] are the most commonly used interestingness measures. In addition to the above, recall has been commonly employed and is one of the most suitable measures to evaluate the goodness of a rule [21, 42]. This measure is also called coverage in a rough set community. In particular, when rules are 100% confident, recall is often used for rule evaluation.

Interestingness measures evaluate the strength of the relation between the conditions and the conclusion of a rule. However, they do not evaluate the strength of the relation between some of the conditions and the conclusion of a rule. Considering the potential of partially-matched data in new objects, a good evaluation of the rule may be performed if the rule maintains the confidence of its conclusion even for partially matched data. Such a rule would also be suitable for design knowledge because the newly designed items may fail to satisfy some conditions of the rule despite the designer's intention. In Chapter 3, we propose novel interestingness measures as a new type of index of rules for evaluating to what extent the interestingness is preserved against the partial matching for the condition.

1.3 Privacy Preservation for Publishing Rules

When rule-based classifiers constructed by rules are used in real-world problems, we may be requested to publish them to ensure fairness and to provide knowledge for field researchers. The development of methods and tools for publishing rules while ensuring the confidentially of personal sensitive and private information is a task of utmost importance. This undertaking is called privacy-preserving data publishing [18]. In previous studies, research communities have responded to this challenge and proposed many approaches [18, 28]. In the field of frequent pattern mining such as Apriori algorithm [4], various kinds of privacy models have been proposed to preserve the privacy of personal information [33, 43, 66, 75].

Induced rules are not always supported by a sufficient number of objects because objects having rare condition attribute values may exist in a dataset. If a rule is supported only by a few objects, some condition or decision attribute values of some particular objects can be revealed. It some of the revealed attribute values are sensitive or identify an object, it may invade the privacy of the object. This identifiability is totally undesirable, especially for sensitive information. That is, the development of techniques for privacy protection is a very important issue for published rules. As a solution, each rule should be supported by at least a certain number of objects in the dataset. For instance, the concept of k-anonymity [65], which requires at least k objects supporting each published data pattern, could be applied to generate the rules. Inuiguchi et al. [30] applied the concept of k-anonymity to the privacy protection for the publication of rules. They proposed a method for inducing the rules satisfying k-anonymity by relaxing the conclusions of rules as the membership to a union of decision classes. They maintained the classification accuracy by taking an intersection of the rules with unions of decision classes. However, this method can be applied only to datasets with more than two decision classes. Namely, the approach is not applicable for datasets with two decision classes. Clustering approaches have been proposed for achieving k-anonymity for privacy protection [31, 32]. These approaches are called k-member clustering methods and protect the data privacy. Their methods have not been applied to publishing rules. Naive approaches that involves erasing rules, whose supports are less than k, often deteriorate the accuracy of the rule-based classifier. In Chapter 4, we investigate a method for anonymizing rules while the rule-based classifier can maintain high classification accuracy when datasets have only two decision classes.

1.4 Utilization of Published Rules

When the rules published by different organizations and companies are available, it is worthwhile to utilize the published rules for constructing a more effective and useful rule-based classifier by incorporating them into the rules induced from the dataset at hand. However, we would not be able to know the dataset from which the published rules originated. The published rules can be induced from some biased dataset. For instance, a number of objects with a condition attribute value exist in the biased dataset compared to our dataset at hand. That is, the published rules may include some special character caused by territorial, cultural, ethnic or religious character, different age strata bias and so on. If we incorporate a published set of rules induced from a potentially biased dataset into the rules induced from the dataset at hand, the performance of the rule-based classifier that incorporates the published set of rules may deteriorate unless the potential bias is properly treated.

Nevertheless, it would be useful to take advantage of the knowledge of the published rules. By utilizing the multiple sets of rules induced from various kinds of datasets, *e.g.*, medical data in different hospitals and questionnaire data from various people, we may construct a more widely applicable and useful rule-based classifier. In previous studies, various rule ensemble models obtained by some ensemble learning techniques, *e.g.*, Bagging [12], have been proposed [8, 9, 63, 70]. However, the sets of rules in these approaches are obtained in bootstrap samples and drawn from a single dataset. Few studies have devoted to the combination of sets of rules obtained from different datasets such as the published rules and rules induced from a dataset at hand. In Chapter 5, we propose a method for utilizing the published set of rules together with rules induced from the dataset at hand.

1.5 Thesis Overview

This thesis studies on three approaches to the utilization of rules induced from datasets. The rules are assumed to be induced by rough set approaches so that they are accurate rules with minimal conditions. The relations between the chapters of this thesis can be explained as shown in Fig 1.1. Chapter 2 presents the basic concepts of the rough set theory. The rule induction method used in this thesis is described. More concretely, MLEM2 algorithm [25], which is a rule induction algorithm, is described. To construct rule-based classifiers using induced rules, LERS algorithm [22], which is a classification algorithm using rules, is described. In addition, interestingness measures [20, 39], which are indices to evaluate the goodness of an if-then rule, are described. The measures

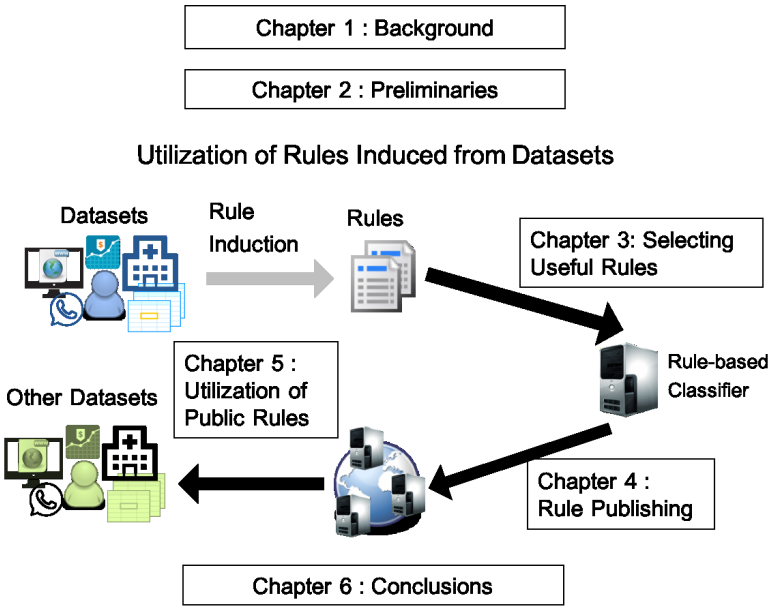


FIGURE 1.1: Structure of this thesis

frequently used in the previous studies are compared to the proposed measure to evaluate the interestingness of a rule in Chapter 3. Techniques for privacy preservation in data publishing are described. k-anonymity [65] is one of the most successful privacy preserving methods. The decision table is said to satisfy k-anonymity if each pattern appearing in the decision table has at least k objects in the decision table. Expectation-Maximization (EM) algorithm [14], which estimates the mixture rates of different distributions from a given dataset, is described. The EM algorithm alternates between executing an expectation (E) step and a maximization (M) step. The E step computes the expectation of the log-likelihood function using the provisional values of parameters, while the M step updates the provisional values of parameters by maximizing the expected log-likelihood.

In Chapter 3, we propose robustness measures as a new type of evaluation index for rules in order to select good rules. This measure evaluates to what extent the interestingness of a rule is preserved against partial matched objects. Various robustness measures are defined depending on the evaluation attitude against uncertainty. Four numerical experiments are conducted to examine the usefulness of the robustness measures. The rules selected according to the robustness measures are compared with those selected according to the recall measure, which is a well-known measure used for selecting good rules. The results reveal that a different aspect of the goodness of a rule is evaluated by robustness measures. This shows that a robustness measure is meaningful as an independent and complementary index of the recall measure.

In Chapter 4, we propose a rule clustering approach to achieving the privacy protection

for rule publication. To protect the privacy, we use the concept of k-anonymity [65], so that only the rules that support at least k objects are used. To achieve k-anonymity, the rules with small support scores are merged if they are similar and the resulting merged rule has high consistency. In this way, a set of k-anonymous rules is generated without any significant deterioration of the classification accuracy of a rule-based classifier. Two numerical experiments are conducted. One is to verify the classification accuracy of the rule-based classifier with the k-anonymized rules, and the other is to evaluate the risk of privacy invasion on decision attribute values. The experimental results demonstrate the usefulness of the proposed method.

In Chapter 5, we propose a mixture model approach to the utilization of published rules for enhancing the effectiveness of a rule-based classifier. The mixture model is used to construct a better rule-based classifier by combining the published rules and the rules induced from a dataset at hand. By applying LERS algorithm to each set of rules, we obtain adequacy scores over decision classes. EM algorithm [14] is then used for estimating the approximate mixture ratios of the adequacy scores. Then the mixed scores over the decision classes is used to estimate the decision class of an object. Numerical experiments are conducted to examine the performance of the proposed approach. The proposed method is compared with four alternative methods using several datasets obtained from public domain. The effectiveness of the proposed method is demonstrated by the numerical experiments.

Finally, in Chapter 6, we summarize the results of these studies, and give the directions for future research.

Chapter 2

Preliminaries

In this chapter, we introduce notions and basic concepts which are necessary in the following chapters. In Section 2.1, the definition of a rough set and its properties are described. In Section 2.2, we explain a rule induction algorithm based on the rough set theory. The rule induction algorithm is called Modified Learning from Examples Module version 2 (MLEM2). In Section 2.3, we explain a classification algorithm using rules. It is called Learning from Examples based on Rough Sets (LERS). In Section 2.4, Interestingness measures evaluating the goodness of a rule are described. In Section 2.5, the concept of k-anonymity which is one of the most successful privacy preserving methods is described. The concept is used in Chapter 4. In Section 2.6, Expectation-Maximization (EM) algorithm which alternates between executing an expectation (E) step and a maximization (M) step is described. The algorithm is used in Chapter 5.

2.1 Rough Sets

2.1.1 Information Table and Indiscernibility Relation

In rough set theory, a data table is modeled by an information table. The information table is defined by a quadruple $\mathcal{T} = \langle U, A, V, \rho \rangle$, where U called universe of discourse is a finite set of objects, A is a finite set of attributes, $V = \bigcup_{a \in A} V_a$ and V_a is a set of all values of attribute $a \in A$, and $\rho : U \times A \to V$ is a total function called an information function.

Given an information table and an attribute set $B \subseteq A$, a B-indiscernibility relation is a relation in which each pair of objects have the same attribute values with respect

to B. Namely, It is defined by

$$I_B = \{(x, y) \in U \times U : \rho(x, a) = \rho(y, a), \forall a \in B\}. \tag{2.1.1}$$

I_B is obviously an equivalence relation, i.e., it satisfies reflexive, symmetric and transitive. The indiscernibility relation I_B divides the object set U into a partition, which is a set of equivalence classes of objects. An equivalence class of $x \in U$ with respect to B is a set of objects which is defined by

$$I_B(x) = \{y \in U : (y, x) \in I_B\}. \tag{2.1.2}$$

Equivalence classes with respect to B are also referred to as B-elementary sets. Elementary sets are constructing blocks to represent concepts. The unions of B-elementary sets are called B-definable sets.

2.1.2 Definitions and Properties of Rough Sets

Let X be a subset of objects. When an object $x \in X$ can be B-indiscernible to another object $y \in U - X$, we say that a given attribute set B is insufficient. Because of the indiscernibility relation, x is certainly classified neither to X nor to $U - X$, namely X is an inconsistent object set. In rough set theory, such inconsistency is captured by using lower and upper approximations of X [59].

Definition 2.1. An information table $\mathcal{T} = \langle U, A, V, \rho \rangle$ is given. For $X \subseteq U$ and $B \subseteq A$, B-lower approximation $\underline{B}(X)$ and B-upper approximation $\overline{B}(X)$ are defined as follows:

$$\underline{B}(X) = \{x \in U : I_B(x) \subseteq X\}, \tag{2.1.3}$$

$$\overline{B}(X) = \{x \in U : I_B(x) \cap X \neq \emptyset\}. \tag{2.1.4}$$

B-lower approximation of X is an object set in which each object x is certainly classified to X, because all objects which are B-indiscernible to x are included in X, namely, there is no object which is against $x \in X$. B-upper approximation of X is interpreted as an object set whose member x is possibly classified to X, because some objects which are B-indiscernible to x are included in X. The difference of B-lower and B-upper approximations of X is called B-boundary region $BND_B(X)$, which is formally defined by

$$BND_B(X) = \overline{B}(X) - \underline{B}(X). \tag{2.1.5}$$

When the boundary region of X is nonempty, X is a rough set; otherwise it is a crisp set. The lower and upper approximations can be expressed by a union of elementary sets [59, 60].

Proposition 2.2. *We have the following equations.*

$$\underline{B}(X) = \bigcup\{I_B(x) : I_B(x) \subseteq X\}, \quad \overline{B}(X) = \bigcup\{I_B(x) : I_B(x) \cap X \neq \emptyset\}. \tag{2.1.6}$$

For any subset $B \subseteq A$, its B-lower and B-upper approximations are B-definable.

We have algebraic properties of lower and upper approximations [59, 60].

Proposition 2.3. *We have the following relations.*

$$\underline{B}(X) \subseteq X \subseteq \overline{B}(X), \tag{2.1.7}$$

$$\underline{B}(\emptyset) = \overline{B}(\emptyset) = \emptyset, \quad \underline{B}(U) = \overline{B}(U) = U, \tag{2.1.8}$$

$$\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y), \quad \overline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y), \tag{2.1.9}$$

$$X \subseteq Y \Rightarrow \underline{B}(X) \subseteq \underline{B}(Y), \quad X \subseteq Y \Rightarrow \overline{B}(X) \subseteq \overline{B}(Y), \tag{2.1.10}$$

$$\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y), \quad \overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y), \tag{2.1.11}$$

$$\underline{B}(U - X) = U - \overline{B}(X), \quad \overline{B}(U - X) = U - \underline{B}(X), \tag{2.1.12}$$

$$\underline{B}(\underline{B}(X)) = \overline{B}(\underline{B}(X)) = \underline{B}(X), \quad \overline{B}(\overline{B}(X)) = \underline{B}(\overline{B}(X)) = \overline{B}(X). \tag{2.1.13}$$

2.1.3 Decision Table

When a set of attributes A are divided into condition attributes C and a decision attribute $\{d\}$. The $\{d\}$-indiscernibility relation derive a classification defined by the partition $\mathcal{D} = \{D_1, \ldots, D_p\}$. An information table $\mathcal{T} = \langle U, C \cup \{d\}, V, \rho\rangle$ is referred to as a decision table. $D_i \in \mathcal{D}$ is called a decision class.

2.2 Rule Induction

Rule induction is to induce accurate if-then rules inferring memberships of decision classes from the decision table. Such if-then rules are called decision rules. Decision rules induced from decision tables are represented as logical expressions of the following form: $E \to H$. Condition part E is comprised of elementary conditions with respect to condition attributes C and conjunctive operations $\wedge$. An elementary condition with respect to $a \in C \cup \{d\}$ is a constraint about a, e.g. $\rho(x, a) = v_a$ or $\rho(x, a) \in W_a$, where $v_a \in V_a$ and $W_a \subseteq V_a$. Conclusion part H is comprised of atomic propositions with

respect to a decision attribute d. Decision rule "$E \rightarrow H$" indicates that an object x satisfying E should also satisfy H. Besides rough set approaches, approaches to rule induction in the field of machine learning techniques have been proposed [5, 22, 35, 64, 68]. However, rough sets approaches treat inconsistencies in a decision table in comparison to other machine learning techniques. If the decision table includes any inconsistent objects, decision rules could be induced from either lower, upper approximations or boundary regions of decision classes. A number of algorithms to induce rules using the rough set approach have been proposed [5, 22, 35, 64].

In this thesis, among previously proposed algorithms, we use MLEM2 algorithm [25] as a rule induction algorithm, which is an extension of LEM2 [22], because it is successfully used in several fields and developed to treat both of nominal and numerical attributes [23, 26, 27]. It is inspired by the ideas of handling numerical attributes earlier introduced in the MODLEM algorithm [24]. In MLEM2 algorithm, elementary condition of decision rules are represented as $\rho(x, a_k) \geq v_{a_k}^{\mathrm{L}}$ or $\rho(x, a_k) < v_{a_k}^{\mathrm{U}}$ for numerical attributes a_k, and as $\rho(x, a_h) = v_{a_h}$ for nominal attributes a_h, where $a_k, a_h \in C$, $v_{a_k}^{\mathrm{L}}, v_{a_k}^{\mathrm{U}} \in V_{a_k}$ and $v_{a_h} \in V_{a_h}$. If the same numerical attribute is chosen twice while constructing a single rule, we obtain the condition $\rho(x, a_k) \in [v_{a_k}^{\mathrm{L}}, v_{a_k}^{\mathrm{U}})$ by taking an intersection of two conditions $\rho(x, a_k) \geq v_{a_k}^{\mathrm{L}}$ and $\rho(x, a_k) < v_{a_k}^{\mathrm{U}}$. Conclusions of decision rules are usually represented as $\rho(x, d) = v_d$, where $v_d \in V_d$.

MLEM2 algorithm is shown in Algorithm 1. This algorithm is based on the idea of a sequential covering and it generates a minimal set of conditions from a rough approximation of each class D_i. The main procedure for rule induction starts by creating the first rule, i.e., sequentially choosing the best elementary conditions according to chosen criteria. We can set an option of MLEM2 with or without taking into account attribute priorities (see Line 6 of Algorithm 1). When MLEM2 is not to take attribute priorities into account, i.e., all attribute priorities are equal to each other, the first criterion is ignored.

When a rule is composed and saved (see Line 17 of Algorithm 1), all learning examples that match this rule are removed from consideration (see Line 18 of Algorithm 1). The process is iteratively repeated as long as some objects from the rough approximation remain still uncovered. MLEM2 induces a minimal set of rules with minimal conditions that covers learning examples from a chosen rough approximation, i.e., objects belonging either to $B = \underline{C}(D_i)$ or $\overline{C}(D_i)$ as an input data. Thus, one can choose between set of certain rules surely concluding D_i by setting $B = \underline{C}(D_i)$, and a set of possible rules possibly concluding D_i by setting $B = \overline{C}(D_i)$.

Algorithm 1 MLEM2

Input: a set B

Output: a single local covering $\mathbb{T}$ of B

1: $\mathbb{T} := \emptyset$;
2: $G := B$;
3: **while** $G \neq \emptyset$ **do**
4: $T := \emptyset$; $T(G) := \{t : [t] \cap G \neq \emptyset\}$;
5: **while** $T = \emptyset$ **or** $[T] \not\subseteq B$ **do**
6: † select $t \in T(G)$ with the highest attribute priority, if a tie occurs, select $t \in T(G)$ such that $|[t] \cap G|$ is maximum; if another tie occurs, select $t \in T(G)$ with smallest cardinality of $[t]$; if a further tie occurs, select a first one;
7: $T := T \cup \{t\}$;
8: $G := [t] \cap G$;
9: $T(G) := \{t : [t] \cap G \neq \emptyset\}$;
10: $T(G) := T(G) - T$;
11: **end while**
12: **for each** t in T **do**
13: **if** $[T - \{t\}] \subseteq B$ **then**
14: $T := T - \{t\}$;
15: **end if**
16: **end for**
17: $\mathbb{T} := \mathbb{T} \cup \{T\}$;
18: $G := B - \bigcup_{T \in \mathbb{T}} [T]$;
19: **end while**
20: **for each** T in $\mathbb{T}$ **do**
21: **if** $\bigcup_{S \in \mathbb{T} - \{T\}} [S] = B$ **then**
22: $\mathbb{T} := \mathbb{T} - \{T\}$;
23: **end if**
24: **end for**

2.3 Classification Algorithm

When a set of rules is obtained from a rule induction algorithm, an unseen object x can be classified based on matching its condition attribute values to the condition parts of rules. However, there is no guarantee that at least one of rules is matched and conclusions of all matched rules are consistent. To classify objects of either unmatched or multiple matching to a set of rules, a rule-based classifier using a classification algorithm should be constructed. We use a classification algorithm implemented in rule induction system Learning from Examples based on Rough Sets (LERS) [22]. Other classification algorithm can be seen in [5, 19, 68]. That classification algorithm is based on strength, specificity and matching factor [25]. $Strength(r)$ is the total number of objects in a given decision table correctly classified by rule r. $Specificity(r)$ is the total number of conditions in the premise of rule r. $Matching\_factor(r)$ is the ratio of the number of matched conditions of rule r to the total number of those of r. When conditions of several rules indicating different decision values are satisfied with a classified object, the

following measure $S(v_i^d)$ is calculated:

$$S(v_i^d) = \sum_{\text{matching rules } r \text{ for } v_i^d} Strength(r) \times Specificity(r), \tag{2.3.1}$$

where r is called a matching rule if the premise of r is satisfied. For convenience, when rules from a particular decision values v_i^d are not matched by the object, we define $S(v_i^d) = 0$. There may be also another ambiguous situation where no conditions of rules obtained by MLEM2 are satisfied with the object. In this case the following measure $M(v_i^d)$ is used:

$$M(v_i^d) = \sum_{\substack{\text{partially matching} \\ \text{rules } r \text{ for } v_i^d}} Matching\_factor(r) \times Strength(r) \times Specificity(r), \tag{2.3.2}$$

where r is called a partially matching rule if a part of the premise of r is satisfied. Classification can be performed as follows: if there exists v_j^d such that $S(v_j^d) > 0$, v_i^d with the largest $S(v_i^d)$ is selected. Otherwise, v_i^d with the largest $M(v_i^d)$ is selected.

Here, we explain the classification method used in LERS. Let R be a subset of induced rules and u an object to be classified. We define $R(D_i)$ as a set of all rules in R inferring the belongingness to a class D_i, $Mat(u, R)$ as a set of rules in R whose conditions completely satisfy object u, and $PM(u, R)$ as a set of rules in R whose conditions are partially satisfied with object u. When all conditions of some rules in R satisfy an object u (i.e., $Mat(u, R) \neq \emptyset$), then the following measure $S(D_i)$ is calculated:

$$S(D_i) = \sum_{\substack{r \in R(D_i) \cap \\ Mat(u,R)}} Strength(r) \times Specificity(r), \tag{2.3.3}$$

where r is a rule, $Strength(r)$ is the total number of objects in the given decision table correctly classified by rule r and $Specificity(r)$ is the total number of conditional attributes in the condition for rule r. For convenience, when $Mat(u, R) = \emptyset$, we define $S(D_i) = 0$. When no rule exists whose conditions completely satisfy object u (i.e., when $Mat(u, R) = \emptyset$), the following measure $M(D_i)$ is calculated:

$$M(D_i) = \sum_{\substack{r \in R(D_i) \cap \\ PM(u,R)}} Match\_factor(r) \times Strength(r) \times Specificity(r), \tag{2.3.4}$$

where $Match\_factor(r)$ is the ratio of the number of matched conditions for conditional attributes of rule r to the total number of conditional attributes used in rule r. The classification is performed as follows: if D_j exists such that $S(D_j) > 0$, the class D_i with the largest $S(D_l)$ is selected; otherwise, the class D_i with the largest $M(D_i)$ is selected.

2.4 Interestingness Measures

To evaluate the goodness of a rule, various measures of rule interestingness called interestingness measure have been proposed in previous studies [20, 39]. These measures are different from one another in the primary characteristics of generality, conciseness, reliability, novelty, surprisingness, utility, actionability, and so on (see reviews [11, 37, 38, 61, 67, 72]). Interestingness measures have been developed to find meaningful patterns and construct an effective classification system composed of several association rules [20, 39]. Support and confidence [3], gain [17], and conviction [13] are the most commonly used interestingness measures. The support of statement E, denoted by $supp(E)$, is defined by the ratio of the number of objects that satisfy E to the number of all objects in a decision table. The support of a rule $E \rightarrow H$ is defined by $supp(E \rightarrow H) = supp(E \wedge H)$. The confidence evaluates to what extent the condition part of a rule is covered by the conclusion of a rule in a decision table. The confidence of a rule $E \rightarrow H$ is defined by $conf(E \rightarrow H) = supp(E \wedge H)/supp(E)$. As a counterpart of the confidence, recall has been also commonly employed and is one of the measures to evaluate the usefulness of a rule under a constant confidence [21, 42]. The recall evaluates to what extent the conclusion part of a rule is covered by the condition part of a rule in the decision table. This is also called coverage in a rough set community. In particular, when rules are 100% confident, recall is often used for rule evaluation. The recall of a rule $E \rightarrow H$ is defined by $rec(E \rightarrow H) = supp(E \wedge H)/supp(H)$.

2.5 k-Anonymity

In this section, we describe k-anonymity which is one of the most successful methods in the field of privacy preserving data mining. A development of privacy preserving data mining has become more important in recent years because of the increasing ability to store personal data about users and the sophistication of data mining algorithms to treat sensitive information of users [1]. In Chapter 4, we propose an approach to achieve k-anonymity of rule for privacy protection. The approach is inspired by conventional concept of k-anonymity on a decision table.

k-anonymity [1, 65] is a property that protect published data against possible re-identification of the objects to whom the published data refer. We may be requested to publish decision tables we have at hand. In such cases, we publish the decision table where data have been de-identified by removing explicit identifiers (e.g., social security number and name). However, some published attributes such as ZIP, Date of birth, Marital status, and Sex can also appear in other published external decision tables in

the external domain. If some combinations of these attribute values are unique or rare, then parties observing their decision tables can link object in one published table to that in published table through the rare combination of attribute values include in both tables. Namely, the parties can reveal more attribute values of an object than those in both tables. If the revealed values are sensitive, it may invade the privacy of the object. This identifiability is totally undesirable, especially for sensitive information.

To protect the privacy of objects from such identification, Sweeney [65] proposed k-anonymity property. k-anonymity states that each object in a decision table must be indiscernible from at least k objects. Two main techniques have been proposed for enforcing k-anonymity on a decision table : *generalization* and *suppression* [1]. *Generalization* consists in replacing attribute values with a generalized version of them. Generalization is based on a domain generalization hierarchy and a corresponding value generalization hierarchy on the values in the domains. Typically, the domain generalization hierarchy is a total order and the corresponding value generalization hierarchy a tree, where the 'parent/child' relationship represents the direct 'generalization/specialization' relationship. *Suppression* consists in protecting sensitive information by removing it. Suppression, which can be applied at the level of a specific attribute value of objects or a set of objects or condition attributes, is a naive technique to be enforced to achieve k-anonymity. The application of these approaches to a decision table may generate a less precise, more general and less complete table. It is important to minimize the information loss of precision and completeness of the decision table caused by the applications of these techniques. A problem of finding minimal k-anonymous decision tables with generalization and suppression has been proved to be computationally hard [2, 41]. Other techniques to achieve k-anonymity developed in recent years and those extensions to rules for privacy preserving algorithm can be reviewed in [1, 18, 31, 33, 43].

2.6 EM Algorithm

In this section, we describe Expectation-Maximization (EM) algorithm which alternates between executing an expectation (E) step and a maximization (M) step. In Chapter 5, this algorithm is applied to construct a mixture model between published rules and rules at hand.

2.6.1 General EM Algorithm

EM algorithm [7, 14] is an efficient iterative procedure to compute the maximum likelihood estimate in the presence of missing or latent valiable. In this subsection, we

describe general EM algorithm used in Chapter 5 based on reference [7]. In maximum likelihood estimation, we wish to estimate model parameters for which the observed data are the most likely. Each iteration of EM algorithm consists of expectation step (E-Step) and maximization step (M-step). In the E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimate of the missing data from the E-step are used in lieu of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration.

The goal of EM algorithm is to find maximum likelihood solutions for models having latent variables. We denote the set of all observed data by $\mathbf{X}$, in which the n row represents $\mathbf{x}_n^{\mathrm{T}}$, and similarly we denote the set of all latent variables by $\mathbf{Z}$, with a corresponding row $\mathbf{z}_n^{\mathrm{T}}$. The set of all model parameters is denoted by $\boldsymbol{\theta}$. The log likelihood function is given by

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log\left\{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})\right\}. \tag{2.6.1}$$

A key observation is that the summation over the latent variables $\mathbf{Z}$ appears inside the logarithm. Even if the joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ belongs to the exponential family, the marginal distribution $p(\mathbf{X}|\boldsymbol{\theta})$ typically does not as a result of this summation. The presence of the sum prevents the logarithm from acting directly on the joint distribution, resulting in complicated expressions for the maximum likelihood solution. Now suppose that, for each observation in $\mathbf{X}$, we were told the corresponding value of the latent variable $\mathbf{Z}$. We shall call $\mathbf{X}, \mathbf{Z}$ the complete data set, and we shall refer to the actual observed data $\mathbf{X}$ as incomplete. The likelihood function for the complete data set simply takes the form ln $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$, and we shall suppose that maximization of this complete-data log likelihood function is straightforward.

However, we are not given the complete data set $\mathbf{X}, \mathbf{Z}$, but only the incomplete data $\mathbf{X}$. Our state of knowledge of the values of the latent variables in $\mathbf{Z}$ is given only by the posterior distribution $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$. Because we cannot use the complete data log likelihood, we consider instead its expected value under the posterior distribution of the latent variable, which corresponds to the E step of EM algorithm. In the subsequent M step, we maximize this expectation. If the current estimate for the parameters is denoted $\boldsymbol{\theta}^{\mathrm{old}}$, then a pair of successive E and M steps gives rise to a revised estimate $\boldsymbol{\theta}^{\mathrm{new}}$. The algorithm is initialized by choosing some starting value for the parameters θ_0.

In the E step, we use the current parameter values $\boldsymbol{\theta}^{\text{old}}$ to find the posterior distribution of the latent variables given by $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$. We then use this posterior distribution to find the expectation of the complete data log likelihood evaluated for some general parameter value $\boldsymbol{\theta}$. This expectation denoted $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ is given by

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}) \log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}). \tag{2.6.2}$$

In the M step, we determine the revised parameter estimate $\boldsymbol{\theta}^{\text{new}}$ by maximizing this function

$$\boldsymbol{\theta}^{\text{new}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}). \tag{2.6.3}$$

Note that in the definition of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$, the logarithm acts directly on the joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$, and so the corresponding M step maximization will be tractable.

The general EM algorithm is summarized below. It has the property, as we shall show later, that each cycle of EM will increase the incomplete data log likelihood.

STEP 0 Given a joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ over observed variables $\mathbf{X}$ and latent-variables $\mathbf{Z}$ governed by parameters $\boldsymbol{\theta}$, the goal is to maximize the likelihood function $p(\mathbf{X}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

STEP 1 Choose an initial setting for the parameters $\boldsymbol{\theta}^{\text{old}}$.

STEP 2 E Step : Comupte $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$.

STEP 3 M Step : Compute $\boldsymbol{\theta}^{\text{new}} = \operatorname{argmax}_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$.

STEP 4 Check for convergence of either the log likelihood or the parameter values. If the convergence criterion is not satisfied, then update $\boldsymbol{\theta}^{\text{new}} \rightarrow \boldsymbol{\theta}^{\text{old}}$ and return to STEP 2.

2.6.2 Mixtures of Bernoulli Distributions

We illustrate EM algorithm for discrete binary variables described by Bernoulli distributions based on reference [7]. This model is also known as latent class analysis [7, 36, 40]. Consider a set of D binary variables x_i, where $i = 1, \ldots, D$, each of which is governed by a Bernoulli distribution with parameter μ_i, so that

$$P(\mathbf{x}|\boldsymbol{\mu}) = \prod_{i=1}^{D} \mu_i^{x_i} (1 - \mu_i)^{(1-x_i)}, \tag{2.6.4}$$

where $\mathbf{x} = (x_1, \ldots, x_D)^T$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_D)^T$. Given $\boldsymbol{\mu}$, we see that the individual variables x_i are independent. Let us consider a finite mixture of these distributions given by

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|\boldsymbol{\mu}_k), \tag{2.6.5}$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$, $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_K\}$, and

$$p(\mathbf{x}|\boldsymbol{\mu}_k) = \prod_{i=1}^{D} \mu_{ki}^{x_i}(1 - \mu_{ki})^{(1-x_i)}. \tag{2.6.6}$$

When we are given a data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, the log likelihood function for this model is given by

$$p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \log\left\{\sum_{k=1}^{K} \pi_k p(\mathbf{x}_n|\boldsymbol{\mu}_k)\right\}. \tag{2.6.7}$$

We confirm the appearance of the summation inside the logarithm. Then, the maximum likelihood solution no longer has closed form. We derive the EM algorithm for maximizing the likelihood function for the mixture of Bernoulli distributions. To do this, we first introduce an explicit latent variable $\mathbf{z}$ associated with each instance of $\mathbf{x}$. $\mathbf{z} = (z_1, ..., z_K)^T$ is a binary K-dimensional variable having a single component equal to 1, with all other components equal to 0. We can then write the conditional distribution of $\mathbf{x}$, given the latent variable, as

$$p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}) = \prod_{k=1}^{K} p(\mathbf{x}|\boldsymbol{\mu}_k)^{z_k}. \tag{2.6.8}$$

The prior distribution for the latent variables is

$$p(\mathbf{z}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_k}. \tag{2.6.9}$$

In order to derive EM algorithm, we first write down the complete data log likelihood function, which is given by

$$\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left\{\log \pi_k + \sum_{i=1}^{D}[x_{ni} \log \mu_{ki} + (1 - x_{ni}) \log(1 - \mu_{ki})]\right\}, \tag{2.6.10}$$

where $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$. Next we take the expectation of the

complete data log likelihood with respect to the posterior distribution of the latent variables to give

$$\mathbb{E}_{\mathbf{z}}[\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\pi})] = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\left\{\log \pi_k + \sum_{i=1}^{D}[x_{ni}\log \mu_{ki} + (1 - x_{ni})\log(1 - \mu_{ki})]\right\}, \tag{2.6.11}$$

where $\gamma(z_{nk}) = \mathbb{E}[z_{nk}]$ is the posterior probability of component k given data point $\mathbf{x}_n$. In the E step, these probabilities are evaluated the form

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \frac{\pi_k p(\mathbf{x}_n|\boldsymbol{\mu}_k)}{\sum_{j=1}^{K}\pi_j p(\mathbf{x}_n|\boldsymbol{\mu}_j)}. \tag{2.6.12}$$

When we consider the sum over n in Equation (2.6.11), two terms are derived as

$$N_k = \sum_{n=1}^{N}\gamma(z_{nk}) \tag{2.6.13}$$

and

$$\overline{\mathbf{x}}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathbf{x}_n. \tag{2.6.14}$$

N_k is the number of data points associated with component k. In the M step, we maximize the expected complete data log likelihood with respect to the parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\pi}$. When we set the derivative of Equation (2.6.11) with respect to $\boldsymbol{\mu}_k$ equal to zero and rearrange the terms, we obtain $\boldsymbol{\mu}_k = \overline{\mathbf{x}}_k$. We confirm that this sets the mean of component k equal to a weighted mean of the data, with weighting coefficients given by the probabilities that component k takes for data points. For the maximization with respect to π_k, we need to introduce a Lagrange multiplier to enforce the constraint $\sum_k \pi_k = 1$. Then we obtain $\pi_k = N_k/N$ which represents the intuitively reasonable result that the mixing coefficient for component k is given by the fraction of points in the data set explained by that component.

Chapter 3

Robustness Measures for Selecting Useful Rules

In this chapter, we propose robustness measures as a new type of interestingness measure for rules. Interestingness measures evaluate the strength of the relation between the conditions and the conclusion of a rule. For objects match the condition of the rule with a big score, the conclusion of the rule is appropriate. However, for the objects that partially match the condition of the rule, the conclusion is not always appropriate. Considering the fact that partially matched rules are often used for estimating the conclusion when no rules are matched for a given object, this potential decay is not desirable. Moreover, when a rule is used for design knowledge, this potential decay may lead to a fail due to slight misunderstanding of the condition of the rule. Therefore, we propose robustness measures for evaluating to what extent the interestingness of a rule is preserved against the partial matching for the condition. Various robustness measures are defined depending on the evaluation attitude against uncertainty in Section 3.1. Numerical experiments are conducted to examine the usefulness of the robustness measures. The rules selected according to the robustness measures are compared with those selected according to the recall measure which is frequently used for selecting good rules. Our results reveal that a different aspect of the goodness of a rule is evaluated by the robustness measure, and therefore, the robustness measure acts as an independent and complementary index of the recall measure. The detailed experimental results are given in Section 3.2. Finally, in Section 3.3, the conclusions of this chapter are given.

3.1 Robustness Measure

3.1.1 The Robustness of a Decision Rule

In rough set approaches for rule induction, except the variable-precision models [76], only 100% confident rules $E \to H$ are induced, where E is minimal. Specifically, for such a rule $E \to H$, we have $conf(E \to H) = 1$ and no other rule $E' \to H$ such that $conf(E' \to H) = 1$ and $conf(E' \to E) = 1$ can in fact be induced. Therefore in the rules induced based on rough set approaches, the confidence is not effective, and support $supp(E \to H)$ and recall $rec(E \to H)$ are often used for evaluating the rules. Note that the premise E of a rule is described by the minimal number of conditional attributes and its conclusion H is described by decision attributes. The confidence and recall are often called "the accuracy and coverage" in the rough set community, respectively. The condition E of a rule $E \to H$ is composed of several elementary conditions e_i, $i = 1, 2, \ldots, p$. Specifically, it is expressed as the conjunction of elementary conditions e_i, $i = 1, 2, \ldots, p$, i.e., $E = e_1 \wedge e_2 \wedge \cdots \wedge e_p$. For simplicity, $E = e_1 \wedge e_2 \wedge \ldots \wedge e_p$ is written as $E = \{e_1, e_2, \ldots, e_p\}$. When $\mathcal{B} = b_1 \wedge b_2 \wedge \ldots \wedge b_q$ is a relaxed condition of E such that $\mathcal{B} = \{b_1, b_2, \ldots, b_q\} \subseteq E$, rule $\mathcal{B} \to H$ is called a coarser rule of $E \to H$. For convenience, we define $supp(\emptyset) = |U|$, where U is the set of objects in the given dataset.

When we evaluate a rule $E \to H$ by one of the interestingness measures described in Section 2.4, the measures do not consider the evaluations of any coarser rules $\mathcal{B} \to H$ with relaxed conditions $\mathcal{B} \subset E$. In a noise- and error-free environment, this evaluation would work extremely well. However, when the observed data includes noises and errors, evaluation of the rules without considering its coarser rules is not sufficient. For example, when an induced rule $E \to H$ shows the condition (E) of popular goods (H), a designer may design a new product that satisfies the condition $E = e_1 \wedge e_2 \wedge \ldots \wedge e_p$. If a designer's understanding of an elementary condition e_i of E is different from that of the customer, the designer may fail to sell the new product as a result of this mismatch. Moreover, when a given dataset is insufficient to express all the cases, the induced rules are imperfect. We may then come across new data that are partially matched to the induced rules. The most useful scenario is if the matched part $\mathcal{B}$ of condition E works sufficiently to conclude H, even when small errors exist.

Considering these facts, we propose robustness measures to evaluate the rules $E \to H$ induced by the rough set approaches. In these proposed robustness measures, we consider the effectiveness of the coarser rules, which are necessary to evaluate the induced rules. In previous studies [11, 38], another kind of robustness measure was defined for association rules. Whereas this robustness measure evaluates the fragility of an

TABLE 3.1: Illustrative decision table

pat	a_1	a_2	a_3	a_4	a_5	d	fr	pat	a_1	a_2	a_3	a_4	a_5	d	fr
p_1	1	1	1	0	0	1	36	p_7	0	1	1	0	1	1	30
p_2	1	1	1	0	0	0	6	p_8	0	1	1	1	1	0	6
p_3	1	1	0	1	0	0	30	p_9	0	1	1	1	0	1	12
p_4	1	0	1	0	0	0	18	p_{10}	1	1	0	0	1	0	2
p_5	0	1	1	0	1	0	4	p_{11}	0	1	1	1	0	0	5
p_6	1	0	0	1	0	0	4	p_{12}	1	0	1	0	1	0	2

TABLE 3.2: Confidence, support, and recall scores of coarser rules

name	rule	confidence	support	recall
r_1	$(a_1 = 1) \wedge (a_2 = 1) \wedge (a_3 = 1) \rightarrow (d = 1)$	0.857	36	0.461
$r_1(\overline{a_3})$	$(a_1 = 1) \wedge (a_2 = 1) \rightarrow (d = 1)$	0.486	36	0.461
$r_1(\overline{a_2})$	$(a_1 = 1) \wedge (a_3 = 1) \rightarrow (d = 1)$	0.581	36	0.461
$r_1(\overline{a_1})$	$(a_2 = 1) \wedge (a_3 = 1) \rightarrow (d = 1)$	0.788	78	1
$r_1(\overline{a_2 a_3})$	$(a_1 = 1) \rightarrow (d = 1)$	0.367	36	0.461
$r_1(\overline{a_1 a_3})$	$(a_2 = 1) \rightarrow (d = 1)$	0.595	78	1
$r_1(\overline{a_1 a_2})$	$(a_3 = 1) \rightarrow (d = 1)$	0.655	78	1
r_2	$(a_2 = 1) \wedge (a_3 = 1) \wedge (a_5 = 1) \rightarrow (d = 1)$	0.75	30	0.385
$r_2(\overline{a_5})$	$(a_2 = 1) \wedge (a_3 = 1) \rightarrow (d = 1)$	0.788	78	1
$r_2(\overline{a_3})$	$(a_2 = 1) \wedge (a_5 = 1) \rightarrow (d = 1)$	0.714	30	0.385
$r_2(\overline{a_2})$	$(a_3 = 1) \wedge (a_5 = 1) \rightarrow (d = 1)$	0.714	30	0.385
$r_2(\overline{a_3 a_5})$	$(a_2 = 1) \rightarrow (d = 1)$	0.595	78	1
$r_2(\overline{a_2 a_5})$	$(a_3 = 1) \rightarrow (d = 1)$	0.655	78	1
$r_2(\overline{a_2 a_3})$	$(a_5 = 1) \rightarrow (d = 1)$	0.682	30	0.385

association rule against noise in the given dataset, the proposed robustness measures evaluate the usefulness of a rule against partially matched data.

3.1.2 Various Robustness Measures

We propose various robustness measures to evaluate the rules induced based on rough set approaches including variable-precision rough set models [76]. To define robustness measures, we consider the effectiveness of coarser rules, which are necessary to evaluate the induced rules. The significance of the robustness measures can be demonstrated by the decision table shown in Table 3.1. In Table 3.1, a_i, $i = 1, 2, \ldots, 5$ are conditional attributes that may explain the decision attribute. The frequency (fr) of each pattern (pat) is shown in the last column.

This table shows that there exists no pattern whose decision-attribute values consistently explain $d = 1$. To explain $d = 1$, the analyst will be interested in rules having high confidence scores. The following rules may be of interest: $r_1 : (a_1 = 1) \wedge (a_2 = 1) \wedge (a_3 = 1) \rightarrow (d = 1)$ with confidence of 0.857 and $r_2 : (a_2 = 1) \wedge (a_3 = 1) \wedge (a_5 = 1) \rightarrow (d = 1)$

with confidence of 0.75. For these rules, let us consider the confidence of their coarser rules with more relaxed conditions. The scores of confidence, support, and recall of the coarser rules are shown in Table 3.2. On comparing the confidence scores of the coarser rules of r_1 and r_2 which have the same number of erased elementary conditions in Table 3.2, we observe that the coarser rules of r_2 have larger confidence scores than those of r_1, but r_2 has lower confidence value than r_1. When we consider only the confidence scores of rules, we are more interested in r_1 than in r_2. However, considering that the induced rule may be used for partially matched data, we would be more interested in r_2. As indicated by Table 3.2, such an advantage of r_2 over r_1 cannot be found, even if we add the evaluation by support and recall.

Based on these facts, we propose various robustness measures. As shown in the previous example, robustness measures are defined by using some interestingness measure $f(E \to H)$. To formulate the robustness measures, we use the following two settings:

S1 The larger $f(\mathcal{B} \to H)$ is, the more interesting rule $\mathcal{B} \to H$ is (gain property)

S2 $f(\mathcal{B} \to H) \leq f(E \to H)$, $\forall \mathcal{B} \subseteq E$ (monotonicity)

S1 means that f shows favorability rather than inadvisability. S2 means that the interestingness does not increase as a result of the relaxation of the condition of rule $E \to H$. Interestingness measures such as support, recall, and relative risk cannot satisfy S2. Instead, interestingness measures such as confidence, and lift ($lift(E \to H) = |U| conf(E \to H) = supp(H)$) may satisfy S2. Moreover, S2 implies that only the minimal rule $E \to H$ is considered.

When we use an adequate interestingness measure f, the preserved f-value from the lack of all conditional attributes in L under rule $E \to H$ is defined by

$$pres_f(E \to H; L) = f(E_{-L} \to H), \tag{3.1.1}$$

where $L \subseteq C$ and C is the set of all conditional attributes. E_{-L} is the subset of the elementary conditions of E, which does not specify the value of the conditional attribute in L. Specifically, consider $E = e_1 \wedge e_2 \ldots \wedge e_p = \bigwedge_{i=1,2,\ldots,p} e_i$ whose elementary condition e_i is represented by $(a_i = \bar{v}_i)$ with $\bar{v}_i \in V_i$ $(i = 1, 2, \ldots, p)$, where V_i is the set of values taken by attribute a_i. We then have $E_L = \bigwedge_{i:\ a_i \in \{a_1, a_2, \ldots, a_p\} \setminus L} e_i$, where $A \setminus B = \{a \in A \mid a \notin B\}$ (difference set).

We note that when $E_{-L} = \emptyset$, $pres_f$ does not always take the minimum. Moreover, for some L such that $E_{-L} \neq E$, we may have $pres_f(E \to H; L) = f(E_{-L} \to H) < f(\emptyset \to H)$. This inequality implies that the rule $E_{-L} \to H$ is less interesting than the rule that

says "everything satisfies H" with respect to the interestingness measure of f. If this inequality holds for some L, rule $E_{-L} \to H$ may weaken the extent of reasoning the conclusion H. For example, in Table 1, the confidence of $r_1(\overline{a2a3})$, or 0.367, is less than the probability of $(d = 1)$, or 0.503. We then define the following aspects of robustness of rule $E \to H$.

Genuine Robustness: Rule $E \to H$ is said to be *robust* regarding f if and only if

$$pres_f(E \to H; L) \geq f(\emptyset \to H),\ \forall L \subseteq C. \tag{3.1.2}$$

In particular, this rule is said to be strongly robust regarding f if and only if Equation (3.1.2) is satisfied with strong inequality.

ϵ-Robustness: Rule $E \to H$ is said to be *ϵ-robust* if and only if

$$pres_f(E \to H; L) \geq f(\emptyset \to H) - \epsilon,\ \forall L \subseteq C \text{ and } \epsilon \geq 0. \tag{3.1.3}$$

kth-order Robustness: Rule $E \to H$ is said to be *kth-order robust* if and only if

$$pres_f(E \to H; L) \geq f(\emptyset \to H), \forall L \subseteq C \text{ such that } |L| \leq k, k \geq 0. \tag{3.1.4}$$

kth-order ϵ-Robustness: Rule $E \to H$ is said to be *kth-order ϵ-robust* if and only if

$$pres_f(E \to H; L) \geq f(\emptyset \to H) - \epsilon, \forall L \subseteq C \text{ such that} |L| \leq k, \epsilon \text{ and } k \geq 0. \tag{3.1.5}$$

Expectation-based Robustness: Rule $E \to H$ is said to be *expectantly robust* if and only if

$$Ex(pres_f(E \to H; L)) = \sum_{L \subseteq C} P(L) pres_f(E \to H; L) \geq f(\emptyset \to H), \tag{3.1.6}$$

where $Ex(\cdot)$ is an expectation operator and $P(L) \geq 0$ is the probability that the values of all conditional attributes in L are missing in the new given object. We assume $\sum_{L \subseteq C} P(L) = 1$.

α-Quantile-based Robustness: Rule $E \to H$ is said to be *100α-robust* if and only if

$$\alpha\text{-}Quantile(pres_f(E \to H; L)) \geq f(\emptyset \to H), \tag{3.1.7}$$

where $\alpha \in (0, 1]$ and α-$Quantile(\cdot)$ is the *α-quantile* operator regarding $P(L)$ that was previously described.

The concepts of ϵ-robustness and kth-order robustness can be applied to the expectation-based and α-quantile-based robustness. For example, the kth-order expectation-based

ϵ-robust rule is a rule $E \to H$ that satisfies

$$Ex(pres_f(E \to H; L)) = \sum_{L \subseteq C} P(L||L| \leq k) pres_f(E \to H; L) \geq f(\emptyset \to H) - \epsilon, \quad (3.1.8)$$

where $P(L||L| \leq k) \geq 0$ is the conditional probability when $|L|$ is at most k and it satisfies $\sum_{L \subseteq C:|L| \leq k} P(L||L| \leq k) = 1$.

Estimating $P(L)$ is not an easy task. However, we may specify it in a manner with certain assumptions. Even if this estimation is not correct, the proposed robustness measures can evaluate the maintenance of the interestingness rule against the partially matched data. Corresponding to each aspect of robustness, we define robustness measures regarding an interestingness measure f satisfying S1 and S2 for some E. The robustness measures corresponding to the aforementioned aspects are given as follows:

Genuine Robustness Measure:

$$rbst(E \to H) = \min_{L \subseteq C} pres_f(E \to H; L) \quad (3.1.9)$$

kth-order Robustness Measure:

$$rbstk(E \to H) = \min_{L \subseteq C:|L| \leq k} pres_f(E \to H; L) \quad (3.1.10)$$

Expectation-based Robustness Measure:

$$Ex\text{-}rbst(E \to H) = \sum_{L \subseteq C} P(L) pres_f(E \to H; L) \quad (3.1.11)$$

α-Quantile-based Robustness Measure:

$$\alpha\text{-}rbst(E \to H) = \alpha\text{-}Quantile(pres_f(E \to H; L)), \ \forall L \subseteq C \quad (3.1.12)$$

We can also define the kth-order expectation-based and the kth-order α-quantile-based robustness measures that correspond to the kth-order expectation-based and the kth-order α-quantile-based robustness, respectively. On comparing the aforementioned robustness measures with $f(\emptyset \to H)$, we find the robustness and ϵ-robustness that were previously defined. Specifically, we prefer a rule with a larger robustness measure because it is more secure against information loss.

When E is a minimal condition that satisfies S2, the measures take values that are smaller than those of $f(E \to H)$. However, we can use the measures for any rule $E \to H$ including non-minimal rules. Therefore, those measures may take larger values than those of $f(E \to H)$. For example, when $E \to H$ is a minimal rule with 100%

confidence, a 100% confident rule $E \cup K \to H$ never takes a smaller value of robustness measure than that of $E \to H$. In this sense, the robustness criterion contrasts with the minimal description criterion as well as with the recall criterion.

3.2 Experiments

3.2.1 Outline

We examine the usefulness of robustness measures in constructing effective rule-based classifiers and their advantage over the recall measure. We use the expectation-based robustness measure and a first-order robustness measure in these experiments. The recall measure is considered the most rational for selecting effective rules from the rules induced by MLEM2 [25]. This is because the confidence scores of those rules take the largest value of 1 (see, for example, [42]). To this end, we conduct four numerical experiments. In these experiments, we use rules induced by MLEM2. This set of induced rules is a minimal set of minimal rules with 100% confidence. In all the experiments, we perform a 10-fold cross validation method [34] repeated 10 times. At each validation stage, we first induce a set of rules by means of MLEM2 algorithm from the training data, which consists of a minimal set of minimal rules with 100% confidence. A set of selected rules is then evaluated by the test data. In the last two experiments, we restrict ourselves to the evaluations of induced rules which matched certain test data. This is necessary because we cannot evaluate unmatched rules by using interesting measures.

The eight datasets listed in Table 3.3 are used for our experiments. They are obtained from UCI Machine Learning Repository [71]. In Table 3.3, $|U|$, $|C|$, and $|V_d|$ represent the number of objects, conditional attributes, and decision classes, respectively. N_{rule} is the number of rules induced by MLEM2. N_{match} denotes the number of induced rules which matched certain test data. *rbst*1 and *Ex-rbst* show the average scores of the first-order and expectation-based robustness measures in the training data, respectively, where a probability distribution $P(L) = (1-0.01)^{|C|-|L|}(0.01)^{|L|}$ is used for *Ex-rbst*. In addition, *rec* shows the average score of the recall measure. *Cor*-1 shows the correlation coefficient between *rbst*1 and *rec*, and *Cor-Ex* shows the correlation coefficient between *rbst*1 and *rec*. Since the absolute values of the correlation coefficients are less than 0.2 except for datasets *iris*, *soybean*, and *zoo*, no strong correlation exists between robustness and recall scores.

In the first two experiments, we select βN_{rule} rules from N_{rule} induced rules based on the robustness and recall measures. However, in the last two experiments, we select

TABLE 3.3: Eight datasets

dataset	$\|U\|$	$\|C\|$	$\|V_d\|$	N_{rule}	N_{match}	$rbst1$	$Ex\text{-}rbst$	rec	$Cor\text{-}1$	$Cor\text{-}Ex$
car	1728	6	4	57.12	49.52	0.526	0.983	0.143	-0.176	-0.030
ecoli	336	7	8	34.38	21.34	0.319	0.985	0.297	-0.034	0.079
glass	214	9	7	21.96	15.13	0.374	0.985	0.261	-0.178	-0.130
iris	150	4	3	7.6	5.1	0.454	0.991	0.476	0.605	0.716
nursery	12960	8	5	533.43	389.46	0.487	0.983	0.014	-0.084	-0.047
soybean	562	35	15	55.74	33.33	0.277	0.977	0.303	-0.374	0.604
wine	178	13	3	4.58	4.26	0.607	0.993	0.768	0.192	0.123
zoo	101	16	7	9.63	5.04	0.190	0.985	0.780	0.442	0.086

βN_{match} rules from N_{match} induced rules which matched certain test data. Using the selected rules, we construct a rule-based classifier based on LERS [22].

3.2.2 Classification Accuracy of New Objects

In the first experiment, we prepare three sets of βN_{rule} rules: the first set is composed of rules that ranked in the top βN_{rule} for robustness score; the second set is composed of rules that ranked in the top βN_{rule} for recall score; the third set is composed of randomly selected βN_{rule} rules, where βN_{rule} is rounded to the nearest integer. We compare those three sets of rules in terms of the average classification accuracies regarding test data. For β, we select 10%, 30%, 50%, 70%, and 90% as the average classification accuracies for five different experiments. The results are shown in Table 3.4. Each entry in these tables represents the average ave and the standard deviation std in the form of $ave_{\pm std}$. The single and double asterisks ("$*$" and "$**$") mean that the average classification accuracy score of the rule-based classifier with the top βN_{rule} rules selected according to the robustness scores is considerably different from that of the rule-based classifier with the top βN_{rule} rules selected according to the a recall score obtained by the paired t-test with significance level $\alpha = 0.05$ and 0.01, respectively. Similarly, the single and double stars ("$\star$" and "$\star\star$") mean that the average classification accuracy score of the rule-based classifier with the top βN_{rule} rules selected according to the robustness scores is considerably different from that of the rule-based classifier with randomly chosen βN_{rule} rules obtained by the paired t-test with significance level $\alpha = 0.05$ and 0.01, respectively.

TABLE 3.4: Classification accuracy scores for new objects

β	car				ecoli			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.970_{\pm 0.013}{}^{}_{**}$	$0.974_{\pm 0.012}{}^{\star}_{**}$	$0.985_{\pm 0.010}$	$0.964_{\pm 0.040}$	$0.776_{\pm 0.071}$	$0.783_{\pm 0.068}{}^{\star}$	$0.783_{\pm 0.068}$	$0.762_{\pm 0.070}$
70%	$0.916_{\pm 0.020}{}^{}_{**}$	$0.932_{\pm 0.020}{}^{\star\star}_{**}$	$0.963_{\pm 0.015}$	$0.904_{\pm 0.067}$	$0.750_{\pm 0.074}{}^{\star\star}_{**}$	$0.771_{\pm 0.068}{}^{\star\star}$	$0.779_{\pm 0.065}$	$0.711_{\pm 0.084}$
50%	$0.870_{\pm 0.024}{}^{\star\star}_{**}$	$0.863_{\pm 0.027}{}^{\star\star}_{**}$	$0.928_{\pm 0.023}$	$0.814_{\pm 0.115}$	$0.673_{\pm 0.084}{}^{\star}_{**}$	$0.734_{\pm 0.078}{}^{\star}$	$0.751_{\pm 0.066}$	$0.647_{\pm 0.102}$
30%	$0.749_{\pm 0.036}$	$0.755_{\pm 0.035}{}^{}_{**}$	$0.831_{\pm 0.063}$	$0.755_{\pm 0.104}$	$0.580_{\pm 0.083}{}^{\star\star}_{**}$	$0.636_{\pm 0.092}{}^{\star\star}_{**}$	$0.692_{\pm 0.078}$	$0.533_{\pm 0.129}$
10%	$0.700_{\pm 0.034}{}^{\star\star}_{**}$	$0.700_{\pm 0.034}{}^{\star\star}_{**}$	$0.593_{\pm 0.035}$	$0.543_{\pm 0.224}$	$0.513_{\pm 0.100}{}^{\star\star}_{**}$	$0.447_{\pm 0.093}{}^{\star\star}_{**}$	$0.093_{\pm 0.096}$	$0.340_{\pm 0.181}$
β	glass				iris			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.650_{\pm 0.116}$	$0.668_{\pm 0.109}$	$0.679_{\pm 0.107}$	$0.658_{\pm 0.111}$	$0.930_{\pm 0.066}$	$0.930_{\pm 0.066}$	$0.930_{\pm 0.066}$	$0.930_{\pm 0.066}$
70%	$0.559_{\pm 0.119}{}^{}_{**}$	$0.615_{\pm 0.121}{}^{}_{**}$	$0.659_{\pm 0.114}$	$0.587_{\pm 0.121}$	$0.891_{\pm 0.142}{}^{\star\star}_{*}$	$0.929_{\pm 0.142}{}^{\star\star}_{**}$	$0.929_{\pm 0.066}$	$0.759_{\pm 0.190}$
50%	$0.517_{\pm 0.126}{}^{}_{**}$	$0.570_{\pm 0.128}{}^{\star\star}_{**}$	$0.617_{\pm 0.106}$	$0.507_{\pm 0.124}$	$0.650_{\pm 0.167}{}^{\star\star}_{**}$	$0.866_{\pm 0.167}{}^{\star\star}_{**}$	$0.926_{\pm 0.068}$	$0.615_{\pm 0.222}$
30%	$0.461_{\pm 0.127}{}^{\star}_{**}$	$0.479_{\pm 0.121}{}^{\star\star}_{**}$	$0.563_{\pm 0.117}$	$0.421_{\pm 0.127}$	$0.593_{\pm 0.129}{}^{\star}_{**}$	$0.699_{\pm 0.129}{}^{\star\star}_{**}$	$0.905_{\pm 0.103}$	$0.533_{\pm 0.206}$
10%	$0.371_{\pm 0.118}{}^{\star\star}_{**}$	$0.393_{\pm 0.131}{}^{\star\star}_{**}$	$0.204_{\pm 0.086}$	$0.323_{\pm 0.124}$	$0.621_{\pm 0.165}{}^{\star\star}_{**}$	$0.645_{\pm 0.165}{}^{\star\star}_{**}$	$0.333_{\pm 0.117}$	$0.431_{\pm 0.188}$
β	nursery				soybean			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.907_{\pm 0.132}{}^{}_{**}$	$0.977_{\pm 0.004}$	$0.976_{\pm 0.004}$	$0.937_{\pm 0.101}$	$0.753_{\pm 0.055}{}^{\star\star}_{**}$	$0.871_{\pm 0.043}{}^{\star\star}$	$0.873_{\pm 0.042}$	$0.825_{\pm 0.059}$
70%	$0.614_{\pm 0.015}{}^{\star\star}_{**}$	$0.946_{\pm 0.007}{}^{\star\star}_{**}$	$0.955_{\pm 0.005}$	$0.821_{\pm 0.154}$	$0.580_{\pm 0.060}{}^{\star\star}_{**}$	$0.862_{\pm 0.042}{}^{\star\star}$	$0.866_{\pm 0.041}$	$0.691_{\pm 0.094}$
50%	$0.493_{\pm 0.014}{}^{\star\star}_{**}$	$0.879_{\pm 0.011}{}^{\star\star}_{**}$	$0.934_{\pm 0.006}$	$0.697_{\pm 0.171}$	$0.422_{\pm 0.075}{}^{\star\star}_{**}$	$0.835_{\pm 0.048}{}^{\star\star}_{**}$	$0.859_{\pm 0.043}$	$0.527_{\pm 0.104}$
30%	$0.407_{\pm 0.014}{}^{\star\star}_{**}$	$0.737_{\pm 0.017}{}^{\star\star}_{**}$	$0.908_{\pm 0.007}$	$0.600_{\pm 0.164}$	$0.307_{\pm 0.082}{}^{\star\star}_{**}$	$0.684_{\pm 0.065}{}^{\star\star}_{**}$	$0.734_{\pm 0.068}$	$0.407_{\pm 0.093}$
10%	$0.354_{\pm 0.022}{}^{\star\star}_{**}$	$0.650_{\pm 0.013}{}^{\star\star}_{**}$	$0.782_{\pm 0.012}$	$0.449_{\pm 0.120}$	$0.223_{\pm 0.070}{}^{}_{*}$	$0.399_{\pm 0.074}{}^{\star\star}_{**}$	$0.243_{\pm 0.053}$	$0.220_{\pm 0.072}$
β	wine				zoo			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.925_{\pm 0.064}$	$0.925_{\pm 0.064}$	$0.925_{\pm 0.064}$	$0.925_{\pm 0.064}$	$0.960_{\pm 0.061}{}^{\star\star}$	$0.944_{\pm 0.068}{}^{\star\star}$	$0.958_{\pm 0.062}$	$0.907_{\pm 0.102}$
70%	$0.764_{\pm 0.156}{}^{}_{**}$	$0.695_{\pm 0.148}{}^{\star\star}_{**}$	$0.916_{\pm 0.068}$	$0.777_{\pm 0.152}$	$0.894_{\pm 0.096}{}^{\star\star}_{**}$	$0.908_{\pm 0.089}{}^{\star\star}$	$0.929_{\pm 0.092}$	$0.675_{\pm 0.224}$
50%	$0.640_{\pm 0.150}{}^{}_{**}$	$0.455_{\pm 0.167}{}^{\star\star}_{**}$	$0.756_{\pm 0.197}$	$0.615_{\pm 0.166}$	$0.750_{\pm 0.158}{}^{\star\star}_{**}$	$0.854_{\pm 0.130}{}^{\star\star}$	$0.839_{\pm 0.110}$	$0.497_{\pm 0.228}$
30%	$0.512_{\pm 0.189}{}^{}_{**}$	$0.418_{\pm 0.130}{}^{\star\star}_{**}$	$0.581_{\pm 0.113}$	$0.505_{\pm 0.146}$	$0.564_{\pm 0.142}{}^{\star\star}_{**}$	$0.672_{\pm 0.143}{}^{\star\star}$	$0.697_{\pm 0.127}$	$0.323_{\pm 0.221}$
10%	$0.384_{\pm 0.114}{}^{\star}_{**}$	$0.402_{\pm 0.116}{}^{\star\star}_{**}$	$0.283_{\pm 0.104}$	$0.351_{\pm 0.111}$	$0.405_{\pm 0.139}{}^{\star}$	$0.405_{\pm 0.139}{}^{\star}$	$0.405_{\pm 0.139}$	$0.335_{\pm 0.256}$

We can now examine the usefulness of the robustness measures. First, we compare the values in columns *rbst*1 and *Ex-rbst* with those in columns *rec* and "Random". Except for cases in which β is small, the rule-based classifiers with the top βN_{rule} rules in recall scores perform better than the rule-based classifier with the top βN_{rule} rules in robustness score. However, when β is small, the rule-based classifiers with the top βN_{rule} rules in robustness score perform better than the other rule-based classifiers including those with randomly chosen rules. This suggests that when few rules are selected, the robustness measure works better than the recall measure does. For example, for design knowledge we cannot select many rules and thus, the robustness measure is useful.

The rule-based classifiers with the top βN_{rule} rules in expectation-based robustness scores perform better than that with the randomly chosen βN_{rule} rules, except for the *wine* dataset. Furthermore, the rule-based classifiers with the top βN_{rule} rules in expectation-based robustness scores perform better than that with the top βN_{rule} rules in the first-order robustness scores.

3.2.3 Classification Accuracy of New Objects with Missing Values

In the previous experiment, the rule-based classifiers with rules selected according to the robustness measures did not always perform well. The rule-based classifier with rules selected according to the recall measure worked better when a sufficient number of rules were selected. The robustness measure evaluates to what extent the interestingness of a rule is preserved against partial matched objects. Specifically, rules with high robustness scores may be effective for classifying objects whose attribute values are only partially known. Therefore, we examine the classification accuracy of rule-based classifiers with rules selected according to the robustness measures. To this end, we compare the rule-based classifiers with rules selected according to the robustness scores to those with rules selected according to the recall scores and at random in terms of the average classification accuracies for new objects with missing values.

To produce a set of new objects with missing values, we delete all the data of a conditional attributes from the test data during the 10-fold cross validation. The deletion is applied to each conditional attribute to obtain $|C|Ch$ test data with missing values, where Ch is the amount of test data. The experimental methods are the same as that in the previous experiment, but we use the test data with missing values at the validation stage. We select 10%, 30%, 50%, 70%, and 90% for β as the average classification accuracy for five different experiments..

TABLE 3.5: Classification accuracy scores for new data with missing values

β	car				ecoli			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.794_{\pm 0.020}$	$0.796_{\pm 0.020}$	$0.796_{\pm 0.020}$	$0.790_{\pm 0.025}$	$0.679_{\pm 0.060}$	$0.682_{\pm 0.060}$	$0.683_{\pm 0.059}$	$0.672_{\pm 0.063}$
70%	$0.793_{\pm 0.023}^{\star\star}{}_{*}$	$0.794_{\pm 0.023}^{\star\star}{}_{**}$	$0.786_{\pm 0.022}$	$0.761_{\pm 0.036}$	$0.659_{\pm 0.064}^{\star}{}_{*}$	$0.673_{\pm 0.061}^{\star\star}$	$0.680_{\pm 0.059}$	$0.635_{\pm 0.076}$
50%	$0.771_{\pm 0.026}^{\star\star}$	$0.773_{\pm 0.026}^{\star\star}$	$0.777_{\pm 0.024}$	$0.717_{\pm 0.078}$	$0.591_{\pm 0.079}{}_{**}$	$0.646_{\pm 0.071}^{\star\star}$	$0.661_{\pm 0.064}$	$0.571_{\pm 0.101}$
30%	$0.711_{\pm 0.033}^{\star\star}{}_{**}$	$0.717_{\pm 0.033}^{\star\star}{}_{**}$	$0.735_{\pm 0.038}$	$0.670_{\pm 0.122}$	$0.518_{\pm 0.082}^{\star}{}_{**}$	$0.567_{\pm 0.083}^{\star\star}$	$0.614_{\pm 0.075}$	$0.483_{\pm 0.140}$
10%	$0.700_{\pm 0.034}^{\star\star}{}_{**}$	$0.700_{\pm 0.034}^{\star\star}{}_{**}$	$0.629_{\pm 0.032}$	$0.512_{\pm 0.200}$	$0.473_{\pm 0.092}^{\star\star}{}_{**}$	$0.439_{\pm 0.092}^{\star\star}{}_{**}$	$0.089_{\pm 0.099}$	$0.322_{\pm 0.172}$
β	glass				iris			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.567_{\pm 0.089}$	$0.582_{\pm 0.082}$	$0.589_{\pm 0.082}$	$0.573_{\pm 0.091}$	$0.764_{\pm 0.066}$	$0.764_{\pm 0.066}$	$0.764_{\pm 0.066}$	$0.764_{\pm 0.066}$
70%	$0.508_{\pm 0.094}^{\star\star}{}_{**}$	$0.547_{\pm 0.096}$	$0.572_{\pm 0.088}$	$0.523_{\pm 0.088}$	$0.724_{\pm 0.116}^{\star\star}{}_{**}$	$0.758_{\pm 0.068}^{\star\star}$	$0.764_{\pm 0.066}$	$0.621_{\pm 0.190}$
50%	$0.480_{\pm 0.103}^{\star\star}{}_{**}$	$0.513_{\pm 0.106}^{\star\star}{}_{*}$	$0.541_{\pm 0.085}$	$0.460_{\pm 0.104}$	$0.546_{\pm 0.130}{}_{**}$	$0.705_{\pm 0.121}^{\star\star}{}_{**}$	$0.748_{\pm 0.068}$	$0.529_{\pm 0.222}$
30%	$0.436_{\pm 0.109}^{\star\star}{}_{**}$	$0.451_{\pm 0.102}^{\star\star}{}_{**}$	$0.494_{\pm 0.099}$	$0.382_{\pm 0.112}$	$0.495_{\pm 0.124}{}_{**}$	$0.581_{\pm 0.127}^{\star\star}{}_{**}$	$0.702_{\pm 0.103}$	$0.479_{\pm 0.206}$
10%	$0.363_{\pm 0.108}^{\star\star}{}_{**}$	$0.379_{\pm 0.118}^{\star\star}{}_{**}$	$0.188_{\pm 0.077}$	$0.304_{\pm 0.124}$	$0.612_{\pm 0.162}^{\star\star}{}_{**}$	$0.635_{\pm 0.139}^{\star\star}{}_{**}$	$0.333_{\pm 0.117}$	$0.412_{\pm 0.188}$
β	nursery				soybean			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.759_{\pm 0.115}{}_{**}$	$0.816_{\pm 0.006}{}_{**}$	$0.809_{\pm 0.007}$	$0.767_{\pm 0.099}$	$0.717_{\pm 0.050}^{\star\star}{}_{**}$	$0.834_{\pm 0.039}^{\star\star}$	$0.835_{\pm 0.037}$	$0.781_{\pm 0.063}$
70%	$0.510_{\pm 0.014}^{\star\star}{}_{**}$	$0.801_{\pm 0.008}^{\star\star}{}_{**}$	$0.791_{\pm 0.007}$	$0.671_{\pm 0.141}$	$0.553_{\pm 0.056}^{\star\star}{}_{**}$	$0.824_{\pm 0.039}^{\star\star}$	$0.828_{\pm 0.038}$	$0.653_{\pm 0.082}$
50%	$0.420_{\pm 0.013}^{\star\star}{}_{**}$	$0.747_{\pm 0.011}^{\star\star}{}_{**}$	$0.773_{\pm 0.008}$	$0.609_{\pm 0.151}$	$0.405_{\pm 0.071}^{\star\star}{}_{**}$	$0.795_{\pm 0.045}^{\star\star}{}_{**}$	$0.818_{\pm 0.040}$	$0.524_{\pm 0.094}$
30%	$0.364_{\pm 0.013}^{\star\star}{}_{**}$	$0.656_{\pm 0.015}^{\star\star}{}_{**}$	$0.762_{\pm 0.008}$	$0.535_{\pm 0.150}$	$0.296_{\pm 0.079}^{\star\star}{}_{**}$	$0.658_{\pm 0.061}^{\star\star}{}_{**}$	$0.702_{\pm 0.063}$	$0.389_{\pm 0.089}$
10%	$0.347_{\pm 0.023}^{\star\star}{}_{**}$	$0.606_{\pm 0.012}^{\star\star}{}_{**}$	$0.697_{\pm 0.011}$	$0.431_{\pm 0.100}$	$0.219_{\pm 0.068}{}_{*}$	$0.388_{\pm 0.071}^{\star\star}{}_{**}$	$0.236_{\pm 0.052}$	$0.203_{\pm 0.081}$
β	wine				zoo			
	*rbst*1	*Ex-rbst*	*rec*	Random	*rbst*1	*Ex-rbst*	*rec*	Random
90%	$0.784_{\pm 0.063}$	$0.784_{\pm 0.063}$	$0.784_{\pm 0.063}$	$0.784_{\pm 0.063}$	$0.877_{\pm 0.056}^{\star\star}$	$0.866_{\pm 0.064}^{\star\star}$	$0.875_{\pm 0.056}$	$0.819_{\pm 0.116}$
70%	$0.688_{\pm 0.125}{}_{**}$	$0.638_{\pm 0.131}^{\star}{}_{**}$	$0.797_{\pm 0.061}$	$0.681_{\pm 0.134}$	$0.825_{\pm 0.089}^{\star\star}$	$0.847_{\pm 0.082}^{\star\star}$	$0.849_{\pm 0.081}$	$0.650_{\pm 0.197}$
50%	$0.611_{\pm 0.143}{}_{*}$	$0.446_{\pm 0.155}^{\star\star}{}_{**}$	$0.666_{\pm 0.176}$	$0.574_{\pm 0.148}$	$0.711_{\pm 0.152}^{\star\star}{}_{**}$	$0.803_{\pm 0.118}^{\star\star}$	$0.777_{\pm 0.100}$	$0.463_{\pm 0.244}$
30%	$0.504_{\pm 0.181}$	$0.415_{\pm 0.125}^{\star\star}{}_{**}$	$0.507_{\pm 0.099}$	$0.520_{\pm 0.150}$	$0.554_{\pm 0.139}^{\star\star}{}_{**}$	$0.655_{\pm 0.141}^{\star\star}$	$0.679_{\pm 0.127}$	$0.318_{\pm 0.199}$
10%	$0.386_{\pm 0.113}{}_{**}$	$0.404_{\pm 0.114}{}_{**}$	$0.283_{\pm 0.104}$	$0.371_{\pm 0.128}$	$0.405_{\pm 0.139}^{\star\star}$	$0.405_{\pm 0.139}^{\star\star}$	$0.405_{\pm 0.139}$	$0.285_{\pm 0.242}$

The results are shown in Table 3.5. Table 3.5 employs the same format as that of Table 3.4. Although the robustness measure evaluates to what extent the interestingness of a rule is preserved against partial matched objects, the results are similar to those in the previous experiment, although all the accuracies decrease. Specifically, we observe that: 1) the rule-based classifiers with the top βN_{rule} rules in recall score perform better than others rule-based classifiers except for cases in which β is small, 2) except for datasets *nursery* and *soybean*, the rule-based classifiers with the top βN_{rule} rules in robustness scores perform better than those with the top βN_{rule} rules in recall score when β is small, and 3) the rule-based classifiers with the top βN_{rule} rules in robustness scores perform better than that with randomly chosen βN_{rule} rules except for datasets *nursery* and *soybean*.

From the results of the first and second experiments, we can conclude that to construct a rule-based classifier with several rules, the recall score works better for selecting effective rules than the robustness score does. However, rules with extremely high robustness scores work extremely well for class estimation and may be useful for design knowledge. Furthermore, the rule-based classifiers with the top βN_{rule} rules in expectation-based robustness score perform better than that with the top βN_{rule} rules in first-order robustness score.

3.2.4 Robustness Scores in Test Data

The previous experiments showed that the selection of rules according to the recall score works better than that according to the robustness score. However, robustness scores may select rules whose confidence scores have not considerably worsened for partially matched data. In the third experiment, we confirm whether rules having high robustness scores in the training data also maintained high robustness scores in the test data.

To confirm this, we calculate the average robustness scores in the test data for the top βN_{match} rules selected according to the robustness score, for the top βN_{match} rules selected according to the recall score, and for randomly chosen βN_{match}. The results are shown in Tables 3.6 and 3.7. In Table 3.6, the first-order robustness measure is used to obtain robustness scores whereas in Table 3.7, the expectation-based robustness measure is used. In these tables, single and double asterisks ("$*$" and "$**$") mean that the average robustness score of the top βN_{rule} rules in robustness scores is considerably different from that of the top βN_{rule} rules in recall score obtained by the paired t-test with significance level $\alpha = 0.05$ and 0.01, respectively.

As shown in Table 3.6, top βN_{match} rules in the first-order robustness score are better than the chosen βN_{match} rules in the recall score. In the same manner as that of the

TABLE 3.6: Average of first-order robustness scores of selected rules

β	car			ecoli		
	*rbst*1	*rec*	Random	*rbst*1	*rec*	Random
50%	$0.648_{\pm 0.207}^{\star\star}$	$0.484_{\pm 0.224}$	$0.369_{\pm 0.294}$	$0.472_{\pm 0.258}^{\star\star}$	$0.387_{\pm 0.266}$	$0.385_{\pm 0.277}$
30%	$0.757_{\pm 0.163}^{\star\star}$	$0.441_{\pm 0.212}$	$0.363_{\pm 0.293}$	$0.569_{\pm 0.232}^{\star\star}$	$0.388_{\pm 0.281}$	$0.378_{\pm 0.266}$
10%	$0.857_{\pm 0.069}^{\star\star}$	$0.445_{\pm 0.214}$	$0.379_{\pm 0.303}$	$0.681_{\pm 0.181}^{\star\star}$	$0.275_{\pm 0.280}$	$0.397_{\pm 0.299}$
β	glass			iris		
	*rbst*1	*rec*	Random	*rbst*1	*rec*	Random
50%	$0.402_{\pm 0.243}^{\star\star}$	$0.319_{\pm 0.227}$	$0.314_{\pm 0.247}$	$0.672_{\pm 0.279}{}_{\star\star}$	$0.576_{\pm 0.267}$	$0.429_{\pm 0.245}$
30%	$0.418_{\pm 0.246}^{\star\star}$	$0.299_{\pm 0.222}$	$0.330_{\pm 0.244}$	$0.736_{\pm 0.249}{}_{\star\star}$	$0.507_{\pm 0.229}$	$0.429_{\pm 0.254}$
10%	$0.461_{\pm 0.250}^{\star\star}$	$0.255_{\pm 0.173}$	$0.314_{\pm 0.243}$	$0.879_{\pm 0.177}^{\star\star}$	$0.333_{\pm 0.117}$	$0.396_{\pm 0.239}$
β	nursery			soybean		
	*rbst*1	*rec*	Random	*rbst*1	*rec*	Random
50%	$0.545_{\pm 0.132}^{\star\star}$	$0.500_{\pm 0.159}$	$0.486_{\pm 0.266}$	$0.330_{\pm 0.222}^{\star\star}$	$0.223_{\pm 0.194}$	$0.239_{\pm 0.211}$
30%	$0.569_{\pm 0.126}{}^{\star\star}$	$0.518_{\pm 0.125}$	$0.484_{\pm 0.256}$	$0.365_{\pm 0.238}^{\star\star}$	$0.154_{\pm 0.148}$	$0.243_{\pm 0.215}$
10%	$0.563_{\pm 0.113}{}^{\star\star}$	$0.507_{\pm 0.134}$	$0.463_{\pm 0.264}$	$0.426_{\pm 0.244}^{\star\star}$	$0.073_{\pm 0.095}$	$0.235_{\pm 0.210}$
β	wine			zoo		
	*rbst*1	*rec*	Random	*rbst*1	*rec*	Random
50%	$0.656_{\pm 0.174}^{\star\star}$	$0.527_{\pm 0.194}$	$0.481_{\pm 0.220}$	$0.347_{\pm 0.186}{}_{\star\star}$	$0.292_{\pm 0.153}$	$0.318_{\pm 0.181}$
30%	$0.663_{\pm 0.171}^{\star\star}$	$0.498_{\pm 0.183}$	$0.477_{\pm 0.205}$	$0.386_{\pm 0.182}^{\star\star}$	$0.311_{\pm 0.156}$	$0.311_{\pm 0.186}$
10%	$0.718_{\pm 0.165}^{\star\star}$	$0.509_{\pm 0.184}$	$0.491_{\pm 0.205}$	$0.405_{\pm 0.139}^{\star\star}$	$0.405_{\pm 0.139}$	$0.988_{\pm 0.020}$

first-order robustness scores, the expectation-based robustness scores of the top βN_{match} rules in the expectation-based robustness score is better than those of the chosen βN_{match} rules in the recall score except for the case of $\beta = 10\%$ in datasets *glass* and *soybean*.

The tables show that the average values gradually decrease as β increases in *rbst*1 and *Ex-rbst*. This implies that rules with high robustness scores in the training data can also have high robustness scores in the test data. This also implies that robustness measures work better for selecting effective rules in the sense that the confidence scores are maintained even when the conditions do not match. Specifically, robustness measures can be useful for selecting design knowledge. However, rules having high recall scores in the training data do not always have high first-order robustness scores in the test data even though the rule-based classifier system employed for them performs well.

3.2.5 Support and Pseudo-support

To determine the differences between robustness and recall measures, we further investigate the support and pseudo-support scores in the fourth experiment. The pseudo-support score in this study is defined by the number of objects that match the rule when a single condition mismatch (at most) exists. In this experiment, we evaluate the average support scores in the training data and average pseudo-support scores in the test data. We compare the top βN_{rule} rules in the robustness score and top βN_{rule}

TABLE 3.7: Average of expectation-based robustness scores of selected rules

β	car			ecoli		
	Ex-rbst	*rec*	Random	*Ex-rbst*	*rec*	Random
50%	$0.990_{\pm 0.008}^{\star\star}$	$0.994_{\pm 0.006}$	$0.976_{\pm 0.021}$	$0.986_{\pm 0.015}^{\star\star}$	$0.981_{\pm 0.019}$	$0.985_{\pm 0.019}$
30%	$0.984_{\pm 0.006}^{\star\star}$	$0.983_{\pm 0.010}$	$0.976_{\pm 0.021}$	$0.989_{\pm 0.012}^{\star\star}$	$0.980_{\pm 0.020}$	$0.984_{\pm 0.019}$
10%	$0.997_{\pm 0.001}^{\star\star}$	$0.984_{\pm 0.011}$	$0.976_{\pm 0.021}$	$0.994_{\pm 0.007}^{\star\star}$	$0.972_{\pm 0.025}$	$0.984_{\pm 0.020}$
β	glass			iris		
	Ex-rbst	*rec*	Random	*Ex-rbst*	*rec*	Random
50%	$0.971_{\pm 0.028}^{\star\star}$	$0.969_{\pm 0.028}$	$0.966_{\pm 0.031}$	$0.995_{\pm 0.006}$	$0.994_{\pm 0.005}$	$0.990_{\pm 0.009}$
30%	$0.973_{\pm 0.027}^{\star\star}$	$0.970_{\pm 0.028}$	$0.966_{\pm 0.030}$	$0.996_{\pm 0.006}{}_{\star\star}$	$0.994_{\pm 0.004}$	$0.990_{\pm 0.009}$
10%	$0.976_{\pm 0.025}^{\star\star}$	$0.979_{\pm 0.021}$	$0.965_{\pm 0.030}$	$0.998_{\pm 0.003}{}_{\star\star}$	$0.993_{\pm 0.001}$	$0.989_{\pm 0.010}$
β	nursery			soybean		
	Ex-rbst	*rec*	Random	*Ex-rbst*	*rec*	Random
50%	$0.990_{\pm 0.003}{}_{\star\star}$	$0.986_{\pm 0.013}$	$0.994_{\pm 0.004}$	$0.973_{\pm 0.046}^{\star\star}$	$0.971_{\pm 0.048}$	$0.925_{\pm 0.105}$
30%	$0.991_{\pm 0.003}^{\star}{}_{\star\star}$	$0.988_{\pm 0.008}$	$0.994_{\pm 0.003}$	$0.979_{\pm 0.032}^{\star\star}{}_{\star}$	$0.976_{\pm 0.038}$	$0.930_{\pm 0.101}$
10%	$0.992_{\pm 0.002}^{\star\star}{}_{\star\star}$	$0.988_{\pm 0.009}$	$0.994_{\pm 0.003}$	$0.982_{\pm 0.025}^{\star\star}{}_{\star}$	$0.985_{\pm 0.011}$	$0.925_{\pm 0.103}$
β	wine			zoo		
	Ex-rbst	*rec*	Random	*Ex-rbst*	*rec*	Random
50%	$0.987_{\pm 0.016}$	$0.985_{\pm 0.014}$	$0.990_{\pm 0.017}$	$0.991_{\pm 0.010}^{\star}$	$0.990_{\pm 0.014}$	$0.988_{\pm 0.021}$
30%	$0.986_{\pm 0.016}$	$0.984_{\pm 0.015}$	$0.990_{\pm 0.012}$	$0.992_{\pm 0.007}^{\star\star}$	$0.992_{\pm 0.009}$	$0.985_{\pm 0.029}$
10%	$0.989_{\pm 0.014}{}_{\star\star}$	$0.983_{\pm 0.016}$	$0.991_{\pm 0.013}$	$0.994_{\pm 0.001}^{\star\star}$	$0.994_{\pm 0.001}$	$0.988_{\pm 0.020}$

rules in the recall score in the training data. The results are shown in Table 3.8. In the table, the single and double asterisks ("$*$" and "$**$") mean that the average support score (with respect to average pseudo-support score) of the top βN_{rule} rules in the first-order robustness score and the expectation-based robustness scores is considerably different from that of the top βN_{rule} rules in the recall score obtained by the paired t-test with significance level $\alpha = 0.05$ and 0.01, respectively. As condition attributes ($|C|$) continued to appear in a dataset, the average support score and average pseudo-support score decrease. From Table 3.8, we can observe a weak tendency in which the average support and average pseudo-support scores of the rules selected according to the robustness scores were greater than those of the rules selected according to the recall scores, except for datasets *iris*, *nursery*, and *zoo*. In these datasets, the differences of average support scores and average pseudo-support scores between the top βN_{rule} rules in the robustness score and the top βN_{rule} rules in the recall score are not considerable. This suggests that rules with high robustness scores in the training data can also have large pseudo-support scores in the test data. Thus, they are useful in test data and as design knowledge.

When we combine the results obtained in the last two experiments, rules with high robustness scores work well individually. However, the rule-based classifier composed of rules with high robustness scores is not as effective as that composed of rules with

TABLE 3.8: Average supports in training data and pseudo-supports in test data

β	car (training)			car (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$97.108_{\pm 125.779}$	$97.167^{**}_{\pm 125.702}$	$84.403_{\pm 130.521}$	$52.7309_{\pm 32.890}$	$51.650^{**}_{\pm 33.244}$	$32.508_{\pm 38.280}$
30%	$131.176^{*}_{\pm 147.791}$	$136.214^{**}_{\pm 147.093}$	$99.519_{\pm 159.581}$	$71.669_{\pm 27.243}$	$71.598^{**}_{\pm 27.480}$	$32.509^{**}_{\pm 43.528}$
10%	$84.053^{**}_{\pm 28.434}$	$120.155^{**}_{\pm 124.409}$	$189.429_{\pm 232.564}$	$66.600_{\pm 7.825}$	$69.011^{**}_{\pm 13.464}$	$48.549_{\pm 56.004}$
β	ecoli (training)			ecoli (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$21.650_{\pm 21.207}$	$23.269_{\pm 20.693}$	$22.615_{\pm 21.101}$	$7.367^{**}_{\pm 4.543}$	$7.921^{**}_{\pm 4.280}$	$6.154_{\pm 4.585}$
30%	$28.461^{**}_{\pm 23.868}$	$30.380^{**}_{\pm 22.927}$	$26.369_{\pm 24.648}$	$8.619^{**}_{\pm 4.294}$	$9.431^{**}_{\pm 4.192}$	$5.835_{\pm 4.795}$
10%	$47.059^{**}_{\pm 24.391}$	$48.128^{**}_{\pm 24.547}$	$6.503_{\pm 7.621}$	$10.122^{**}_{\pm 4.276}$	$11.492^{**}_{\pm 3.322}$	$3.122_{\pm 3.637}$
β	glass (training)			glass (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$11.910^{**}_{\pm 7.197}$	$12.107^{**}_{\pm 8.082}$	$13.245_{\pm 7.235}$	$3.921_{\pm 1.676}$	$4.089_{\pm 1.869}$	$2.961_{\pm 1.977}$
30%	$14.038_{\pm 7.547}$	$13.167^{**}_{\pm 8.750}$	$14.280_{\pm 8.438}$	$3.826_{\pm 1.601}$	$4.302_{\pm 1.927}$	$2.649_{\pm 1.910}$
10%	$16.837^{**}_{\pm 8.367}$	$15.611^{*}_{\pm 10.308}$	$13.970_{\pm 6.574}$	$3.921_{\pm 1.554}$	$4.579_{\pm 1.966}$	$1.985_{\pm 1.225}$
β	iris (training)			iris (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$28.322^{**}_{\pm 17.003}$	$36.172_{\pm 12.035}$	$37.463_{\pm 10.603}$	$4.543_{\pm 1.726}$	$4.608_{\pm 1.726}$	$4.841_{\pm 1.726}$
30%	$33.086^{**}_{\pm 13.639}$	$38.700^{**}_{\pm 6.589}$	$42.045_{\pm 3.223}$	$4.547_{\pm 1.739}$	$4.542_{\pm 1.744}$	$4.909_{\pm 1.744}$
10%	$39.150^{**}_{\pm 2.311}$	$38.580^{**}_{\pm 2.349}$	$45.000_{\pm 1.752}$	$4.753_{\pm 1.614}$	$4.768_{\pm 1.650}$	$5.000_{\pm 1.650}$
β	nursey (training)			nursey (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$45.679^{**}_{\pm 52.24}$	$67.719_{\pm 240.78}$	$66.014_{\pm 241.62}$	$153.873^{**}_{\pm 31.10}$	$160.792^{**}_{\pm 35.35}$	$148.210_{\pm 59.37}$
30%	$54.154^{**}_{\pm 59.37}$	$90.586^{**}_{\pm 306.44}$	$100.621_{\pm 306.57}$	$156.781^{**}_{\pm 31.56}$	$170.439^{**}_{\pm 39.58}$	$165.571_{\pm 61.24}$
10%	$31.794^{**}_{\pm 26.53}$	$163.126^{**}_{\pm 517.84}$	$197.562_{\pm 514.91}$	$141.478^{**}_{\pm 12.04}$	$187.984^{**}_{\pm 51.29}$	$176.614_{\pm 89.29}$
β	soybean (training)			soybean (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$11.592^{**}_{\pm 11.941}$	$18.374_{\pm 11.819}$	$18.869_{\pm 11.343}$	$5.897^{**}_{\pm 2.806}$	$5.119^{**}_{\pm 3.227}$	$4.218_{\pm 3.147}$
30%	$10.967^{**}_{\pm 9.724}$	$22.605_{\pm 12.006}$	$22.109_{\pm 12.058}$	$5.854^{**}_{\pm 2.601}$	$4.570^{**}_{\pm 2.976}$	$2.978_{\pm 2.158}$
10%	$13.057^{**}_{\pm 10.416}$	$30.920^{**}_{\pm 13.933}$	$21.709_{\pm 8.123}$	$5.847^{**}_{\pm 2.181}$	$5.694^{**}_{\pm 3.026}$	$2.801_{\pm 1.693}$
β	wine (training)			wine (test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$48.422_{\pm 8.706}$	$40.419^{**}_{\pm 17.692}$	$49.446_{\pm 5.329}$	$5.839^{*}_{\pm 1.901}$	$5.928^{**}_{\pm 2.061}$	$5.403_{\pm 1.910}$
30%	$48.255_{\pm 8.573}$	$45.875^{*}_{\pm 12.515}$	$51.000_{\pm 5.271}$	$5.830^{**}_{\pm 1.933}$	$5.933^{**}_{\pm 2.087}$	$5.265_{\pm 1.899}$
10%	$49.690_{\pm 7.577}$	$51.720_{\pm 5.765}$	$51.000_{\pm 4.420}$	$6.069^{**}_{\pm 1.911}$	$6.423^{**}_{\pm 1.904}$	$4.980_{\pm 1.850}$
β	zoo(training)			zoo(test)		
	*rbst*1	*Ex-rbst*	*rec*	*rbst*1	*Ex-rbst*	*rec*
50%	$15.161_{\pm 12.024}$	$15.970_{\pm 11.441}$	$15.575_{\pm 11.798}$	$2.465_{\pm 1.623}$	$2.627_{\pm 1.592}$	$2.609_{\pm 1.605}$
30%	$18.500^{**}_{\pm 13.444}$	$21.420_{\pm 11.776}$	$22.017_{\pm 11.013}$	$2.750_{\pm 1.742}$	$3.033_{\pm 1.613}$	$3.035_{\pm 1.611}$
10%	$36.040_{\pm 5.322}$	$36.900_{\pm 1.425}$	$36.900_{\pm 1.425}$	$4.100_{\pm 1.425}$	$4.100_{\pm 1.425}$	$4.100_{\pm 1.425}$

high recall scores. This may be caused by the non-uniform distribution of rules over the conditional and decision attribute space.

3.3 Conclusions

We introduced robustness measures that evaluate to what extent the interestingness of a rule is preserved against partially matched objects. Four numerical experiments

were conducted to examine the performance of rules selected according to the robustness scores and recall scores. The results revealed that the robustness measure was an independent interestingness measure of recall and may be effective in selecting useful rules for design knowledge. Rules with extremely high robustness scores are useful for design knowledge, whereas rules with high recall scores are useful for constructing a rule-based classifier. Specifically, rules with very high robustness scores performed well in estimating the decision attribute value but a rule-based classifier composed of such rules did not.

Chapter 4

A k-Anonymous Rule Clustering for Data Publishing

In this chapter, we propose a rule clustering approach to achieving the k-anonymity for privacy protection in rule publication. When rule-based classifiers are used in real-world problems, we may be requested to publish them to ensure fairness and provide knowledge for field researchers. To develop methods and tools for publishing rules while ensuring the confidentiality of personal sensitive and private information is a task of the utmost importance. Induced rules are not always supported by a sufficient number of objects because objects with rare condition attribute values may exist in a dataset. If a rule is supported only by a few objects, some condition or decision attribute values of some particular objects can be revealed. If some of the revealed attribute values are sensitive or identify an object, it may invade the privacy of the object. This identifiability is totally undesirable, especially for sensitive information. That is, the development of techniques for privacy protection is a very important issue for published rules. To protect the privacy, achieving the k-anonymity [65], i.e., using only rules supported at least k objects is useful. Namely, we propose to merge rules with low support based on the similarity and consistency measures of rules. By this way, we try to achieve the k-anonymity while aiming to maintain the classification accuracy through the use of similarity and consistency measures. The approach is described in Section 4.1. In Section 4.2, two numerical experiments are conducted. One is to verify the accuracy of the rule-based classifier with the k-anonymized rules, and the other is to evaluate the risk of privacy invasion of decision attribute values. The effectiveness of the proposed approach are demonstrated by these experiments. Finally, conclusions are given in Section 4.3.

4.1 k-Anonymous Rule Clustering Approach

In this section, we propose a k-anonymous rule clustering algorithm. The k-anonymity for a rule is defined as follows.

Definition 4.1. Let r be an induced rule expressed as "if $f(u, a_1) \in Cond_1, f(u, a_2) \in Cond_2, \ldots, f(u, a_l) \in Cond_l$ then $u \in D_j$", where $f(u, a_i)$ is the attribute value of u in attribute a_i, $Cond_i$ is a condition about the attribute valllue $f(u, a_i)$ and D_j is a decision class. Rule r is said to satisfy k-anonymity if and only if

$$Strength(r) = |\{u \in U : u \in D_j, f(u, a_i) \in Cond_i,\ i = 1, 2, \ldots, l\}| \geq k, \qquad (4.1.1)$$

where $Strength(r)$ is the number of objects that satisfy a rule r in the decision table. Equation (4.1.1) implies that rule r satisfies k-anonymity when rule r is supported at least k objects in the decision table. Rules satisfying k-anonymities are called k-anonymous rules. In the field of frequent pattern mining, Apriori algorithm [4] and FP-growth algorithm [29] can induce rules satisfying k-anonymity by adopting the minimum support as one of rule induction criteria. However, the rules induced by those algorithms did not always cover all objects in the decision table (dataset). Therefore, those induced rules are useful for discovering knowledge but not alway for constructing a rule-based classifier. Some of rules constructing a rule-based classifier may be supported only by a few objects, i.e., their support scores are less than k. Without those rules, the classification accuracy of the rule-based classifier will be decreased. In order to satisfy the k-anonymity for all rules constructing the rule-based classifier, rules with small support scores are combined until the merged rules satisfy the k-anonymity. We note that by the combination, the confidence of rules constructing the rule-based classifier decreases. For rule combination, the following similarity measure between rule r and r' is used.

Definition 4.2. Let r and r' be rules with a same conclusion. Similarity measure between r and r' is defined by

$$s(r, r') = \frac{1}{|C|} \sum_{a \in C} s(r, r' : a), \qquad (4.1.2)$$

where

$$s(r,r':a)=\begin{cases}1, \text{ if } \rho(r,a)\cup\rho(r',a)=\emptyset\\ 0, \text{ if } (\rho(r,a)=\emptyset \text{ and } \rho(r',a)\neq\emptyset) \text{ or } (\rho(r,a)\neq\emptyset \text{ and } \rho(r',a)=\emptyset)\\ \max\{Truth(\rho(r,a)\subseteq\rho(r',a)), Truth(\rho(r',a)\subseteq\rho(r,a))\},\\ \qquad \text{if } a \text{ is nominal and } \rho(r,a)\neq\emptyset \text{ and } \rho(r',a)\neq\emptyset\\ \max\{1-\frac{|\bar{\rho}(r,a)-\bar{\rho}(r',a)|}{3\sigma_a}, 0\},\\ \qquad \text{if } a \text{ is numerical and } \rho(r,a)\neq\emptyset \text{ and } \rho(r',a)\neq\emptyset.\end{cases} \tag{4.1.3}$$

$\rho(r,a)$ denotes the set of values satisfying the condition about condition attribute a in rule r. In Equation (4.1.3), Truth(*proposition*) is 1 if *proposition* is true and 0 otherwise. When a condition attribute a is numerical, $\rho(r,a)$ become a closed interval, and σ_a and $\bar{\rho}(r,a)$ are defined as the standard deviation of attribute values of a in the decision table and the center value of closed interval $\rho(r,a)$, i.e., $\bar{\rho}(r,a) = 0.5\inf(\rho(r,a)) + 0.5\sup(\rho(r,a))$.

When two rules are combined, the resulting rule may be composed of the inconsistent conditions that do not corresponding to the original rules. Then, we define a consistency measure between two rules r and r' to reduce the number of inconsistent conditions.

Definition 4.3. Let r and r' be rules with the same conclusion. The consistency measure is defined by

$$c(r,r')=\frac{\sum_{r''\in R}\max\{Truth(\forall a\in A(r), \rho(r'',a)\subseteq\rho(r,a)),\\ \qquad Truth(\forall a\in A(r'), \rho(r'',a)\subseteq\rho(r',a))\}}{2^{|\{a\in A(r)\cap A(r') \,:\, a \text{ is numerical}\}|}\times\prod_{a\in A(r)\cap A(r') \,:\, a \text{ is nominal}}|\rho(r,a)\cup\rho(r',a)|}, \tag{4.1.4}$$

where $A(r)$ is the set of condition attributes used in the premise of rule r. The combination of rules r and r' generates the following rule:

$$\text{if } f(x,a_1)\in\rho(r,a_1)\widetilde{\cup}\rho(r',a_1),\ldots,f(x,a_l)\in\rho(r,a_p)\widetilde{\cup}\rho(r',a_p) \text{ then } x\in D_j. \tag{4.1.5}$$

$\rho(r,a_1)\widetilde{\cup}\rho(r',a_1)$ is defined by $\rho(r,a)\cup\rho(r',a)$ if a is nominal and $[\min(\rho(r,a),\rho(r',a)), \max(\rho(r,a),\rho(r',a))]$ if a is numerical. R is a set of rules whose premises are composed of $f(x,a_i) = v_{a_i}$ with $v_{a_i}\in\rho(r,a_i)\cup\rho(r',a_i)$ if a_i is nominal, and $v_{a_i} = \rho(r,a_i)$ or $v_{a_i} = \rho(r',a_i)$ if a_i is numerical. Namely, R has exactly $2^{|\{a_i\in A(r)\cap A(r') \,:\, a_i \text{ is numerical}\}|}\times\prod_{a_i\in A(r)\cap A(r') \,:\, a_i \text{ is nominal}}|\rho(r,a_i)\cup\rho(r',a_i)|$ rules. We note that if r is a merged rule of $r_1,r_2,\ldots,r_p$, $\rho(r,a_i)$ is defined by a set of $\{\rho(r_1,a_i),\rho(r_2,a_i),\ldots,\rho(r_p,a_i)$ for nominal

condition attribute a_i and an interval $[\min \bigcup_{j=1,2,\ldots,p} \rho(r_j, a_i), \max \bigcup_{j=1,2,\ldots,p} \rho(r_j, a_i)]$ for numerical condition attribute a_i.

The proposed clustering algorithm for k-anonymization of a set of rules R is given in Algorithm 1. This algorithm is a kind of agglomerative hierarchical clustering algorithm in which rules to be merged are selected by the minimum strength criterion and a criterion maximizing the product of the similarity defined by Equation (4.1.3) and the consistency defined by Equation (4.1.4). The set R_j of rules which do not satisfy k-anonymity, i.e., whose strength scores are less than k is generated for each decision class D_j. This rule set $R_j^{'}$ is updated by replacing rules r and r' with the merged rule r_{new}. Rule r is selected by by the minimum strength criterion. Then rule r' is selected by maximizing the product of the similarity and the consistency score with respect to rule r. The algorithm merging two rules r and r' is shown in Algorithm 2. By this algorithm, r_{new} is created. These procedures are continued until all rules in $R_j^{'}$ satisfy k-anonymity. Algorithm 1 is terminated after k-anonymization process is applied to all decision classed $D_j, j = 1, 2, \ldots, n$. As shown in Algorithm 2, the condition attributes used in the merged rule r_{new} are all condition attributes commonly used in rules r and r', more concretely, $A(r_{new}) = A(r) \cap A(r')$. The set of attribute rules for condition attribute $a \in A(r_{new})$ is defined by the minimum closed interval including both closed intervals with respect to condition attribute a given in rules r and r' if a is numerical. On the other hand, the set of attribute rules for condition attribute $a \in A(r_{new})$ is defined by the union of sets of attribute values for condition attribute a given in rules r and r' if a is nominal. We note that the strength score of r_{new} is larger than strength scores of r and r'. Because the condition of r_{new} is weaken than condition of r and r', all objects covered by rules r and r' are also covered by the merged rule r_{new}. Eventually, all objects covered by the original rule set R_j are also covered by the k-anonymized rule set R_j. As we assume that k is not more than the number of objects in the smallest decision class, i.e., $k \leq \min_{j=1,2,\ldots,n} |D_j|$, it is guaranteed that Algorithm 1 is always terminated when obtaining a set of k-anonymized rules R'.

4.2 Experiments

4.2.1 Experimental Setup

We used five datasets obtained from UCI Machine Learning Repository [71]. The parameters of the datasets are listed in Table 4.1. Column *prop* of Table 4.1 lists the ratios of rules with $Strength(r) = 1$ to all rules induced by the MLEM2 algorithm. The dataset

Algorithm 2 k-anonymous rule clustering

Input: a set of rules R and a threshold value k
Output: a set of k-anonymized rules R'

$R' \leftarrow \emptyset$
for each deicision class D_j, $j = 1, \ldots, n$ **do**
  $R'_j \leftarrow \{r \in R(D_j) : Strength(r) < k\}$
  $R' \leftarrow (R' \cup R(D_j)) \setminus R'_j$
  $min = \min\{Strength(r) : r \in R'_j\}$
  while $min < k$ **do**
    select r such that $Strength(r) = min$, if a tie occurs, randomly select one of them.
    select r' having the maximum of multiplication similarity and consistency measure $s(r, r') \times c(r, r')$, $r' \in R'_j \setminus \{r\}$; if a tie occurs, select r' with smallest $Strength(r')$; if a further tie occurs, randomly select one of them.
    $r_{new} \leftarrow$merge(r,r')
    $R'_j \leftarrow R'_j \setminus \{r, r'\}$
    $R'_j \leftarrow R'_j \cup \{r_{new}\}$
    $min = \min\{Strength(r) : r \in R'_j\}$
  end while
  $R' \leftarrow R' \cup R'_j$
end for

Algorithm 3 Rule merging : merge

Input: two rules r and r'
Output: a combined rule r''

$A \leftarrow A(r) \cap A(r')$
for each $c \in A$ **do**
  if c is numeric **then**
    $min = \min\{\rho(r, c) \cup \rho(r', c)\}$
    $max = \max\{\rho(r, c) \cup \rho(r', c)\}$
    $e_c \leftarrow f(x, c) \in [min, max]$
  else
    $e_c \leftarrow f(x, c) \in \rho(r, c) \cup \rho(r', c)$
  end if
end for
$r'' \leftarrow (\bigwedge_{c \in C} e_c \rightarrow x \in D_j)$

on census income *adult* is used to estimate whether a person earns more than fifty thousand dollars per year from census data and demographic information of the person. We use only eight condition attributes considered as quasi-identifiers in previous studies [31]. Dataset *default* is used to estimate the possibility of a credit default. Namely, whether a person is credible or not is estimated from this dataset. The dataset on german-credit data (*german-credit*) consists of 1,000 records and 20 attributes (excluding the decision attribute) of bank account holders. The decision attribute in this dataset takes either good or bad in credit classification of holders. The three above-mentioned datasets are well-known real datasets containing both numerical and categorical attributes, and are frequently used in the field of data privacy. In addition, we use other two datasets *hayes-roth* and *nursery*.

We apply a ten-fold cross validation method [34] in these experiments. Specifically, we divide each dataset into ten subsets. Nine subsets are used for training and the remaining one for a testing. Therefore, we obtain ten different evaluations by the combinations of

the nine subsets. This validation method is applied ten times with different divisions of a dataset. The average and standard deviation are calculated for each evaluation. We use MLEM2 [25] for rule induction in the experiments.

4.2.2 Classification Accuracy

In this experiment, we investigate the classification accuracy of the k-anonymous rules obtained using the proposed rule clustering. The classification accuracy is the ratio of the number of correctly classified objects to the number of all objects. We use LERS algorithm [22] to classify objects of a test dataset.

To verify the effectiveness of the proposed *similarity* and *consistency* measures, we compared the k-anonymous rule clustering approach using both measures and that using only one measure. Moreover, to verify the performance of the proposed approach, we compared it to the following three methods :

1) Random rule merging (*Random*): This method generates a set of k-anonymous rules by merging two randomly selected rules with the same conclusion. The rule-based classifier is constructed in the same way as the proposed method.

2) k-anonymization by using the number of matched condition attributes for selection of two rules to be merged (*Match*): In this method, a set of k-anonymized rules are generated from rule set R by a rule clustering but two rules to be merged are selected by maximizing the number of common condition attributes, i.e., $|A(r) \cap A(r')|$. After the k-anonymization a rule-based classifier is constructed in the same way as the proposed method.

3) Removing rules which does not satisfying k-anonymity (*Only k*): In this method, all rules whose strength scores are less than k are removed from R'. Then, a rule-based classifier is constructed only with the remaining rules in the same way as the proposed method.

TABLE 4.1: Dataset parameters and the average number, length, and strength of the rules induced by MLEM2

dataset	$\|U\|$	$\|C\|$	$\|V_d\|$	prop	number	length	strength
adult	48844	8	2	28 %	$4689.410_{\pm 42.246}$	$4.910_{\pm 0.009}$	$8.455_{\pm 0.177}$
default	30000	11	2	29 %	$4158.380_{\pm 38.443}$	$6.339_{\pm 0.009}$	$4.608_{\pm 0.049}$
german-credit	1000	20	2	11 %	$258.440_{\pm 5.968}$	$6.475_{\pm 0.051}$	$5.039_{\pm 0.177}$
hayes-roth	160	5	3	3 %	$23.240_{\pm 1.485}$	$2.692_{\pm 0.048}$	$5.869_{\pm 0.267}$
nursery	12960	8	5	14 %	$574.020_{\pm 27.414}$	$5.473_{\pm 0.019}$	$33.617_{\pm 2.786}$

In Tables 4.2 and 4.3, $Sim \times Con$, Sim, and Con represent the rule-based classifiers with k-anonymized rules obtained by the proposed rule clustering using both similarity and consistency measures, only similarity measure, and only consistency measure for the selection of a rule to be merged, respectively. *Random*, $Match$, and *Only k* are the alternative rule-based classifiers described earlier above. We examine several parameter settings of k for each dataset. As the sizes of datasets differ values of k depend on datasets. The average classification accuracies (acc) and the standard deviation (dev) are shown in the form of $acc_{\pm dev}$. The average classification accuracies in bold typeface indicate that they are statistically larger than those obtained for any of the alternative methods in a paired t-test with 5 % of significant level. The underlined values indicate that they are larger than the average classification accuracy of the rule-based classifier with all rules induced by MLEM2 shown next to the name of dataset.

We observe that the average classification accuracy scores of the proposed methods are larger than those of the three alternative methods in every dataset. Moreover, the average classification accuracies of the proposed approaches are statistically larger for all k values in dataset *adult* and some k values in datasets *default*, *german-credit* and *hayes-roth*. Method Con shows either larger or almost the some accuracy scores as method Sim, whereas method $Sim \times Con$ shows the largest accuracy scores. This implies that the consistency measure works better than the similarity and that the product of consistency and similarity measures performs best. In datasets, *adult* and *nursery*, as k values increases the average classification accuracy score decreases. This can be explained from the fact that increasing k values generate more anonymous rules that are more general. In this dataset, it seems that the characters of individual data are rather diverse because the average classification accuracy decreases as rules are generalized. On the other hand, the average classification accuracy scores of the proposed approaches in dataset *german-credit* are not affected by the value of k. Furthermore, we observe that the average classification accuracy scores of the proposed approaches in datasets *adult*, *default*, *german-credit*, and *hayes-roth* are larger than those of the rule-based classifiers with rules induced by MLEM2 for many values of k. Given that the average classification accuracy scores of method *Only k* are largest among the three alternative methods, we may consider that rules not satisfying k-anonymity are not very essential for accurate classification. However, even in such datasets, the proposed methods $Sim \times Con$ and Con perform better than method *Only k*. This implies that methods $Sim \times Con$ and Con select rules to be merged effectively so that the rule-based classifiers overcome the potential overfitting.

TABLE 4.2: Average and standard deviation of classification accuracy scores in datasets *adult*, *default*, and *german-credit*

adult : $0.767_{\pm 0.005}$									
method	$k=5$	$k=10$	$k=15$	$k=20$	$k=25$	$k=30$	$k=35$	$k=40$	$k=45$
Sim × Con	$\mathbf{0.769_{\pm 0.006}}$	$\mathbf{0.764_{\pm 0.006}}$	$\mathbf{0.763_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.761_{\pm 0.006}}$	$\mathbf{0.761_{\pm 0.006}}$
Sim	$\mathbf{0.769_{\pm 0.006}}$	$\mathbf{0.765_{\pm 0.005}}$	$0.754_{\pm 0.038}$	$0.753_{\pm 0.043}$	$0.752_{\pm 0.045}$	$0.751_{\pm 0.047}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.761_{\pm 0.007}}$	$\mathbf{0.761_{\pm 0.006}}$
Con	$\mathbf{0.769_{\pm 0.006}}$	$\mathbf{0.764_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.762_{\pm 0.006}}$	$\mathbf{0.761_{\pm 0.006}}$	$\mathbf{0.761_{\pm 0.007}}$	$\mathbf{0.761_{\pm 0.007}}$
Random	$0.720_{\pm 0.008}$	$0.720_{\pm 0.008}$	$0.720_{\pm 0.007}$	$0.717_{\pm 0.013}$	$0.719_{\pm 0.009}$	$0.719_{\pm 0.008}$	$0.719_{\pm 0.010}$	$0.720_{\pm 0.008}$	$0.721_{\pm 0.006}$
Match	$0.731_{\pm 0.006}$	$0.729_{\pm 0.006}$	$0.728_{\pm 0.006}$	$0.726_{\pm 0.006}$	$0.726_{\pm 0.006}$	$0.726_{\pm 0.006}$	$0.726_{\pm 0.006}$	$0.723_{\pm 0.006}$	$0.724_{\pm 0.006}$
Only k	$0.757_{\pm 0.006}$	$0.752_{\pm 0.006}$	$0.751_{\pm 0.006}$	$0.749_{\pm 0.006}$	$0.749_{\pm 0.006}$	$0.749_{\pm 0.006}$	$0.749_{\pm 0.006}$	$0.742_{\pm 0.006}$	$0.742_{\pm 0.006}$
default : $0.782_{\pm 0.007}$									
method	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$	$k=12$	$k=14$	$k=16$	$k=18$
Sim × Con	$0.792_{\pm 0.010}$	$\mathbf{0.798_{\pm 0.009}}$	$\mathbf{0.795_{\pm 0.007}}$	$\mathbf{0.792_{\pm 0.013}}$	$0.785_{\pm 0.019}$	$0.784_{\pm 0.008}$	$0.783_{\pm 0.013}$	$0.781_{\pm 0.004}$	$0.781_{\pm 0.015}$
Sim	$0.792_{\pm 0.004}$	$0.794_{\pm 0.003}$	$0.792_{\pm 0.014}$	$0.788_{\pm 0.006}$	$0.781_{\pm 0.003}$	$0.781_{\pm 0.015}$	$0.781_{\pm 0.012}$	$0.781_{\pm 0.014}$	$0.781_{\pm 0.011}$
Con	$0.792_{\pm 0.007}$	$\mathbf{0.798_{\pm 0.004}}$	$\mathbf{0.794_{\pm 0.011}}$	$\mathbf{0.791_{\pm 0.010}}$	$0.784_{\pm 0.013}$	$0.784_{\pm 0.016}$	$0.781_{\pm 0.014}$	$0.781_{\pm 0.017}$	$0.781_{\pm 0.011}$
Random	$0.776_{\pm 0.017}$	$0.778_{\pm 0.012}$	$0.770_{\pm 0.023}$	$0.762_{\pm 0.029}$	$0.745_{\pm 0.060}$	$0.719_{\pm 0.082}$	$0.715_{\pm 0.074}$	$0.713_{\pm 0.078}$	$0.705_{\pm 0.092}$
Match	$0.788_{\pm 0.007}$	$0.773_{\pm 0.019}$	$0.777_{\pm 0.011}$	$0.780_{\pm 0.006}$	$0.779_{\pm 0.004}$	$0.780_{\pm 0.004}$	$0.780_{\pm 0.003}$	$0.780_{\pm 0.003}$	$0.779_{\pm 0.003}$
Only k	$0.789_{\pm 0.005}$	$0.792_{\pm 0.004}$	$0.790_{\pm 0.003}$	$0.786_{\pm 0.003}$	$0.784_{\pm 0.003}$	$0.783_{\pm 0.003}$	$0.782_{\pm 0.003}$	$0.781_{\pm 0.003}$	$0.781_{\pm 0.003}$
german-credit : $0.689_{\pm 0.026}$									
method	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$	$k=12$	$k=14$	$k=16$	$k=18$
Sim × Con	$0.692_{\pm 0.031}$	$0.704_{\pm 0.023}$	$0.699_{\pm 0.020}$	$\mathbf{0.701_{\pm 0.016}}$	$\mathbf{0.702_{\pm 0.015}}$	$\mathbf{0.703_{\pm 0.016}}$	$\mathbf{0.702_{\pm 0.016}}$	$\mathbf{0.703_{\pm 0.016}}$	$\mathbf{0.703_{\pm 0.016}}$
Sim	$0.676_{\pm 0.037}$	$0.670_{\pm 0.050}$	$0.646_{\pm 0.048}$	$0.648_{\pm 0.057}$	$0.649_{\pm 0.065}$	$0.646_{\pm 0.070}$	$0.654_{\pm 0.087}$	$0.651_{\pm 0.075}$	$0.643_{\pm 0.096}$
Con	$0.693_{\pm 0.029}$	$0.704_{\pm 0.023}$	$0.699_{\pm 0.020}$	$\mathbf{0.701_{\pm 0.016}}$	$\mathbf{0.702_{\pm 0.015}}$	$\mathbf{0.702_{\pm 0.015}}$	$\mathbf{0.702_{\pm 0.015}}$	$\mathbf{0.703_{\pm 0.015}}$	$\mathbf{0.703_{\pm 0.015}}$
Random	$0.656_{\pm 0.042}$	$0.668_{\pm 0.049}$	$0.681_{\pm 0.038}$	$0.685_{\pm 0.042}$	$0.671_{\pm 0.072}$	$0.676_{\pm 0.059}$	$0.680_{\pm 0.065}$	$0.678_{\pm 0.068}$	$0.668_{\pm 0.083}$
Match	$0.690_{\pm 0.033}$	$0.699_{\pm 0.030}$	$0.682_{\pm 0.065}$	$0.654_{\pm 0.092}$	$0.641_{\pm 0.104}$	$0.584_{\pm 0.140}$	$0.563_{\pm 0.149}$	$0.555_{\pm 0.157}$	$0.531_{\pm 0.154}$
Only k	$0.690_{\pm 0.030}$	$0.700_{\pm 0.023}$	$0.694_{\pm 0.020}$	$0.695_{\pm 0.016}$	$0.695_{\pm 0.015}$	$0.695_{\pm 0.015}$	$0.695_{\pm 0.015}$	$0.698_{\pm 0.015}$	$0.698_{\pm 0.015}$

TABLE 4.3: Average and standard deviation of classification accuracy scores in datasets *hayes-roth* and *nursery*

hayes-roth : $0.809_{\pm 0.088}$									
method	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
Sim × Con	$\underline{0.809}_{\pm 0.083}$	$\underline{0.809}_{\pm 0.083}$	$\underline{0.825}_{\pm 0.096}$	$\underline{0.833}_{\pm 0.089}$	$\underline{0.833}_{\pm 0.094}$	$\mathbf{0.808_{\pm 0.148}}$	$\mathbf{0.806_{\pm 0.147}}$	$\mathbf{0.700_{\pm 0.156}}$	$0.594_{\pm 0.129}$
Sim	$\underline{0.810}_{\pm 0.083}$	$\underline{0.810}_{\pm 0.083}$	$\underline{0.825}_{\pm 0.096}$	$\underline{0.835}_{\pm 0.090}$	$\underline{0.835}_{\pm 0.090}$	$\mathbf{0.804_{\pm 0.149}}$	$\mathbf{0.800_{\pm 0.148}}$	$\mathbf{0.694_{\pm 0.155}}$	$0.612_{\pm 0.096}$
Con	$\underline{0.810}_{\pm 0.083}$	$\underline{0.810}_{\pm 0.083}$	$\underline{0.825}_{\pm 0.096}$	$\underline{0.831}_{\pm 0.089}$	$\underline{0.835}_{\pm 0.094}$	$\underline{\mathbf{0.810}}_{\pm \mathbf{0.148}}$	$\mathbf{0.806_{\pm 0.147}}$	$\mathbf{0.700_{\pm 0.156}}$	$0.594_{\pm 0.129}$
Random	$0.776_{\pm 0.109}$	$0.707_{\pm 0.132}$	$0.698_{\pm 0.140}$	$0.711_{\pm 0.126}$	$0.650_{\pm 0.131}$	$0.708_{\pm 0.129}$	$0.675_{\pm 0.152}$	$0.633_{\pm 0.146}$	$0.631_{\pm 0.129}$
Match	$0.702_{\pm 0.104}$	$0.699_{\pm 0.073}$	$0.682_{\pm 0.093}$	$0.606_{\pm 0.127}$	$0.588_{\pm 0.137}$	$0.588_{\pm 0.137}$	$0.610_{\pm 0.130}$	$0.622_{\pm 0.137}$	$0.641_{\pm 0.156}$
Only k	$0.801_{\pm 0.083}$	$0.801_{\pm 0.083}$	$\underline{0.822}_{\pm 0.096}$	$\underline{0.830}_{\pm 0.089}$	$\underline{0.830}_{\pm 0.089}$	$0.792_{\pm 0.148}$	$0.791_{\pm 0.148}$	$0.675_{\pm 0.166}$	$0.550_{\pm 0.134}$
nursery : $0.910_{\pm 0.130}$									
method	$k=3$	$k=6$	$k=9$	$k=12$	$k=15$	$k=18$	$k=21$	$k=24$	$k=27$
Sim × Con	$0.906_{\pm 0.131}$	$0.898_{\pm 0.131}$	$0.898_{\pm 0.131}$	$0.896_{\pm 0.130}$	$0.887_{\pm 0.130}$	$0.883_{\pm 0.130}$	$0.885_{\pm 0.129}$	$0.866_{\pm 0.132}$	$0.865_{\pm 0.131}$
Sim	$0.906_{\pm 0.130}$	$0.899_{\pm 0.130}$	$0.899_{\pm 0.131}$	$0.896_{\pm 0.131}$	$0.886_{\pm 0.131}$	$0.881_{\pm 0.133}$	$0.881_{\pm 0.133}$	$0.875_{\pm 0.135}$	$0.875_{\pm 0.134}$
Con	$0.906_{\pm 0.131}$	$0.899_{\pm 0.131}$	$0.898_{\pm 0.131}$	$0.896_{\pm 0.130}$	$0.886_{\pm 0.130}$	$0.883_{\pm 0.131}$	$0.885_{\pm 0.130}$	$0.866_{\pm 0.132}$	$0.866_{\pm 0.132}$
Random	$0.890_{\pm 0.130}$	$0.882_{\pm 0.132}$	$0.862_{\pm 0.139}$	$0.846_{\pm 0.145}$	$0.839_{\pm 0.144}$	$0.827_{\pm 0.139}$	$0.822_{\pm 0.145}$	$0.799_{\pm 0.147}$	$0.791_{\pm 0.143}$
Match	$0.899_{\pm 0.130}$	$0.889_{\pm 0.130}$	$0.878_{\pm 0.131}$	$0.865_{\pm 0.131}$	$0.845_{\pm 0.131}$	$0.831_{\pm 0.132}$	$0.831_{\pm 0.132}$	$0.823_{\pm 0.133}$	$0.823_{\pm 0.133}$
Only k	$0.902_{\pm 0.130}$	$0.893_{\pm 0.130}$	$0.893_{\pm 0.130}$	$0.891_{\pm 0.131}$	$0.882_{\pm 0.131}$	$0.881_{\pm 0.132}$	$0.881_{\pm 0.132}$	$0.873_{\pm 0.133}$	$0.873_{\pm 0.133}$

The average length, number, and strength of the rules induced by the proposed methods and the three alternative methods are shown in Fig 4.1, 4.2, and 4.3. The length of rule is the total number of condition attributes in the rule. We observe that the average length of the rules induced by method *Only k* is larger than that of any other method in datasets *adult*, *default*, and *german-credit*, whereas method *Random* shows the smaller lengths among all methods. As k increases, the number of induced rules decreases. This is not surprising because the rules are more merged as k increases. The numbers of rules in methods $Sim \times Con$, Con, and *Only k* are small compared to those in other methods. In addition, the average strength scores of rules in methods $Sim \times Con$ and Con are larger than those in other methods for every dataset. These facts implies that in methods $Sim \times Con$ and Con original rules are merged to rules with large strength scores. Namely, k-anonymization is achieved by merging rules with small strength scores to rules with large scores in methods $Sim \times Con$ and Con. The differences of the classification accuracies between the classification with rules induced by MLEM2 algorithm and all rule-based classifiers with k-anonymous rules are not very clear in dataset *nursery*. As shown in Table 5.1, the strength scores of dataset *nursery* are large compared to other datasets. Each decision class would be estimated easily in dataset *nursery*. From the overall results, we can conclude that the proposed methods maintain a sufficiently large accuracy scores for all the evaluated datasets.

4.2.3 Privacy Preservation

In the second experiment, we verify the performance of privacy preservation. Assuming that the set of rules is published and some condition attributes of an object are revealed by an attacker, we examine to what amount decision classes are revealed from the published set of rules and the condition attribute values known by the attacker. In this experiment, we do not use the rule-based classifier based on LERS for estimating a unique decision class and instead estimate the decision class simply by taking the conjunction of conclusions of rules. The attack is successful when a unique decision class is estimated. Let the condition attribute sets known by the attacker be $P = \{a_{p(1)}, a_{p(2)}, \ldots, a_{p(q)}\}$. The possible number of combinations of condition attribute values is obtained as $V(P) = \prod_{j=1,2,\ldots,q} |V_{a_j}|$. Considering all possible $P \subseteq A$, the total number of possible combinations of revealed condition attribute values is obtained as $\sum_{P \subseteq A} V(P)$. We calculate the ratio of the number of the success levels to the number of all the combinations of the condition attribute values. A smaller ratio is better in privacy preservation.

The successful attack ratios for the five datasets are shown in Tables 4.4 and 4.5. We set $|P| = 1, 2$, and 3, and k by the values used in the previous experiments. The

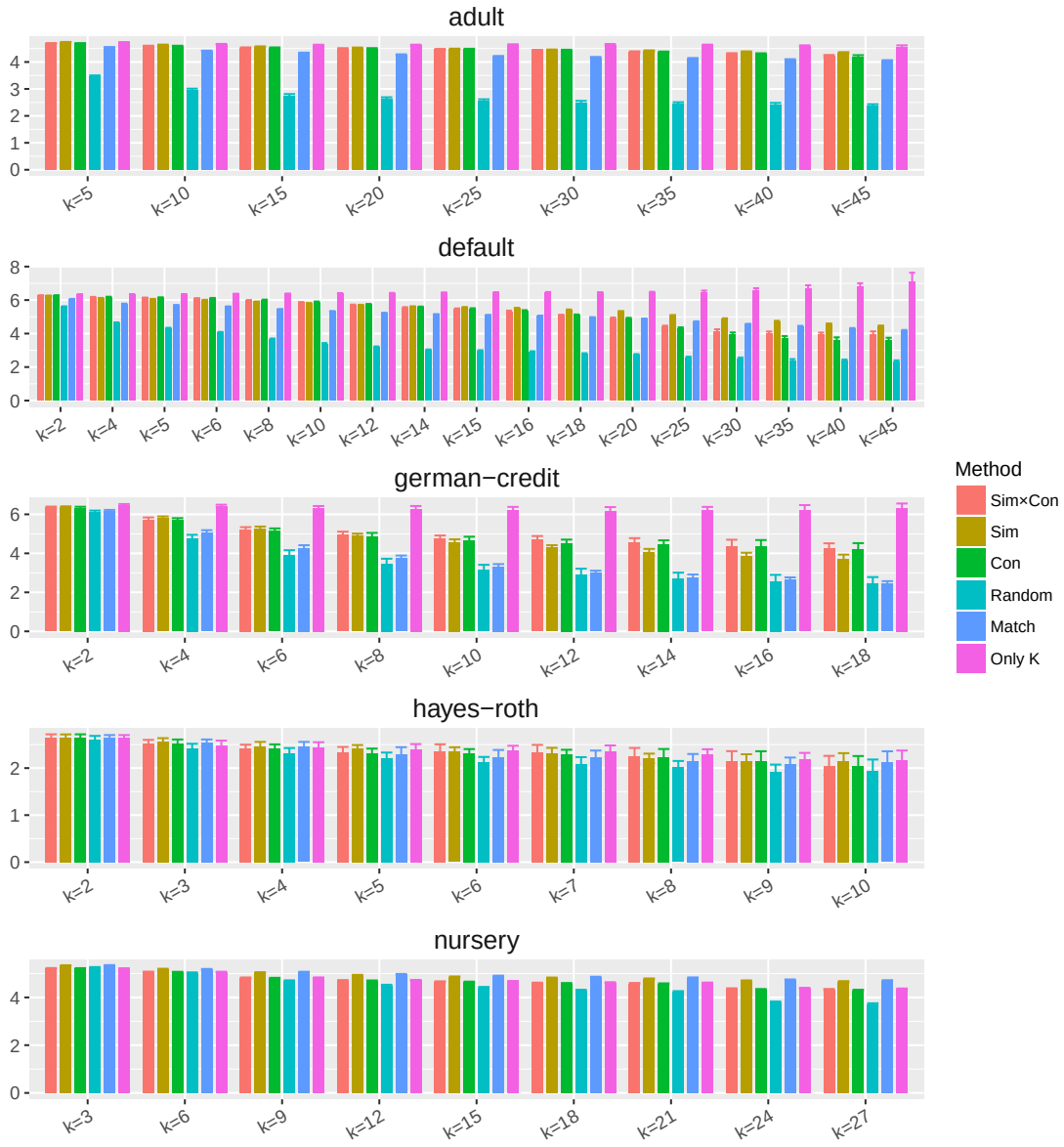


FIGURE 4.1: Average length of rules for all methods

underlined ratios indicate the largest among all methods. The ratios in bold typeface indicate that they are statistically larger than those obtained any other methods in a paired t-test with 5 % significant level. We compared the proposed method $Sim \times Con$ with the three alternative methods described in Section 4.2.2.

Except for dataset *german-credit*, we observed that successful attack ratio in method *Match* was smaller than those in the other methods. In contract, the successful attack ratio in method *Only k* is large except for dataset *hayes-roth*. The successful attack ratio of the proposed method is smaller than those of the three alternative methods in the $|P| = 3$ of dataset *german-credit*. Method *Match* merges rules based on the number of common condition attributes. As the number of elementary conditions increase,

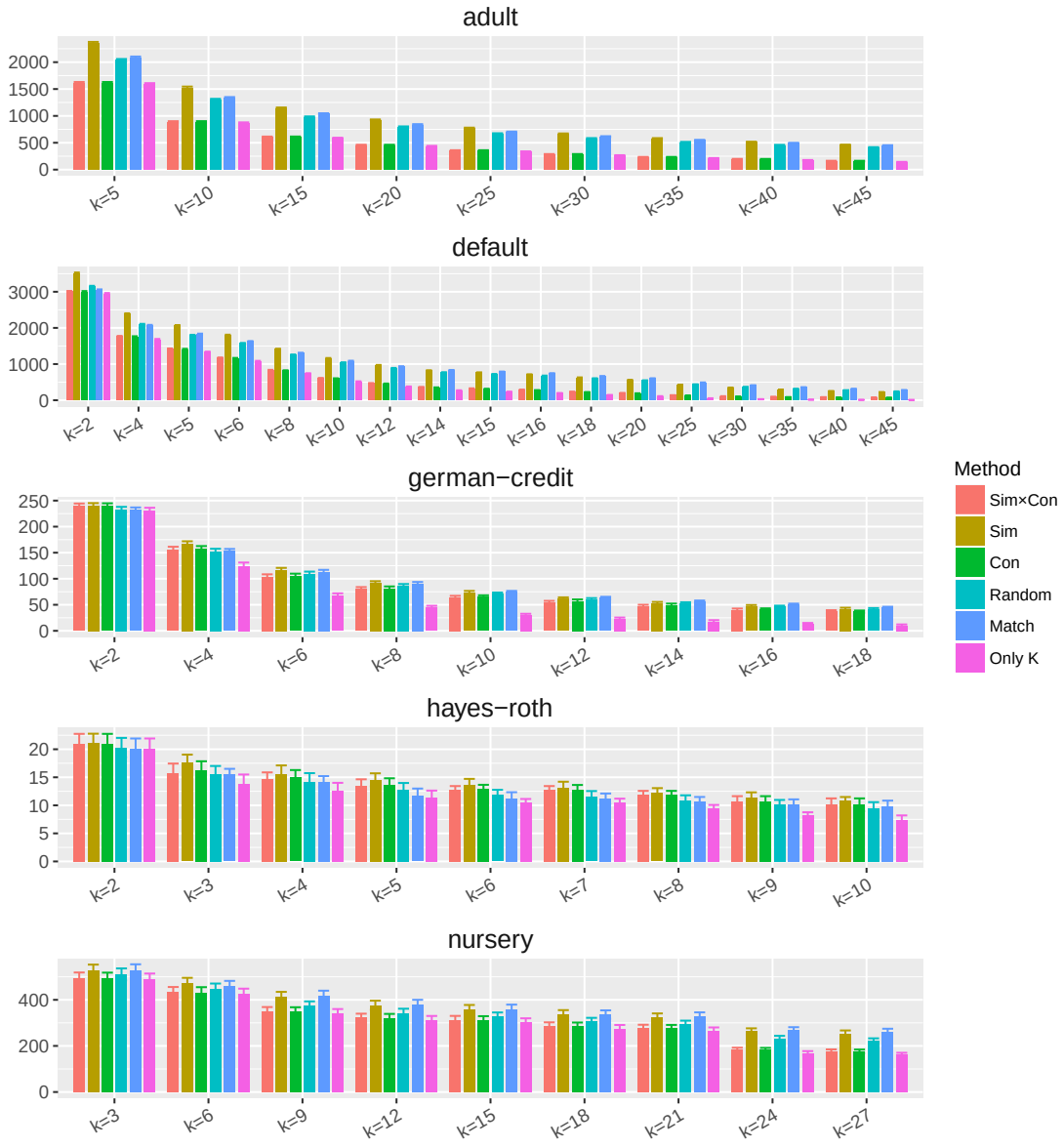


FIGURE 4.2: Average numbers of rules for all methods

the number of rules having some of condition attribute values known by the attacker increases. In addition, the decision classes in the conclusions of these rules tend to differ from each other. Therefore, the successful attack ratio decreases. To maintain the classification accuracy score, the proposed method preserves the original rules if they satisfy k-anonymity and generates merged rules to satisfy k-anonymity. As k increases, the number of merged rules increases while both the number of original rules and length of the rules decrease. Consequently, the number of rules having some of the condition attribute values known by the attacker is small. Thus, the successful attack ratio increases. We observe that our proposed method takes larger accuracy score than method *Match* and a smaller successful attack ratio than method *Only k*. There is no

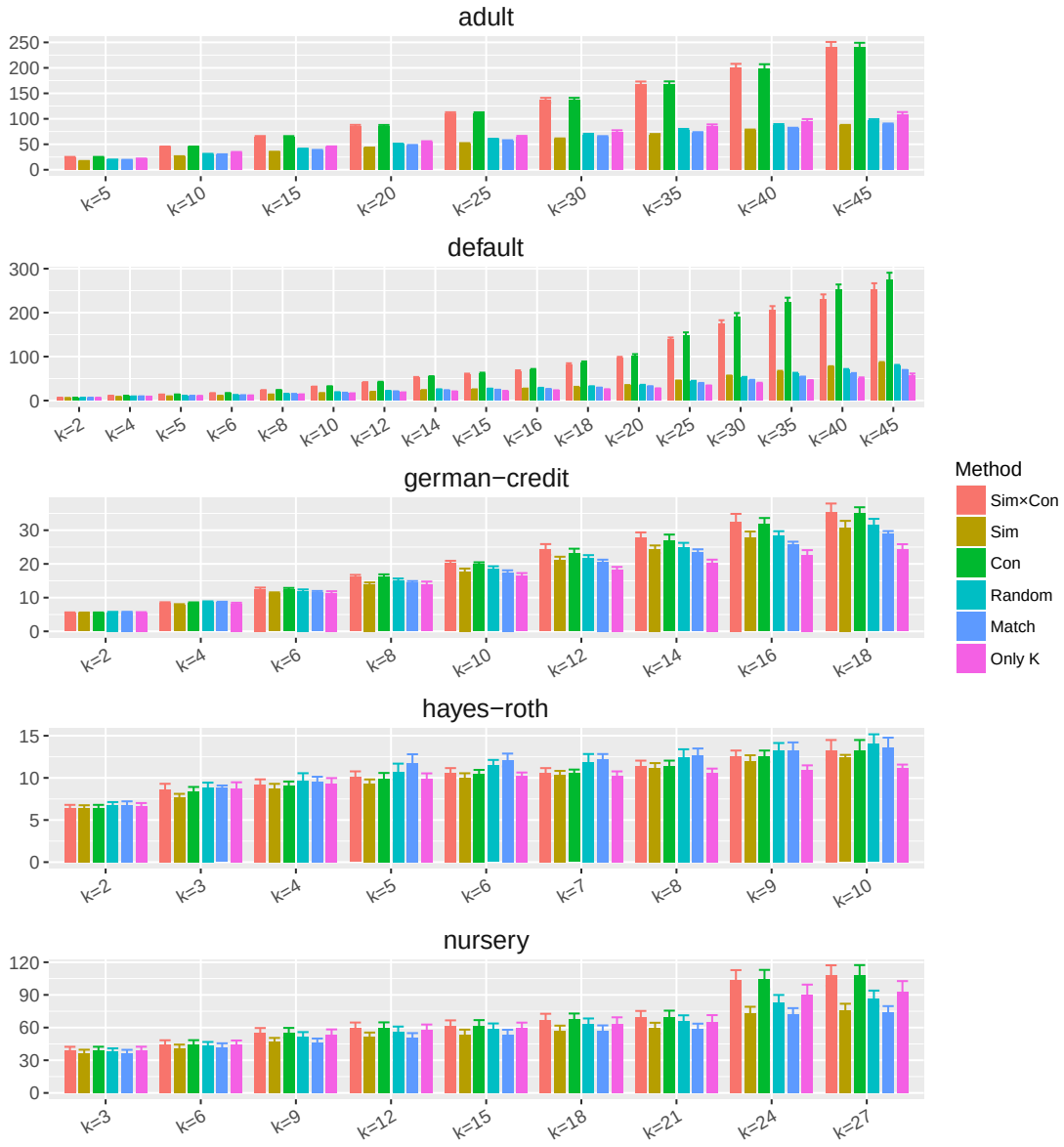


FIGURE 4.3: Average strength of rules for all methods

alternative method that dominates the proposed method in both classification accuracy and successful attack ratio.

4.3 Conclusions

In this chapter, we proposed a rule clustering approach to achieve k-anonymity in rules to be published for privacy protection. In the proposed method, rules selected by the similarity and consistency measures are merged iteratively until k-anonymity is achieved for

TABLE 4.4: Successful attack ratios in datasets *adult*, *default*, and *german-credit*

adult						
$\|P\|$	method	$k=5$	$k=15$	$k=25$	$k=35$	$k=45$
1	Sim × Con	$0.072_{\pm 0.022}$	$0.090_{\pm 0.035}$	$0.097_{\pm 0.038}$	$0.099_{\pm 0.036}$	$0.108_{\pm 0.037}$
	Random	$0.110_{\pm 0.027}$	$0.218_{\pm 0.049}$	$0.271_{\pm 0.057}$	$0.316_{\pm 0.070}$	$0.346_{\pm 0.089}$
	Match	$\mathbf{0.025_{\pm 0.010}}$	$\mathbf{0.026_{\pm 0.010}}$	$\mathbf{0.027_{\pm 0.010}}$	$\mathbf{0.029_{\pm 0.009}}$	$\mathbf{0.030_{\pm 0.009}}$
	Only K	$\underline{0.536}_{\pm 0.026}$	$\underline{0.825}_{\pm 0.020}$	$\underline{0.916}_{\pm 0.014}$	$\underline{0.927}_{\pm 0.024}$	$\underline{1.000}_{\pm 0.000}$
2	Sim × Con	$0.636_{\pm 0.038}$	$0.764_{\pm 0.069}$	$0.782_{\pm 0.067}$	$0.767_{\pm 0.076}$	$0.799_{\pm 0.065}$
	Random	$0.513_{\pm 0.021}$	$0.605_{\pm 0.049}$	$0.655_{\pm 0.064}$	$0.685_{\pm 0.079}$	$0.710_{\pm 0.085}$
	Match	$\mathbf{0.308_{\pm 0.006}}$	$\mathbf{0.256_{\pm 0.009}}$	$\mathbf{0.242_{\pm 0.012}}$	$\mathbf{0.243_{\pm 0.016}}$	$\mathbf{0.258_{\pm 0.017}}$
	Only K	$\underline{0.803}_{\pm 0.011}$	$\underline{0.967}_{\pm 0.008}$	$\underline{0.992}_{\pm 0.003}$	$\underline{0.993}_{\pm 0.005}$	$\underline{1.000}_{\pm 0.000}$
3	Sim × Con	$0.925_{\pm 0.014}$	$0.992_{\pm 0.006}$	$0.995_{\pm 0.008}$	$0.987_{\pm 0.021}$	$0.992_{\pm 0.009}$
	Random	$0.785_{\pm 0.022}$	$0.859_{\pm 0.030}$	$0.886_{\pm 0.037}$	$0.901_{\pm 0.041}$	$0.910_{\pm 0.043}$
	Match	$\mathbf{0.539_{\pm 0.008}}$	$\mathbf{0.506_{\pm 0.011}}$	$\mathbf{0.514_{\pm 0.015}}$	$\mathbf{0.550_{\pm 0.021}}$	$\mathbf{0.604_{\pm 0.023}}$
	Only K	$\underline{0.931}_{\pm 0.007}$	$\underline{0.996}_{\pm 0.002}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$
default						
$\|P\|$	method	$k=2$	$k=6$	$k=10$	$k=14$	$k=18$
1	Sim × Con	$0.132_{\pm 0.025}$	$0.148_{\pm 0.020}$	$0.133_{\pm 0.026}$	$0.135_{\pm 0.024}$	$0.134_{\pm 0.025}$
	Random	$0.132_{\pm 0.026}$	$0.175_{\pm 0.031}$	$0.184_{\pm 0.033}$	$0.194_{\pm 0.034}$	$0.203_{\pm 0.038}$
	Match	$\mathbf{0.127_{\pm 0.012}}$	$\mathbf{0.128_{\pm 0.011}}$	$\mathbf{0.128_{\pm 0.014}}$	$\mathbf{0.127_{\pm 0.013}}$	$\mathbf{0.126_{\pm 0.012}}$
	Only K	$\underline{0.176}_{\pm 0.019}$	$\underline{0.394}_{\pm 0.035}$	$\underline{0.513}_{\pm 0.021}$	$\underline{0.650}_{\pm 0.047}$	$\underline{0.787}_{\pm 0.063}$
2	Sim × Con	$\underline{0.340}_{\pm 0.016}$	$0.393_{\pm 0.019}$	$0.402_{\pm 0.021}$	$0.397_{\pm 0.027}$	$0.406_{\pm 0.027}$
	Random	$0.285_{\pm 0.011}$	$0.358_{\pm 0.018}$	$0.402_{\pm 0.024}$	$0.437_{\pm 0.030}$	$0.484_{\pm 0.042}$
	Match	$\mathbf{0.279_{\pm 0.009}}$	$\mathbf{0.270_{\pm 0.012}}$	$\mathbf{0.272_{\pm 0.015}}$	$\mathbf{0.277_{\pm 0.014}}$	$\mathbf{0.283_{\pm 0.014}}$
	Only K	$0.286_{\pm 0.010}$	$\underline{0.691}_{\pm 0.011}$	$\underline{0.827}_{\pm 0.015}$	$\underline{0.935}_{\pm 0.019}$	$\underline{0.962}_{\pm 0.013}$
3	Sim × Con	$\underline{0.487}_{\pm 0.009}$	$0.771_{\pm 0.024}$	$0.796_{\pm 0.049}$	$0.806_{\pm 0.039}$	$0.809_{\pm 0.044}$
	Random	$0.420_{\pm 0.009}$	$0.630_{\pm 0.021}$	$0.695_{\pm 0.026}$	$0.738_{\pm 0.028}$	$0.777_{\pm 0.039}$
	Match	$\mathbf{0.351_{\pm 0.007}}$	$\mathbf{0.348_{\pm 0.010}}$	$\mathbf{0.369_{\pm 0.012}}$	$\mathbf{0.396_{\pm 0.012}}$	$\mathbf{0.438_{\pm 0.015}}$
	Only K	$0.449_{\pm 0.007}$	$\underline{0.876}_{\pm 0.006}$	$\underline{0.964}_{\pm 0.007}$	$\underline{0.990}_{\pm 0.005}$	$\underline{0.993}_{\pm 0.003}$
german-credit						
$\|P\|$	method	$k=2$	$k=6$	$k=10$	$k=14$	$k=18$
1	Sim × Con	$0.166_{\pm 0.026}$	$0.294_{\pm 0.053}$	$\mathbf{0.338_{\pm 0.066}}$	$\mathbf{0.385_{\pm 0.071}}$	$\mathbf{0.427_{\pm 0.073}}$
	Random	$\underline{0.170}_{\pm 0.027}$	$0.452_{\pm 0.060}$	$0.604_{\pm 0.082}$	$0.688_{\pm 0.081}$	$0.755_{\pm 0.077}$
	Match	$\mathbf{0.156_{\pm 0.024}}$	$\mathbf{0.276_{\pm 0.035}}$	$0.411_{\pm 0.050}$	$0.509_{\pm 0.066}$	$0.572_{\pm 0.067}$
	Only K	$0.169_{\pm 0.022}$	$\underline{0.705}_{\pm 0.098}$	$\underline{0.981}_{\pm 0.050}$	$\underline{0.995}_{\pm 0.021}$	$\underline{1.000}_{\pm 0.000}$
2	Sim × Con	$0.622_{\pm 0.013}$	$\mathbf{0.793_{\pm 0.023}}$	$\mathbf{0.817_{\pm 0.031}}$	$\mathbf{0.827_{\pm 0.030}}$	$\mathbf{0.847_{\pm 0.034}}$
	Random	$0.634_{\pm 0.016}$	$0.836_{\pm 0.035}$	$0.896_{\pm 0.035}$	$0.927_{\pm 0.034}$	$0.949_{\pm 0.031}$
	Match	$\underline{0.638}_{\pm 0.017}$	$0.800_{\pm 0.024}$	$0.886_{\pm 0.025}$	$0.935_{\pm 0.022}$	$0.955_{\pm 0.021}$
	Only K	$0.620_{\pm 0.014}$	$\underline{0.907}_{\pm 0.040}$	$\underline{0.996}_{\pm 0.013}$	$\underline{0.999}_{\pm 0.004}$	$\underline{1.000}_{\pm 0.000}$
3	Sim × Con	$\mathbf{0.837_{\pm 0.010}}$	$\mathbf{0.938_{\pm 0.011}}$	$\mathbf{0.949_{\pm 0.017}}$	$\mathbf{0.954_{\pm 0.015}}$	$\mathbf{0.961_{\pm 0.020}}$
	Random	$0.846_{\pm 0.010}$	$0.949_{\pm 0.019}$	$0.974_{\pm 0.016}$	$0.984_{\pm 0.014}$	$0.991_{\pm 0.011}$
	Match	$\underline{0.856}_{\pm 0.013}$	$0.948_{\pm 0.008}$	$0.977_{\pm 0.013}$	$0.995_{\pm 0.008}$	$0.999_{\pm 0.004}$
	Only K	$0.836_{\pm 0.010}$	$\underline{0.964}_{\pm 0.019}$	$\underline{0.999}_{\pm 0.003}$	$\underline{1.000}_{\pm 0.001}$	$\underline{1.000}_{\pm 0.000}$

all the rules to be published. By using five datasets obtained from UCI Machine Learning Repository, the classification accuracy of the rule-based classifier and the degree of privacy protection of rules obtained using the proposed method were compared to those of three alternative methods. The results showed that the proposed method was advantageous over the three alternative methods for obtaining k-anonymized rules. Moreover,

TABLE 4.5: Successful attack ratios in datasets *hayes-roth* and *nursery*

hayes-roth						
$\|P\|$	method	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$
1	Sim × Con	$0.335_{\pm 0.089}$	$0.393_{\pm 0.067}$	$\underline{0.402}_{\pm 0.119}$	$0.420_{\pm 0.115}$	$0.486_{\pm 0.122}$
	Random	$0.338_{\pm 0.080}$	$0.366_{\pm 0.072}$	$0.399_{\pm 0.092}$	$0.421_{\pm 0.084}$	$0.466_{\pm 0.106}$
	Match	$\mathbf{0.326_{\pm 0.083}}$	$\mathbf{0.333_{\pm 0.081}}$	$\mathbf{0.334_{\pm 0.096}}$	$\mathbf{0.375_{\pm 0.095}}$	$\mathbf{0.392_{\pm 0.099}}$
	Only K	$\underline{0.349}_{\pm 0.086}$	$\underline{0.463}_{\pm 0.088}$	$0.396_{\pm 0.116}$	$\underline{0.423}_{\pm 0.122}$	$\underline{0.548}_{\pm 0.145}$
2	Sim × Con	$0.757_{\pm 0.038}$	$\underline{0.785}_{\pm 0.037}$	$\underline{0.740}_{\pm 0.075}$	$\underline{0.729}_{\pm 0.073}$	$\underline{0.766}_{\pm 0.098}$
	Random	$0.719_{\pm 0.050}$	$0.681_{\pm 0.092}$	$0.685_{\pm 0.123}$	$0.723_{\pm 0.120}$	$0.758_{\pm 0.142}$
	Match	$\mathbf{0.687_{\pm 0.044}}$	$\mathbf{0.597_{\pm 0.066}}$	$\mathbf{0.559_{\pm 0.122}}$	$\mathbf{0.573_{\pm 0.134}}$	$\mathbf{0.596_{\pm 0.135}}$
	Only K	$\underline{0.808}_{\pm 0.027}$	$0.740_{\pm 0.064}$	$0.675_{\pm 0.085}$	$0.623_{\pm 0.125}$	$0.724_{\pm 0.094}$
3	Sim × Con	$0.967_{\pm 0.027}$	$0.994_{\pm 0.027}$	$0.994_{\pm 0.027}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.004}$
	Random	$0.954_{\pm 0.047}$	$0.933_{\pm 0.078}$	$0.923_{\pm 0.092}$	$0.948_{\pm 0.085}$	$0.946_{\pm 0.095}$
	Match	$\mathbf{0.934_{\pm 0.054}}$	$\mathbf{0.876_{\pm 0.079}}$	$\mathbf{0.797_{\pm 0.152}}$	$\mathbf{0.813_{\pm 0.154}}$	$\mathbf{0.839_{\pm 0.132}}$
	Only K	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$	$\underline{1.000}_{\pm 0.000}$
nursery						
$\|P\|$	method	$k=3$	$k=9$	$k=15$	$k=21$	$k=27$
1	Sim × Con	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$	$0.033_{\pm 0.014}$	$0.042_{\pm 0.020}$	$0.046_{\pm 0.022}$
	Random	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$	$0.036_{\pm 0.017}$	$0.048_{\pm 0.022}$	$0.058_{\pm 0.028}$
	Match	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$	$0.031_{\pm 0.013}$
	Only K	$0.031_{\pm 0.013}$	$\underline{0.045}_{\pm 0.022}$	$\underline{0.049}_{\pm 0.023}$	$\underline{0.166}_{\pm 0.045}$	$\underline{0.301}_{\pm 0.039}$
2	Sim × Con	$0.208_{\pm 0.027}$	$0.281_{\pm 0.039}$	$0.366_{\pm 0.052}$	$0.398_{\pm 0.064}$	$0.434_{\pm 0.073}$
	Random	$0.202_{\pm 0.023}$	$0.188_{\pm 0.028}$	$0.233_{\pm 0.032}$	$0.304_{\pm 0.048}$	$0.380_{\pm 0.056}$
	Match	$\mathbf{0.188_{\pm 0.023}}$	$\mathbf{0.158_{\pm 0.025}}$	$\mathbf{0.156_{\pm 0.028}}$	$\mathbf{0.161_{\pm 0.027}}$	$\mathbf{0.159_{\pm 0.027}}$
	Only K	$\underline{0.269}_{\pm 0.028}$	$\underline{0.480}_{\pm 0.027}$	$\underline{0.557}_{\pm 0.024}$	$\underline{0.688}_{\pm 0.023}$	$\underline{0.804}_{\pm 0.019}$
3	Sim × Con	$0.616_{\pm 0.035}$	$0.729_{\pm 0.044}$	$0.784_{\pm 0.037}$	$0.789_{\pm 0.040}$	$0.813_{\pm 0.041}$
	Random	$0.588_{\pm 0.017}$	$0.545_{\pm 0.024}$	$0.605_{\pm 0.043}$	$0.675_{\pm 0.042}$	$0.726_{\pm 0.035}$
	Match	$\mathbf{0.571_{\pm 0.025}}$	$\mathbf{0.498_{\pm 0.024}}$	$\mathbf{0.471_{\pm 0.030}}$	$\mathbf{0.469_{\pm 0.030}}$	$\mathbf{0.460_{\pm 0.029}}$
	Only K	$\underline{0.729}_{\pm 0.015}$	$\underline{0.848}_{\pm 0.010}$	$\underline{0.880}_{\pm 0.009}$	$\underline{0.913}_{\pm 0.010}$	$\underline{0.957}_{\pm 0.008}$

compared to the original MLEM2, the set of rules obtained using the proposed method maintained a suitable level of classification accuracy. Furthermore, the proposed method maintained successful attack ratios comparable to those of the conceivable alternative methods. Thus, the rules obtained using the proposed method protect the privacy to an acceptable level.

Chapter 5

Rule-based Classifier Using Bernoulli Mixture Model for Utilizing Published Rules

In this chapter, we propose a mixture model approach to constructing an effective rule-based classifier by utilizing published rules. When rules published by different organizations and companies are available, it is worthwhile to utilize the published rules for constructing a more effective and useful rule-based classifier by incorporating them into the rules induced from the dataset at hand. However, we would not be able to know the originating dataset of the published rules. Then, published rules may be induced from some biased dataset. For instance, a number of objects with a condition attribute value exist in the biased dataset compared to our dataset at hand. The published rules may include some special characters resulting from biases such as territorial, cultural, ethnic, and religious characters, different age strata bias and so on. If we incorporated published set of rules induced from such biased dataset into the rules induced from the dataset at hand, the performance of a rule-based classifier constructed by both published and owned rules may deteriorate. Nevertheless, it would be advantageous to utilize the published set of rules because some useful knowledge not obtained from the dataset at hand may be included in the published rules. By utilizing the multiple sets of rules induced from various kinds of datasets, *e.g.*, medical data in different hospitals and questionnaire data obtained from various people, we may construct a more widely applicable and useful rule-based classifier.

We apply the concept of the mixture model in order to overcome the potential biases in the published rules. By applying LERS algorithm to each set of rules, we obtain adequacy scores over decision classes. EM algorithm [14] is then used for estimating

the approximate mixture ratios of the adequacy scores. The proposed approach is explained in Section 5.1. In Section 5.2, numerical experiments are conducted to examine the performance of the proposed approach. The proposed approach is compared with four alternative methods in several datasets obtained from public domain. The effectiveness of the proposed approach is demonstrated through the numerical experiments. Conclusions are given in Section 5.3.

5.1 Mixture Model Approach

In this section, we propose a mixture model approach in order to utilize both published rules and the rules induced from dataset at hand. In this thesis, we consider a two-decision class case, *i.e.*, binary classification problem. Assuming that K sets of rules are induced from K decision tables owned by different organizations and ourselves, we can obtain the K sets of rules. In addition, the score of $Strength(r)$ is also published for each rule r among K sets of rules.

5.1.1 Adequacy Score based on LERS Algorithm

We define an adequacy score of a set of rules R_k induced from k-th decision table to a decision class D_i for a object u based on LERS [22]. When all conditions of some rule in R_k are satisfied with object u, the adequacy score is defined by $q_k^u(D_i) = S_k^u(D_i)/\sum_i S_k^u(D_i)$, where $S_k^u(D_i)$ is calculated using Equation (2.3.3) for each object u by using a set of rules R_k. The adequacy score $q_k^u(D_i)$ evaluates the score showing to what extent the object u belongs in each decision class D_i. When there is no rule whose conditions are totally satisfied with object u, the adequacy score is defined by $q_k^u(D_i) = M_k^u(D_i)/\sum_i M_k^u(D_i)$, where $M_k^u(D_i)$ is calculated using Equation (2.3.4) for each object u using a set of rules R_k.

5.1.2 Mixture Model and EM Algorithm

In the binary classification problem, we assume $D_1 \cup D_2 = U$ and $q_k^u(D_1) = 1 - q_k^u(D_2)$. For convenience, let $q_{k,u} = q_k^u(D_1)$ and the class labels $y_u \in \{0, 1\}$ be given. When an object u belongs to D_1, we set u as $y_u = 1$, otherwise $y_u = 0$. Let Y be a set of binary variables y_u, which is governed by a Bernoulli distribution with parameter $q_{k,u}$. Variable y_u and $u \in U$ are independent of one another. Namely, variable y_u is not affected by another variable $y_v, v \neq u$ Next, we pre-set initial mixture ratio π_k^0. We first set the

complete likelihood and log-likelihood:

$$P(\mathbf{Y}, \mathbf{Z}) = \prod_{u \in U} \prod_{k \in K} (\pi_k q_{k,u}^{y_u} (1 - q_{k,u})^{1-y_u})^{z_{u,k}} \tag{5.1.1}$$

$$\log P(\mathbf{Y}, \mathbf{Z} | \boldsymbol{q}, \boldsymbol{\pi}) = \sum_{u \in U} \sum_{k \in K} z_{u,k} (\log \pi_k + y_u \log q_{k,u} + (1 - y_u) \log(1 - q_{k,u})) \tag{5.1.2}$$

where $\pi_k, k \in K$ are assumed to be known and satisfy $\pi_k \geq 0$, $\sum_{k \in K} \pi_k = 1$, and K is an index set of rule-based classifiers. Variables $z_{u,k}$ are hidden variables and representing the component assignment to object y_u. Namely, if y_u is drawn from the k-th component, $z_{u,k} = 1$ while the remainings are all 0. The form of the log-likelihood function in Equation (5.1.2) is generally preferred because it makes EM algorithm [14] available. The expectation step computes the expected value of the hidden variable by giving the current parameter values [7]. The expected value of the hidden variable $z_{u,k}$ is then given by

$$\gamma(z_{u,k}) = \frac{\pi_k q_{k,u}^{y_u} (1 - q_{k,u})^{1-y_u}}{\sum_{k \in K} q_{k,u}^{y_u} (1 - q_{k,u})^{1-y_u}}. \tag{5.1.3}$$

The expected value of the complete log-likelihood with respect to $\mathbf{Z}$ is given by

$$\mathbb{E}_{\mathbf{z}}[\log P(\mathbf{Y}, \mathbf{Z})] = \sum_{u \in U} \sum_{k \in K} \gamma(z_{u,k}) (\log \pi_k + y_u \log q_{k,u} + (1 - y_u) \log(1 - q_{k,u})), \tag{5.1.4}$$

where $\gamma(z_{u,k}) = \mathbb{E}(z_{u,k})$. The maximization step determines the parameter values maximizing Equation (5.1.4). In this step, we should maximize the complete log-likelihood. Thus, we consider the derivative with respect to π_k and set it to 0, obtaining

$$\pi_k = \frac{\sum_{u \in U} \gamma(z_{u,k})}{|U|}. \tag{5.1.5}$$

To summarize the two steps of EM algorithm for the mixture of Bernoulli distributions :

E step : Compute $\gamma(z_{u,k})$ by using current parameter values of π_k^{old}

M step : Update π_k, i.e., obtain update parameter values π_k^{new} through Equation (5.1.5)

5.1.3 Classification by Mixture Rule-based Classifier

The classification of a object u is performed by

$$y_u = \begin{cases} 1, & \text{if } \sum_{k \in K} \pi_k q_{k,u} \geq \theta \\ 0, & \text{otherwise} \end{cases} \tag{5.1.6}$$

where θ is a threshold for classification taking a value in $[0, 1]$. In the experiments described later, we use $\theta = 0.5$.

5.2 Experiments

5.2.1 Dataset

We use four datasets obtained from UCI Machine Learning Repository [71]. The datasets are shown in Table 5.1, where $|U|$, $|C|$, and $|V_d|$ show the number of objects in the given data table, number of condition attributes, and number of decision classes treated in this experiment, respectively. Datasets *adult*, *default*, *german-credit*, and *nursery* are described in Section 4.2. We note that dataset *nursery* is not originally a binary classification dataset. Hence, we use 8,310 objects with the decision class as "priority" or "spec_prior".

5.2.2 Three Situations of Dataset

In this experiment, we assume that we can obtain two types of sets of rules : one from public domain and the other from dataset at hand. For convenience, the two organizations used in the numerical experiments are called "A" and "B". We assume that we only have set of rules R_B of organization "B" obtained from its public domain, although we have dataset $\mathcal{D}_A$ of organization "A". In our experiments, we generate four datasets by sampling objects from each dataset: dataset $\mathcal{D}_A$ of organization "A", dataset $\mathcal{D}_B$ of "B", dataset $\mathcal{D}_M$ for estimating the mixture ratio π_k, and dataset $\mathcal{D}_T$ for testing.

TABLE 5.1: Four datasets

dataset	$\|U\|$	$\|C\|$	$\|V_d\|$
adult	48844	8	2
default	30000	11	2
german-credit	1000	20	2
nursery	8310	8	2

TABLE 5.2: Selected attribute-value pairs

dataset	a	v_a
adult	sex	male
default	sex marital status	male : married / widowed
german-credit	sex	male
nursery	health	priority

Based on our problem setting, we assume that we do not know dataset $\mathcal{D}_B$ but a set of rules R_B induced from dataset $\mathcal{D}_B$. On the other hand, we know dataset $\mathcal{D}_A$ and also the set of rules R_A induced from dataset $\mathcal{D}_A$. Dataset $\mathcal{D}_T$ is used for evaluation of the usefulness of the approach. In the division, $\mathcal{D}_A$, $\mathcal{D}_B$, and $\mathcal{D}_M$ are composed of 30 % of data in a dataset, respectively. $\mathcal{D}_T$ is composed of the remaining 10 % of data. As both datasets $\mathcal{D}_A$ and $\mathcal{D}_B$ may be biased, we generate datasets for this experiment by assuming the following three situations :

1) Biased situation1 : We assume that distributions of $\mathcal{D}_A$ and $\mathcal{D}_B$ are significantly different. We consider a condition attribute a and an attribute value v_a. $\mathcal{D}_A$ is generated so that 85 % of their objects take v_a for condition attribute a and the remaining 15 % of their objects take other values for condition attribute a. $\mathcal{D}_B$ is generated so that 15 % of their objects take v_a for condition attribute a and the remaining 85 % of their objects take other values for condition attribute a. $\mathcal{D}_M$ and $\mathcal{D}_T$ are generated so that half of their objects take v_a for condition attribute a and the other half take other values for condition attribute a.

2) Biased situation2 : We assume that distributions of $\mathcal{D}_A$, $\mathcal{D}_M$, and $\mathcal{D}_T$ are similar but that of $\mathcal{D}_B$ is uniquely different from those of others. Then, we consider a condition attribute a and an attribute value v_a. $\mathcal{D}_A$, $\mathcal{D}_M$, and $\mathcal{D}_T$ are generated so that 85 % of their objects take v_a for condition attribute a and the remaining 15 % of their objects take other values for condition attribute a. $\mathcal{D}_B$ is generated so that 15 % of their objects take v_a for condition attribute a and the remaining 85 % of their objects take other values for condition attribute a.

3) Unbiased situation : $\mathcal{D}_A$, $\mathcal{D}_B$, $\mathcal{D}_M$, and $\mathcal{D}_T$ are generated by randomly selected objects from each dataset.

For the selection of $a \in C$ and v_a, many pairs may be conceivable. However, we consider only a pair for each dataset as shown in Table 5.2. These pairs are selected so as to have a sufficient number of objects matched to the attribute-value pair in each dataset. In each situation, we generated fifty sets of four datasets in order to enhance the experimental accuracy. Therefore, we obtain 150 sets for each dataset in total.

5.2.3 Alternative Methods

We use MLEM2 [25] for inducing rules in the experiments. An initial mixture ratio π_k obtained through Equation (5.1.5) for EM algorithm is 0.5. The number of iterations in EM Algorithm is set to 100. Threshold θ used in Equation (5.1.6) for classification is 0.5. To confirm the effectiveness of the proposed approach (EM), we compare the proposed method to the following four alternative rule-based classifiers :

1) Only A : A rule-based classifier based on LERS algorithm [22] using only a set of rules R_A.

2) Only B : A rule-based classifier based on LERS algorithm [22] using only a set of rules R_B.

3) A+B : A rule-based classifier based on LERS algorithm [22] using both sets of rules R_A and R_B.

4) A+B(α) : A rule-based classifier based on LERS algorithm [22] using a set of rules R_A and the largest α percent of rules in R_B selected according to support. We use $\alpha = 10, 30$, and 50 for all experiments.

5.2.4 Classification Performance

In the experiments, we evaluate the proposed method and alternative rule-based classifiers through classification accuracy and AUC (Area Under the Receiver Operating Characteristic Curve). The classification accuracy is the rate of the number of correctly classified objects to the total number of objects given to the rule-based classifier. AUC is known an evaluation index for a binary classifier and its value shows the size of area under the ROC Curve [10]. If a classifier could classify all objects accurately, the AUC takes the maximal value 1. If a classifier classified all objects randomly, the AUC takes 0.5. The averages and standard deviations of classification accuracy, AUC, and the number of rules are calculated. The obtained statistics are shown in Tables 5.3, 5.4, and 5.5. The proposed method is shown as "EM". The average *ave* and standard deviation *dev* are shown as $ave_{\pm dev}$ in three tables. The values in bold typeface represent the largest scores among all methods. In addition, the values with mark "$*$" show that they are statistically better than values obtained through the other methods by using a paired t-test at a level of 5 %.

TABLE 5.3: Averages and standard deviations of classification accuracy scores in all methods

situation and dataset		Only A	Only B	A+B	A+B(10)	A+B(30)	A+B(50)	EM
biased situation1	adult	$0.727^{*}_{\pm 0.006}$	$0.743_{\pm 0.007}$	$\mathbf{0.754_{\pm 0.006}}$	$0.729^{*}_{\pm 0.006}$	$0.744_{\pm 0.007}$	$0.753_{\pm 0.006}$	$0.744_{\pm 0.007}$
	default	$0.820^{*}_{\pm 0.008}$	$0.821^{*}_{\pm 0.009}$	$0.826_{\pm 0.008}$	$0.821^{*}_{\pm 0.008}$	$0.823_{\pm 0.008}$	$0.825_{\pm 0.008}$	$\mathbf{0.828_{\pm 0.008}}$
	german-credit	$0.698_{\pm 0.050}$	$0.692^{*}_{\pm 0.043}$	$0.698_{\pm 0.043}$	$0.700_{\pm 0.046}$	$0.698_{\pm 0.045}$	$0.698_{\pm 0.043}$	$\mathbf{0.713_{\pm 0.050}}$
	nursery	$0.935^{*}_{\pm 0.013}$	$0.934^{*}_{\pm 0.014}$	$\mathbf{0.972_{\pm 0.006}}$	$0.957^{*}_{\pm 0.008}$	$0.965_{\pm 0.006}$	$0.967_{\pm 0.006}$	$0.968_{\pm 0.008}$
biased situation2	adult	$0.730_{\pm 0.007}$	$0.711^{*}_{\pm 0.006}$	$\mathbf{0.742_{\pm 0.007}}$	$0.714^{*}_{\pm 0.006}$	$0.730_{\pm 0.007}$	$0.741_{\pm 0.007}$	$0.734_{\pm 0.011}$
	default	$0.832_{\pm 0.008}$	$0.810^{*}_{\pm 0.008}$	$0.825^{*}_{\pm 0.007}$	$0.813^{*}_{\pm 0.008}$	$0.818^{*}_{\pm 0.007}$	$0.821^{*}_{\pm 0.007}$	$\mathbf{0.835_{\pm 0.008}}$
	german-credit	$0.708_{\pm 0.036}$	$0.704^{*}_{\pm 0.041}$	$0.708_{\pm 0.036}$	$0.707^{*}_{\pm 0.042}$	$0.711_{\pm 0.042}$	$0.711_{\pm 0.039}$	$\mathbf{0.717_{\pm 0.035}}$
	nursery	$0.963_{\pm 0.021}$	$0.881^{*}_{\pm 0.038}$	$0.968_{\pm 0.019}$	$0.935^{*}_{\pm 0.029}$	$0.951^{*}_{\pm 0.023}$	$0.960^{*}_{\pm 0.021}$	$\mathbf{0.970_{\pm 0.020}}$
unbiased situation	adult	$0.771_{\pm 0.006}$	$0.772_{\pm 0.006}$	$\mathbf{0.784_{\pm 0.006}}$	$0.775_{\pm 0.006}$	$0.781_{\pm 0.006}$	$\mathbf{0.784_{\pm 0.006}}$	$0.779_{\pm 0.008}$
	default	$0.780^{*}_{\pm 0.007}$	$0.779^{*}_{\pm 0.008}$	$0.795_{\pm 0.008}$	$0.787_{\pm 0.007}$	$0.795_{\pm 0.008}$	$\mathbf{0.797_{\pm 0.007}}$	$0.790_{\pm 0.008}$
	german-credit	$0.700_{\pm 0.041}$	$0.690^{*}_{\pm 0.048}$	$0.700_{\pm 0.048}$	$\mathbf{0.707_{\pm 0.043}}$	$0.707_{\pm 0.044}$	$0.705_{\pm 0.046}$	$0.702_{\pm 0.048}$
	nursery	$0.958^{*}_{\pm 0.007}$	$0.959^{*}_{\pm 0.008}$	$0.973_{\pm 0.006}$	$0.960^{*}_{\pm 0.006}$	$0.964^{*}_{\pm 0.005}$	$0.966^{*}_{\pm 0.005}$	$\mathbf{0.978_{\pm 0.006}}$

TABLE 5.4: Averages and standard deviations of AUC scores in all methods

situation and dataset		Only A	Only B	A+B	A+B(10)	A+B(30)	A+B(50)	EM
biased situation1	adult	$0.768^{*}_{\pm 0.009}$	$0.745^{*}_{\pm 0.009}$	$0.759^{*}_{\pm 0.010}$	$0.774^{*}_{\pm 0.010}$	$0.778^{*}_{\pm 0.008}$	$0.774^{*}_{\pm 0.009}$	$\mathbf{0.786_{\pm 0.008}}$
	default	$0.701^{*}_{\pm 0.014}$	$0.703^{*}_{\pm 0.013}$	$0.718^{*}_{\pm 0.013}$	$0.696^{*}_{\pm 0.013}$	$0.696^{*}_{\pm 0.013}$	$0.701^{*}_{\pm 0.013}$	$\mathbf{0.732_{\pm 0.012}}$
	german-credit	$0.623_{\pm 0.056}$	$0.610_{\pm 0.072}$	$0.628_{\pm 0.053}$	$0.624_{\pm 0.059}$	$0.621_{\pm 0.062}$	$0.622_{\pm 0.055}$	$\mathbf{0.632_{\pm 0.070}}$
	nursery	$0.973^{*}_{\pm 0.009}$	$0.977^{*}_{\pm 0.008}$	$\mathbf{0.994_{\pm 0.003}}$	$0.983^{*}_{\pm 0.005}$	$0.988^{*}_{\pm 0.004}$	$0.990_{\pm 0.004}$	$0.992_{\pm 0.002}$
biased situation2	adult	$0.759^{*}_{\pm 0.011}$	$0.740^{*}_{\pm 0.010}$	$0.752^{*}_{\pm 0.009}$	$0.766^{*}_{\pm 0.010}$	$0.771^{*}_{\pm 0.010}$	$0.768^{*}_{\pm 0.009}$	$\mathbf{0.779_{\pm 0.008}}$
	default	$0.728^{*}_{\pm 0.015}$	$0.669^{*}_{\pm 0.018}$	$0.714^{*}_{\pm 0.016}$	$0.673^{*}_{\pm 0.017}$	$0.677^{*}_{\pm 0.017}$	$0.687^{*}_{\pm 0.017}$	$\mathbf{0.746_{\pm 0.014}}$
	german-credit	$0.632_{\pm 0.059}$	$0.612^{*}_{\pm 0.062}$	$0.630_{\pm 0.049}$	$0.634_{\pm 0.047}$	$0.636_{\pm 0.043}$	$0.636_{\pm 0.045}$	$\mathbf{0.637_{\pm 0.058}}$
	nursery	$0.990_{\pm 0.010}$	$0.946^{*}_{\pm 0.026}$	$0.991_{\pm 0.009}$	$0.967^{*}_{\pm 0.018}$	$0.978^{*}_{\pm 0.014}$	$0.984^{*}_{\pm 0.012}$	$\mathbf{0.995_{\pm 0.005}}$
unbiased situation	adult	$0.783^{*}_{\pm 0.009}$	$0.786^{*}_{\pm 0.009}$	$0.791^{*}_{\pm 0.007}$	$0.791^{*}_{\pm 0.008}$	$0.793^{*}_{\pm 0.008}$	$0.794^{*}_{\pm 0.007}$	$\mathbf{0.812_{\pm 0.008}}$
	default	$0.627^{*}_{\pm 0.013}$	$0.626^{*}_{\pm 0.013}$	$\mathbf{0.651_{\pm 0.011}}$	$0.634^{*}_{\pm 0.013}$	$0.641^{*}_{\pm 0.013}$	$0.645_{\pm 0.012}$	$0.648_{\pm 0.014}$
	german-credit	$0.630_{\pm 0.066}$	$0.625^{*}_{\pm 0.052}$	$0.631_{\pm 0.064}$	$0.636_{\pm 0.066}$	$0.630_{\pm 0.061}$	$0.628^{*}_{\pm 0.062}$	$\mathbf{0.653_{\pm 0.050}}$
	nursery	$0.988^{*}_{\pm 0.004}$	$0.989^{*}_{\pm 0.004}$	$\mathbf{0.995_{\pm 0.003}}$	$0.988^{*}_{\pm 0.004}$	$0.991_{\pm 0.003}$	$0.992_{\pm 0.003}$	$\mathbf{0.995_{\pm 0.002}}$

TABLE 5.5: Averages and standard deviations of the number of rules in all methods

situation and dataset		Only A	Only B	A+B	A+B(10)	A+B(30)	A+B(50)	EM
biased situation1	adult	$2232.88_{\pm 39.10}$	$1959.90_{\pm 36.51}$	$4192.78_{\pm 51.93}$	$2429.30_{\pm 39.05}$	$2821.24_{\pm 40.00}$	$3213.08_{\pm 42.19}$	-
	default	$1606.48_{\pm 28.77}$	$1606.36_{\pm 25.47}$	$3212.84_{\pm 36.50}$	$1767.56_{\pm 28.67}$	$2088.82_{\pm 29.03}$	$2409.98_{\pm 30.25}$	-
	german-credit	$83.66_{\pm 5.92}$	$96.20_{\pm 4.13}$	$179.86_{\pm 7.12}$	$93.68_{\pm 6.01}$	$112.96_{\pm 6.14}$	$132.04_{\pm 6.21}$	-
	nursery	$207.54_{\pm 9.70}$	$254.02_{\pm 12.32}$	$461.56_{\pm 13.88}$	$233.40_{\pm 9.50}$	$284.14_{\pm 9.63}$	$334.82_{\pm 10.26}$	-
biased situation2	adult	$2226.84_{\pm 36.17}$	$1977.42_{\pm 35.47}$	$4204.26_{\pm 47.52}$	$2425.04_{\pm 35.91}$	$2820.52_{\pm 36.48}$	$3215.82_{\pm 38.34}$	-
	default	$1601.12_{\pm 26.69}$	$1614.42_{\pm 32.00}$	$3215.54_{\pm 45.11}$	$1763.06_{\pm 27.43}$	$2085.90_{\pm 29.90}$	$2408.62_{\pm 33.46}$	-
	german-credit	$83.94_{\pm 5.27}$	$94.92_{\pm 4.21}$	$178.86_{\pm 6.42}$	$93.92_{\pm 5.28}$	$112.90_{\pm 5.33}$	$131.64_{\pm 5.48}$	-
	nursery	$205.46_{\pm 9.09}$	$251.80_{\pm 10.70}$	$457.26_{\pm 10.79}$	$231.06_{\pm 8.73}$	$281.48_{\pm 8.40}$	$331.56_{\pm 8.43}$	-
unbiased situation	adult	$2232.52_{\pm 42.40}$	$2235.62_{\pm 37.55}$	$4468.14_{\pm 55.33}$	$2456.52_{\pm 42.40}$	$2903.70_{\pm 43.38}$	$3350.58_{\pm 45.59}$	-
	default	$1869.32_{\pm 34.12}$	$1867.56_{\pm 33.52}$	$3736.88_{\pm 41.61}$	$2056.54_{\pm 33.54}$	$2430.08_{\pm 33.15}$	$2803.36_{\pm 34.18}$	-
	german-credit	$87.46_{\pm 5.45}$	$86.86_{\pm 5.79}$	$174.32_{\pm 6.96}$	$96.60_{\pm 5.28}$	$113.98_{\pm 5.32}$	$131.14_{\pm 5.51}$	-
	nursery	$250.14_{\pm 12.00}$	$251.38_{\pm 10.03}$	$501.52_{\pm 14.92}$	$275.72_{\pm 11.97}$	$326.08_{\pm 12.13}$	$376.08_{\pm 12.66}$	-

As shown in Tables 5.3 and 5.4, the proposed method (EM) provides the largest average classification accuracy scores compared with those of four alternative methods in datasets *default* and *german-credit* under biased situation1 and datasets *default*, *german-credit*, and *nursery* under biased situation2. Moreover, the proposed method provides the largest average AUC scores compared with those of obtained using the other four alternative methods in datasets *adult*, *default*, and *german-credit* under biased situation1 and in all datasets under biased situation2. The results showing that method "EM" outperforms method "A+B" imply the effect of combining two sets of rules by using EM Algorithm. However, method "A+B" provides the largest average classification accuracy in dataset *adult* under all situations. As shown in Table 5.5, a large number of rules is induced because of the size of $|U|$. This result shows that more number of rules are effective for classification when utilizing the published rules induced from a large dataset. Method "A+B" provides larger average classification accuracy scores than methods "Only A" and "Only B" in almost all datasets. This result indicates that the utilization of the published rules regardless of the situation is advantageous in classification. Some differences are observed in the classification between methods "Only A" and "Only B" in datasets *default* and *nursery* under biased situation2. This is caused by the difference of data distribution in datasets $\mathcal{D}_A$ and $\mathcal{D}_B$, respectively. The result shows that "EM" provides larger average classification accuracy scores than methods "Only A" and "Only B" by employing the computed mixture ratio effectively.

Based on the results obtained under unbiased situation shown in Tables 5.3 and 5.4, no significant differences are observed in classification accuracy in any pairs of the methods. The experimental results confirm that the proposed method "EM" is advantageous for obtaining better rule-based classifiers by utilizing the published sets of rules.

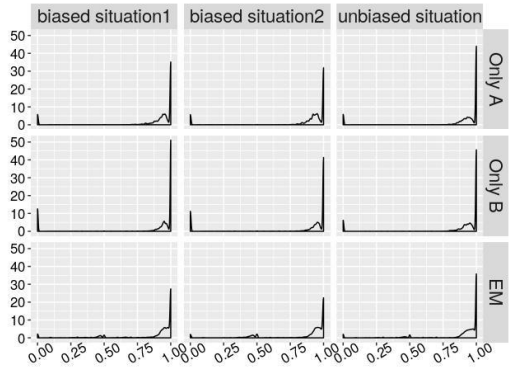
5.2.5 Results of Mixture Ratios and Adequacy Scores

The average and its standard deviation of the mixture ratio π_A obtained using EM algorithm are shown in Table 5.6, where π_A is the mixture ratio for the rule-based classifier using R_A. The mixture ratios in three datasets under biased situation2 are relatively large because $\mathcal{D}_T$ is generated using the same distribution on dataset. In contrast, the mixture ratios π_A under unbiased situation are nearly 0.5, implying that R_A and R_B are almost equally treated.

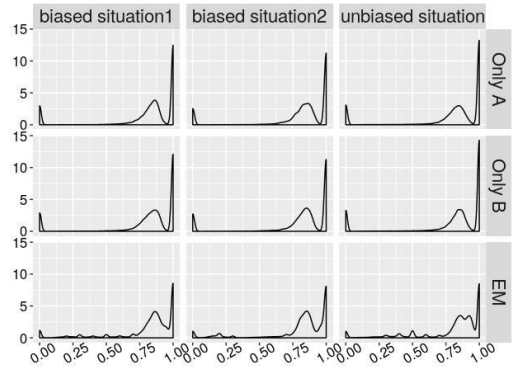
The average of Pearson's correlation coefficient and its standard deviation are shown in Table 5.7. As shown in the column of methods "Only A v.s. Only B" in Table 5.7, the coefficients between biased situation1 and situation2 are approximately 11.9 % and 13.2 % smaller than those in the unbiased situation, except dataset *default*. This result

TABLE 5.6: Averages and standard deviations of π_A

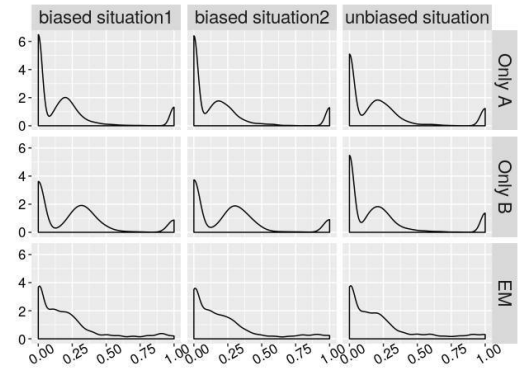
dataset	biased situation1	biased situation2	unbiased situation
adult	$0.494_{\pm 0.042}$	$0.506_{\pm 0.042}$	$0.500_{\pm 0.000}$
default	$0.512_{\pm 0.039}$	$0.516_{\pm 0.042}$	$0.504_{\pm 0.028}$
german-credit	$0.414_{\pm 0.333}$	$0.482_{\pm 0.343}$	$0.508_{\pm 0.319}$
nursery	$0.518_{\pm 0.069}$	$0.600_{\pm 0.049}$	$0.502_{\pm 0.112}$



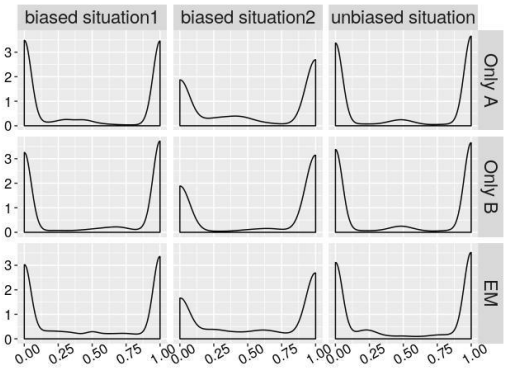
(A) adult



(B) default



(C) german-credit



(D) nursery

FIGURE 5.1: The frequency distributions of adequacy scores of decision class D_1

shows that the adequacy scores of decision class D_j in R_A and R_B are different in biased situations.

5.3 Conclusions

In this chapter we proposed the application of the mixture model for utilization of sets of rules obtained in a public domain. We applied the concept of the mixture model in order to utilize both the published rules and the rules induced from dataset at hand. By applying LERS algorithm to each set of rules, we obtained two normalized score

TABLE 5.7: Averages and standard deviations of Pearson's correlation coefficients with adequacy scores of decision class D_1

adult	Only A v.s. Only B	Only A v.s. EM	Only B v.s. EM
biased situation1	$0.295_{\pm 0.037}$	$0.769_{\pm 0.124}$	$0.795_{\pm 0.155}$
biased situation2	$0.293_{\pm 0.031}$	$0.830_{\pm 0.122}$	$0.749_{\pm 0.092}$
unbiased situation	$0.349_{\pm 0.034}$	$0.797_{\pm 0.113}$	$0.814_{\pm 0.109}$
default	Only A v.s. Only B	Only A v.s. EM	Only B v.s. EM
biased situation1	$0.384_{\pm 0.026}$	$0.805_{\pm 0.141}$	$0.810_{\pm 0.131}$
biased situation2	$0.385_{\pm 0.034}$	$0.970_{\pm 0.015}$	$0.592_{\pm 0.051}$
unbiased situation	$0.320_{\pm 0.032}$	$0.800_{\pm 0.098}$	$0.803_{\pm 0.098}$
german-credit	Only A v.s. Only B	Only A v.s. EM	Only B v.s. EM
biased situation1	$0.132_{\pm 0.096}$	$0.890_{\pm 0.148}$	$0.469_{\pm 0.239}$
biased situation2	$0.133_{\pm 0.122}$	$0.826_{\pm 0.250}$	$0.521_{\pm 0.276}$
unbiased situation	$0.149_{\pm 0.126}$	$0.635_{\pm 0.287}$	$0.736_{\pm 0.282}$
nursery	Only A v.s. Only B	Only A v.s. EM	Only B v.s. EM
biased situation1	$0.836_{\pm 0.027}$	$0.957_{\pm 0.013}$	$0.959_{\pm 0.010}$
biased situation2	$0.798_{\pm 0.053}$	$0.971_{\pm 0.013}$	$0.916_{\pm 0.029}$
unbiased situation	$0.917_{\pm 0.011}$	$0.978_{\pm 0.008}$	$0.979_{\pm 0.007}$

distributions over decision classes. EM algorithm was then used for estimating the mixture ratios for constructing a rule-based classifier. By using the four datasets frequently used in privacy preserving data mining and publishing, we considered three biased situations to generate datasets for the experiments. The classification performance of the proposed method was compared with those of other alternative methods. It was shown that the classification accuracy scores and AUC scores of the proposed mixture models are advantageous over the other alternative methods.

Chapter 6

Conclusions

In this thesis, we have studied on three approaches to the utilization of rules induced from datasets. The contributions of this thesis are summarized as follows.

In Chapter 3, we proposed robustness measures as a new type of interestingness measures for rules. We provided an example that demonstrates the necessity of a robustness measure, and then described several concepts of robustness. The robustness measures were defined depending on the evaluation attitude against uncertainty. The measures evaluate to what extent the interestingness of a rule is preserved against partially matched objects. Four numerical experiments using eight datasets obtained from UCI Machine Learning Repository [71] were conducted to examine the usefulness of the robustness measures and the difference between the robustness measures and recall measure. The results revealed that the robustness measure was an independent interestingness measure of recall and may be effective in selecting useful rules for design knowledge. Rules with extremely high robustness scores are useful for design knowledge, whereas rules with high recall scores are useful for constructing a rule-based classifier. Specifically, rules with very high robustness scores performed well in estimating a decision attribute value.

In Chapter 4, we proposed a rule clustering approach to privacy protection in publishing rules. We introduced k-anonymity to rule publication so that each rule covers at least k objects in the decision table at hand. In the proposed method, rules selected by the similarity and consistency measures were merged iteratively until k-anonymity is achieved with respect to all rules to be published. Using five datasets obtained from UCI Machine Learning Repository, the classification accuracy of the rule-based classifier and the degree of privacy protection of rules obtained by the proposed method were compared to three alternative methods. The results showed that the proposed method was advantageous over three alternative methods. Moreover, compared to the original

MLEM2, the set of rules obtained by the proposed method maintained an acceptable level of classification accuracy. Furthermore, the successful attack ratios to the rules obtained by the proposed method were comparable to those of the alternative methods. Thus, the rules obtained by the proposed method protected the privacy to an acceptable level.

In Chapter 5, we proposed a mixture model approach to constructing an effective rule-based classifier by utilizing published rules. We applied the concept of the mixture model in order to utilize both the published rules and the induced rules from dataset at hand. By applying LERS algorithm to each set of rules, we obtained adequacy scores over decision classes. EM algorithm was then used for estimating the mixture ratios for constructing a rule-based classifier. By using four datasets frequently used in privacy preserving data mining and publishing, we conducted numerical experiments under three different biased situations. Classification performance of the proposed method were compared to other alternative methods. It was shown that the classification accuracy scores and AUC scores of the proposed mixture models are advantageous over the other alternative methods.

As shown above in this thesis, we tackled three problems toward the utilization off rules induced from datasets : selection of good rules, privacy preservation for publishing rules, and utilization of published rules. The following research topics are expected as extensions of this studies. First, we may attempt to identify the reason that a rule-based classifier composed of rules with extremely high robustness scores does not perform well. A modified approach to selecting rules in order to construct a rule-based classifier using a robustness measure would be proposed. As another topic, we may select rules based on both robustness measures and other interestingness measures. The selected rules would be useful in constructing an effective rule-based classifier system.

Second, the improvement of the classification performance of the rule clustering approach to achieving k-anonymity will be an important future work. The consistency scores and the similarity scores of the rules to be merged decreases gradually as the number of iterations increases. This method is based on agglomerative hierarchical clustering algorithm so that their scores of rules to be combined may be very small values at the end of the iterations. This may cause a degradation of the performance of the proposed method. Therefore, we will address this disadvantage by developing other algorithms to improve the classification performance. In addition to the improvement, we will conduct experiments under various attack situations to verify the performance of the privacy preservation.

Third, the mixture model approach could be extended to a mixture multinomial distribution for multi-class classification problems. To verify the effectiveness, we would

compare the mixture model approach to ensemble learning techniques in previous studies. Moreover, the usefulness would be verified under various biased situations.

Finally, it will be an important task to deploy a decision support system using the proposed approaches to promote the utilization of rules induced from datasets in real-world problems. These studies in this thesis will be helpful for the utilization of induced rules.

Bibliography

[1] C. C. Aggarwal and P. S. Yu. *Privacy-Preserving Data Mining: Models and Algorithms*. Springer, 2008.

[2] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu. Approximation algorithms for k-anonymity. *Journal of Privacy Technology*, 2005.

[3] R. Agrawal, T. Imieliski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD international conference on Management of data*, pp. 207–216, 1993.

[4] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487–499, 1994.

[5] J.G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, and J. Wroblewski. Rough set algorithms in classification problems. *Transactions on Rough Sets IX*, pp. 49–88, 2000.

[6] M. J. Beynon and M. J. Peel. Variable precision rough set theory and data discretization : an application to corporate failure prediction. *Omega*, Vol. 29, pp. 561–576, 2001.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[8] J. Blaszczynski, R. Slowinski, and J. Stefanowski. Ordinal classification with monotonicity constraints by variable consistency bagging. In *Rough Sets and Current Trends in Computing, LNCS 6086*, pp. 392–401. Springer, Berlin, Heidelberg, 2010.

[9] J. Blaszczynski, R. Slowinski, and J. Stefanowski. Variable consistency bagging ensembles. In *Transactions on Rough Sets XI, LNCS 5946*, pp. 40–52. Springer, Berlin, Heidelberg, 2010.

[10] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, Vol. 30, No. 7, pp. 1145–1159, 1997.

[11] Y. L. Bras, P. Meyer, P. Lenca, and S. Lallich. A robustness mea- sure of association rules. In *Machine Learning and Knowledge Discovery in Databases, LNCS 6322*, pp. 227–242, 2010.

[12] L. Breiman. Bagging predictors. *Machine Learning*, Vol. 24, No. 2, pp. 123–140, 1996.

[13] S. Brin, R. Motwani, J. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *Proceedings of the ACM SIGMOD international conference on Management of data*, pp. 255–264, 1997.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38, 1977.

[15] A. I. Dimitras, R. Slowinski, R. Susmaga, and C. Zopounidis. Business failure prediction using rough sets. *European Journal of Operational Research*, Vol. 114, pp. 263–280, 1999.

[16] Q. Duan, D. Miao, and M. Chen. Web document classification based on rough set. In *Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, pp. 240 – 247, 2007.

[17] T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. Data mining using two-dimensional optimized association rules: Scheme, algorithms, and visualization. In *Proceedings of the ACM SIGMOD international conference on Management of data*, pp. 13–23, 1996.

[18] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, Vol. 42, No. 14, 2010.

[19] J. Furnkranz. Separate-and-conquer rule learning. *Artificial Intelligence Review*, Vol. 13, No. 1, pp. 3–54, 1999.

[20] L. Geng and H. J. Hamilton. Interestingness measure for data mining: A survey. *ACM Computing Surveys*, Vol. 38, No. 9, pp. 1–32, 2006.

[21] S. Greco, R. Slowinski, and I. Szczch. Properties of rule interestingness measure and alternative approaches to normalization of measures. *Information Sciences*, Vol. 216, pp. 1–16, 2012.

[22] J. W. Grzymala-Busse. Lers: A system for learning from examples based on rough sets. In *Intelligent Decision Support: Handbook of Applications and Advance of the Rough Sets Theory*, pp. 3–18. Kluwer Academic Publishers, 1992.

[23] J. W. Grzymala-Busse. A new version of the rule induction system lers. *Fundamenta Informaticae*, Vol. 31, No. 1, pp. 27–39, 1997.

[24] J. W. Grzymala-Busse. The rough set based rule induction technique for classification problems. In *Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT'98*, pp. 109–113, 1998.

[25] J. W. Grzymala-Busse. Mlem2: Discretization during rule induction. In *Proceedings of the IIPWM 2003*, pp. 499–508, 2003.

[26] J. W. Grzymala-Busse. An experimental comparison of three rough set approaches to missing attribute values. *Transactions on Rough Sets VI*, Vol. 1, , 2007.

[27] J. W. Grzymala-Busse, J. Stefanowski, and W. Ziarko. Rough sets: Facts versus misconceptions. *Informatica*, Vol. 20, pp. 455–464, 1996.

[28] S. Hajian, J. Domingo-Ferrer, A. Monreale, D. Pedreschi, and F. Giannotti. Discrimination- and privacy-aware patterns. *Data Mining and Knowledge Discovery*, Vol. 29, No. 6, pp. 1733–1782, 2015.

[29] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 1–12, 2000.

[30] M. Inuiguchi, T. Hamakawa, and S. Ubukata. Imprecise rules for data privacy. In *Rough Sets and Knowledge Technology, LNCS 9436*, pp. 129–139. Springer International Publishing, 2015.

[31] B. Ji-Won, K. Ashish, B. Elisa, and L Ninghui. Efficient k-anonymization using clustering techniques. In *12th International Conference on Database Systems for Advanced Applications*, Vol. 4443, pp. 188–200, 2007.

[32] A. Kawano, K. Honda, H. Kasugai, and A. Notsu. A greedy algorithm for k-member co-clustering and its applicability to collaborative filtering. In *17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems*, Vol. 22, pp. 477–484, 2013.

[33] B. J. Khyati, V. Jignesh, and R. P. Dhiren. A survey on association rule hiding methods. *International Journal of Computer Application*, Vol. 13, No. 82, pp. 20–25, 2013.

[34] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence*, Vol. 2, pp. 1137–1143, 1995.

[35] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron. Rough sets: A tutorial. In *Rough Fuzzy Hybridization: A new trend in decision making*, pp. 3–98, 1999.

[36] P. F. Lazarsfeld and N. W. Henry. *Latent Structure Analysis.* Houghton Mifflin, 1968.

[37] P. Lenca, P. Meyer, B. Vaillant, and S. Lallich. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operation Research*, Vol. 184, pp. 610–626, 2008.

[38] P. Lenca, B. Vaillant, and S. Lallich. On the robustness of association rules. In *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, pp. 596–601, 2006.

[39] K. Macgarry. A survey of interestingness measure for knowledge discovery. *The Knowledge Engineering Review*, pp. 1–24, 2005.

[40] G. J. McLachlan and D. Peel. *Finite Mixture Models.* Wiley, 2000.

[41] A. Meyerson and R Williams. On the complexity of optimal k-anonymity. In *Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2004.

[42] N. Mori, H. Tanaka, and K. Inoue. *Rough Sets and Kansei (in Japanese).* Keibundo, 2006.

[43] J. G. Narges and N. D. Mohammad. A survey on privacy preserving association rule mining. *Advances in Computer Science: an International Journal*, Vol. 4, No. 14, pp. 41–48, 2015.

[44] M. Ohki, T. Harada, and M. Inuiguchi. Decision rule visualization system for knowledge discovery by means of the rough set approach. *Journal of Japan Society for Fuzzy Theory and Systems*, Vol. 24, No. 2, pp. 660–670, 2012.

[45] M. Ohki and M. Inuiguchi. Measuring the effect of conditions to the conclusion in a decision rule (in Japanese). In *Proceedings of the 28th Fuzzy System Symposium*, pp. 294–297, 2012.

[46] M. Ohki and M. Inuiguchi. Robustness measure and its application to the analysis of decision rules (in Japanese). In *Proceedings of the 23rd Soft Science Workshop*, pp. 22–25, 2013.

[47] M. Ohki and M. Inuiguchi. Robustness measure of decision rule. In *Rough Sets and Knowledge Technology, LNCS 8171*, pp. 166–177. Springer International Publishing, 2013.

[48] M. Ohki and M. Inuiguchi. A clustering approach for decision rule publishing. In *Proceedings of the 7th International Symposium of Innovative Management, Information and Production, and The Thirteenth International Symposium on Management Engineering*, pp. 45–50, 2016.

[49] M. Ohki and M. Inuiguchi. Decision rule clustering for data publishing. The 19th Czech - Japan Seminar on Data Analysis and Decision Making under Uncertainty, pp. 104–111, 2016.

[50] M. Ohki and M. Inuiguchi. Decision rule clustering for k-anonymization (in Japanese). In *Proceedings of the 32nd Fuzzy System Symposium*, pp. 343–348, 2016.

[51] M. Ohki and M. Inuiguchi. Construction of fairness-aware rule classifier taking care of confidence degradation by eliminating discriminatory conditions (in Japanese). In *Proceedings of the 33rd Fuzzy System Symposium*, 2017.

[52] M. Ohki and M. Inuiguchi. A k-anonymous rule clustering approach for data publishing. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 21, No. 6, pp. 980–988, 2017.

[53] M. Ohki and M. Inuiguchi. Rule clustering for data privacy (in Japanese). In *Proceedings of the 60th Annual Conference of the Institute of Systems, Control and Information Engineers [CD-ROM]*, 362-5.pdf, 2016.

[54] M. Ohki, M. Inuiguchi, and T. Harada. Decision rule visualizaition for knowledge discovery by means of rough set approach. In *2011 IEEE International Conference on Granular Computing*, pp. 502–507, 2011.

[55] M. Ohki, E. Sekiya, and M. Inuiguchi. Role of robustness measure in rule induction. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 20, No. 4, pp. 580–589, 2016.

[56] M. Oki and M. Inuiguchi. A mixture model approach to utilizing published decision rules. In *Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*. IEEE, 2017.

[57] M. Oki and M. Inuiguchi. Rule classifier using bernoulli mixture model for utilizing published rules (in Japanese). *Journal of Japan Society for Fuzzy Theory and Systems*, to appear.

[58] M. Oki, K. Takeuchi, and Y. Uematsu. Mobile network failure event detection and forecasting with multiple user activity data sets. In *The Thirtieth Conference*

on Innovative Applications of Artificial Intelligence (IAAI-18). Association for the Advancement of Artificial Intelligence, to appear.

[59] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, Vol. 11, No. 5, pp. 341–356, 1982.

[60] Z. Pawlak and A. Skowron. Rudiments of rough sets. *Information Sciences*, Vol. 177, No. 1, pp. 3–27, 2007.

[61] G. P. Shapiro. Discovery, analysis and presentation of strong rules. In *Knowledge Discovery in Databases*, pp. 229–248, 1991.

[62] R. Slowinski, C. Zopounidis, and A. I. Dimitras. Prediction of company acquisition in greece by means of the rough set approach. *European Journal of Operational Research*, Vol. 100, pp. 1–15, 1997.

[63] Sebastian Stawicki, Dominik Slezak, Andrzej Janusz, and Sebastian Widz. Decision bireducts and decision reducts : a comparison. *International Journal of Approximate Reasoning*, Vol. 84, pp. 75–109, 2017.

[64] J. Stefanowski. On rough set based approaches to induction of decision rules. *Rough sets in knowledge discovery*, Vol. 1, No. 1, pp. 500–529, 1998.

[65] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty Fuzziness and Knowledge-based System*, Vol. 10, No. 5, pp. 557–570, 2002.

[66] C. H. Tai, P. S.Yu, and M. S. Chen. k-support anonymity based on pseudo taxonomy for outsourcing of frequent itemset mining. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 473–482, 2010.

[67] P. N. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 32–41, 2002.

[68] P. Tan, M. Steinbach, and V. Kumar. Classification : Alternative techniques. In *Introduction to data mining*, pp. 207–223, 2005.

[69] S. Tsumoto. Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model. *Information Sciences*, Vol. 162, pp. 65–80, 2004.

[70] S. Ubukata, T. Miyazaki, A. Notsu, K. Honda, and M. Inuiguchi. An ensemble learning approach based on rough set preserving the qualities of approximations. In *4th International Symposium on Integrated Uncertainty in Knowledge Modeling and Decision Making*, pp. 89–101, 2015.

[71] UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/index.php.

[72] B. Vaillant, S. Lallich, and P. Lenca. Modeling of the counter-examples and association rules interestingness measures behavior. In *International Conference on Data Mining*, pp. 132–137, 2006.

[73] N. Yamaguchi, M. Wu, M. Nakata, and H. Sakai. Application of rough set-based information analysis to questionnaire data. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 18, No. 6, pp. 953–961, 2014.

[74] J. Zakuski, R. Szoszkiewicz, J. Krysinski, and J. Stefanowski. Rough set theory and decision rules in data analysis of breast cancer patients. In *Transactions on Rough Sets*, pp. 375–391, 2004.

[75] Z. Zhu and W. Du. k-anonymous association rule hiding. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, pp. 305–309, 2010.

[76] W. Ziarko. Variable precision rough set model. *International Journal of Computer and System Sciences*, Vol. 46, No. 1, pp. 39–59, 1993.

Acknowledgements

Foremost, I would like to express my sincere gratitude to Professor Masahiro Inuiguchi, for his kind support, endless encouragements and valuable discussion. Without his help, this study would never be completed.

Besides my advisers, I am very grateful to Professor Toshimitsu Ushio and Professor Youji Iiguni for accepting to be a chair of my dissertation committee. I also would like to express my gratitude to them for giving a valuable suggestions and comments to complete this thesis.

My sincere thanks also goes to Mr. Eiji Sekiya, who conducted a part of numerical experiments to verify the performance of our proposed methods of this study.

Furthermore, I would like to express my special thanks to Associate professor Tatsushi Nishi, Assistant Professor Hirosato Seki, Assistant Professor Yoshifumi Kusunoki, and Assistant Professor Puchit Sariddichainunta of Osaka University for his kind supports, helpful advices, and friendly conversations. I am very thankful to all students in Inuguchi's Laboratory for their helpful supports. In addition, I am very grateful to Professor Toshinobu Harada of Wakayama Univerisity for teaching me an enjoyment of studying the topic of Rough Sets. I sincerely enjoyed the new step he gave me to plan my course in my career.

Apart from the academic circles, I wish to take this opportunity to express my gratitude to my managers, Satoshi Kamei and Yukio Uematsu of NTT Communications, for great supports at work and cheerful encouragements to my academic life. Further, I would like to offer special thanks to my co-workers for their encouragements.

Finally, I would like to thank my beloved parents and my wife for their cheerful encouragements and support. They encouraged and believed in me with great enthusiasm.

List of Publications

Refereed Journal Articles

- M. Oki and M. Inuiguchi. Rule classifier using bernoulli mixture model for utilizing published rules (in Japanese). *Journal of Japan Society for Fuzzy Theory and Systems*, to appear

- M. Ohki and M. Inuiguchi. A k-anonymous rule clustering approach for data publishing. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 21, No. 6, pp. 980–988, 2017

- M. Ohki, E. Sekiya, and M. Inuiguchi. Role of robustness measure in rule induction. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 20, No. 4, pp. 580–589, 2016

Refereed International Conference Papers

- M. Oki and M. Inuiguchi. A mixture model approach to utilizing published decision rules. In *Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*. IEEE, 2017

- M. Ohki and M. Inuiguchi. A clustering approach for decision rule publishing. In *Proceedings of the 7th International Symposium of Innovative Management, Information and Production, and The Thirteenth International Symposium on Management Engineering*, pp. 45–50, 2016

- M. Ohki and M. Inuiguchi. Decision rule clustering for data publishing. The 19th Czech - Japan Seminar on Data Analysis and Decision Making under Uncertainty, pp. 104–111, 2016

- M. Ohki and M. Inuiguchi. Robustness measure of decision rule. In *Rough Sets and Knowledge Technology, LNCS 8171*, pp. 166–177. Springer International Publishing, 2013

Domestic Conference Papers

- M. Ohki and M. Inuiguchi. Decision rule clustering for k-anonymization (in Japanese). In *Proceedings of the 32nd Fuzzy System Symposium*, pp. 343–348, 2016

- M. Ohki and M. Inuiguchi. Rule clustering for data privacy (in Japanese). In *Proceedings of the 60th Annual Conference of the Institute of Systems, Control and Information Engineers [CD-ROM]*, 362-5.pdf, 2016

- M. Ohki and M. Inuiguchi. Robustness measure and its application to the analysis of decision rules (in Japanese). In *Proceedings of the 23rd Soft Science Workshop*, pp. 22–25, 2013

- M. Ohki and M. Inuiguchi. Measuring the effect of conditions to the conclusion in a decision rule (in Japanese). In *Proceedings of the 28th Fuzzy System Symposium*, pp. 294–297, 2012

Awards

- SCI Young Researcher Award, 2017.5

- Student Excellent Paper Award, Thirteenth International Symposium on Management Engineering 2016, 2016.10

- Student-Presentation Award in Systems, Control and Information 2016, 2016.5

- 23th Soft Science Workshop Best Presentation Award, 2013.3

Others

- Refereed Journal Articles

  – M. Ohki, T. Harada, and M. Inuiguchi. Decision rule visualization system for knowledge discovery by means of the rough set approach. *Journal of Japan Society for Fuzzy Theory and Systems*, Vol. 24, No. 2, pp. 660–670, 2012

- Refereed International Conference Papers

  – M. Oki, K. Takeuchi, and Y. Uematsu. Mobile network failure event detection and forecasting with multiple user activity data sets. In *The Thirtieth Conference on Innovative Applications of Artificial Intelligence (IAAI-18)*. Association for the Advancement of Artificial Intelligence, to appear

  – M. Ohki, M. Inuiguchi, and T. Harada. Decision rule visualizaition for knowledge discovery by means of rough set approach. In *2011 IEEE International Conference on Granular Computing*, pp. 502–507, 2011

- Domestic Conference Papers
  - M. Ohki and M. Inuiguchi. Construction of fairness-aware rule classifier taking care of confidence degradation by eliminating discriminatory conditions (in Japanese). In *Proceedings of the 33rd Fuzzy System Symposium*, 2017
- Awards
  - JSAI Cup 2017 5th Place Winners, 2017.05
  - Nomura Research Institute Marketing Contest 2013 "Honorable Mention Award", Analysis of Consumption Behavior between Facebook and Twitter Users and Propose the advertisement strategy, 2013.11