A Comparison of Global and Saturated Probabilistic Approximations Using Characteristic Sets in Mining Incomplete Data

Patrick G. Clark
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA
Email: patrick.g.clark@gmail.com

Jerzy W. Grzymala-Busse
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA
Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, Rzeszow, Poland
Email: jerzy@ku.edu

Teresa Mroczek and Rafal Niemiec
Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, Rzeszow, Poland
Email: tmroczek@wsiz.rzeszow.pl and rniemiec@wsiz.rzeszow.pl

***Abstract*—Data mining systems form granules of information from data sets. Methods used to construct these granules can significantly impact the overall accuracy of the resulting model. In this paper, we study incomplete data sets with two interpretations of missing attribute values, lost values and "do not care" conditions, to determine the best method between two approaches and achieve the highest accuracy. For such incomplete data sets, we apply data mining based on two probabilistic approximations, global and saturated. The main objective of our paper is to compare both approaches in terms of an error rate, evaluated by ten-fold cross validation. Saturated probabilistic approximations are closer to the concept than global probabilistic approximations, so the corresponding error rate should be smaller. Using a 5% level of significance, our main result shows that there are differences between both approaches. However, in general neither is better for all data sets and thus, both approaches should be tried for each data set with the best selected for rule induction.**

Keywords–Data mining; rough set theory; probabilistic approximations; MLEM2 rule induction algorithm; lost values; "do not care" conditions.

I. INTRODUCTION

In this paper, we use two interpretations of a missing attribute value: lost values and "do not care" conditions. Lost values indicate that the original values were erased, and as a result we should use only existing, specified attribute values for rule induction. A lost value is denoted by "?". "Do not care" conditions mean that the missing attribute value may be replaced by any specified attribute value. A "do not care" condition is denoted by "*".

We use for data mining probabilistic approximations, a generalization of the idea of lower and upper approximations known in rough set theory. A probabilistic approximation is associated with a parameter (probability) α, if $\alpha = 1$, a probabilistic approximation is reduced to the lower approximation; if α is a small positive number, e.g., 0.001, the probabilistic approximation becomes the upper approximation. Usually, probabilistic approximations are applied to completely specified data sets [1]–[9]. Such approximations were generalized to incomplete data sets in [10].

Characteristic sets were introduced in [11] for incomplete data sets with any interpretation of missing attribute values. One of the methods being used in this paper, global probabilistic approximations, were introduced in [12]. This prior work expanded the ideas of characteristic sets and also studied their performance as a data mining tool. To further improve our methodology, we introduce a new idea of saturated probabilistic approximations. Our main objective is to compare both approaches in terms of an error rate, evaluated by ten-fold cross validation. The Modified Learning from Examples Module, version 2 (MLEM2) [13] was used for rule induction.

This paper starts with a discussion on incomplete data in Section II where we define attribute-value blocks and characteristic sets. In Section III, we present two types of probabilistic approximations, global and saturated. Section IV contains the details of our experiments. Finally, conclusions are presented in Section V.

II. INCOMPLETE DATA

We assume that the input data sets are presented in the form of a decision table. An example of a decision table is shown in Table I. Rows of the decision table represent cases, while columns are labeled by variables. The set of all cases will be denoted by U. In Table I, U = {1, 2, 3, 4, 5, 6, 7, 8}. Independent variables are called attributes and a dependent variable is called a decision and is denoted by d. The set of all attributes will be denoted by A. In Table I, A = {*Temperature*, *Headache*, *Cough*}. The value for a case x and an attribute a will be denoted by $a(x)$. For example, *Temperature*(1) = *high*.

The set X of all cases defined by the same value of the decision d is called a *concept*. For example, a concept associated with the value *yes* of the decision *Flu* is the set {1, 2, 3, 4}.

For a variable a and its value v, (a, v) is called a variable-value pair. A *block* of (a, v), denoted by $[(a, v)]$, is the set $\{x \in U \mid a(x) = v\}$ [14]. For incomplete decision tables, the

 ISBN: 978-1-61208-723-8

TABLE I. A DECISION TABLE

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	high	yes	*	yes
2	*	no	no	yes
3	very-high	*	yes	yes
4	high	no	*	yes
5	normal	yes	no	no
6	high	*	no	no
7	?	no	*	no
8	normal	yes	yes	no

definition of a block of an attribute-value pair is modified in the following way:

- if for an attribute a and a case x we have $a(x) = ?$, the case x should not be included in any blocks $[(a, v)]$ for all values v of attribute a;
- if for an attribute a and a case x we have $a(x) = *$, the case x should be included in blocks $[(a, v)]$ for all specified values v of attribute a.

For the data set from Table I, the blocks of attribute-value pairs are:

[(Temperature, normal)] = {2, 5, 8},
[(Temperature, high)] = {1, 2, 4, 6},
[(Temperature, very-high)] = {2, 3},
[(Headache, no)] = {2, 3, 4, 6, 7},
[(Headache, yes)] = {1, 3, 5, 6, 8},
[(Cough, no)] = {1, 2, 4, 5, 6, 7}, and
[(Cough, yes)] = {1, 3, 4, 7, 8}.

For a case $x \in U$ and $B \subseteq A$, the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where the set $K(x, a)$ is defined in the following way:

- if $a(x)$ is specified, then $K(x, a)$ is the block $[(a, a(x))]$ of attribute a and its value $a(x)$;
- if $a(x) = ?$ or $a(x) = *$, then $K(x, a) = U$.

For example, for Table I and $B = A$,

$$K(1, Temperature) = [(Temperature, high)] = \{1, 2, 4, 6\},$$

$$K(1, Headache) = [(Headache, yes)] = \{1, 3, 5, 6, 8\}$$

and

$$K(1, Cough) = U$$

so $K_A(1) = \{1, 2, 4, 6\} \cap \{1, 3, 5, 6, 8\} \cap U = \{1, 6\}$.

Similarly,

$K_A(2)$ = {2, 4, 6, 7},
$K_A(3)$ = {3},
$K_A(4)$ = {2, 4, 6},
$K_A(5)$ = {5},
$K_A(6)$ = {1, 2, 4, 6},
$K_A(7)$ = {2, 3, 4, 6, 7}, and
$K_A(8)$ = {8}.

III. Probabilistic Approximations

In this section, we will discuss two types of probabilistic approximations: global and saturated.

A. Global Probabilistic Approximations

An idea of the global probabilistic approximation, though restricted only to lower and upper approximations, was introduced in [15][16], and presented in a general form in [12]. Let X be a concept, $X \subseteq U$. A *B-global probabilistic approximation* of the concept X, based on characteristic sets, with the parameter α and denoted by $appr^{global}_{\alpha,B}(X)$ is defined as the following set

$$\bigcup \{K_B(x) \mid \exists\ Y \subseteq U\ \forall\ x \in Y,\ Pr(X|K_B(x)) \geq \alpha\}. \quad (1)$$

In general, for given sets B and X and the parameter α, there exist many B-global probabilistic approximations of X. Additionally, the algorithm for computing B-global probabilistic approximations is of exponential computational complexity. Therefore, we decided to use a heuristic version of the definition of B-global probabilistic approximation, called a MLEM2 B-global probabilistic approximation of the concept X, associated with a parameter α and denoted by $appr^{mlem2}_{\alpha,B}(X)$, [12]. This definition is based on the rule induction algorithm MLEM2 [17]. The MLEM2 algorithm is used in the Learning from Examples using Rough Sets (LERS) data mining system [17]–[19]. The approximation $appr^{mlem2}_{\alpha,B}(X)$ is constructed from characteristic sets $K_B(y)$, the most relevant to the concept X, i.e., with $|X \cap K_B(y)|$ as large as possible and $Pr(X|K_B(y)) \geq \alpha$, where $y \in U$. If more than one characteristic set $K_B(y)$ satisfies both conditions, we pick the characteristic set $K_B(y)$ with the largest $Pr(X|K_B(y))$. If this criterion ends up with a tie, a characteristic set is picked up heuristically, as the first on the list [12].

In this paper, we study MLEM2 B-global probabilistic approximations based on characteristic sets, with $B = A$, and calling them, for simplicity, *global probabilistic approximations* associated with the parameter α, denoted by $appr^{mlem2}_{\alpha}(X)$. Similarly, for $B = A$, the characteristic set $K_B(X)$ is denoted by $K(x)$.

Let $E_\alpha(X)$ be the set of all eligible characteristic sets defined as follows

$$\{K(x) \mid x \in U, Pr(X|K(x)) \geq \alpha\}. \quad (2)$$

A heuristic version of the MLEM2 global probabilistic approximation is computed using the algorithm specified in Figure 1.

For Table I, all distinct MLEM2 global probabilistic approximations are

$$appr^{mlem2}_{1}(\{1, 2, 3, 4\}) = \{3\},$$

$$appr^{mlem2}_{0.75}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6\},$$

$$appr^{mlem2}_{0.6}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6, 7\},$$

$$appr^{mlem2}_{1}(\{5, 6, 7, 8\}) = \{5, 8\},$$

$$appr^{mlem2}_{0.5}(\{5, 6, 7, 8\}) = \{2, 4, 5, 6, 7, 8\} \text{ and}$$

$$appr^{mlem2}_{0.4}(\{5, 6, 7, 8\}) = \{2, 3, 4, 5, 6, 7, 8\}.$$

 ISBN: 978-1-61208-723-8

MLEM2 global probabilistic approximation algorithm
input: a set X (a concept), a set $E_\alpha(X)$,
output: a set T ($appr_\alpha^{mlem2}(X)$)
begin
  $G := X$;
  $T := \emptyset$;
  $Y := E_\alpha(X)$;
  while $G \neq \emptyset$ **and** $Y \neq \emptyset$
    begin
      select a characteristic set $K(x) \in Y$
      such that $|K(x) \cap X|$ is maximum;
      if a tie occurs, select $K(x) \in Y$
      with the smallest cardinality;
      if another tie occurs, select the first $K(x)$;
      $T := T \cup K(x)$;
      $G := G - T$;
      $Y := Y - K(x)$
    end
end

Figure 1. MLEM2 Global Approximation Algorithm

Saturated probabilistic approximation algorithm
input: a set X (a concept), index m,
a set $E_i(x)$ for i = 1, 2,..., n and $x \in U$,
output: a set T ($appr_{\alpha_m}^{saturated}(X)$)
begin
  $T := \emptyset$;
  $Y_i(x)$:= $E_i(x)$ for all i = 1, 2,..., m and $x \in U$;
  for j = 1, 2,..., m **do**
    while $Y_j(x) \neq \emptyset$
      begin
        select a characteristic set $K(x) \in Y_j(x)$
        such that $|K(x) \cap X|$ is maximum;
        if a tie occurs, select the first $K(x)$;
        $Y_j(x) := Y_j(x) - K(x)$;
        if $(K(x) - T) \cap X \neq \emptyset$
          then $T := T \cup K(x)$;
        if $X \subseteq T$ **then exit**
      end
end

Figure 2. Saturated Probabilistic Approximation Algorithm

B. Saturated Probabilistic Approximations

Another heuristic version of the probabilistic approximation is based on selection of characteristic sets while giving higher priority to characteristic sets with larger conditional probability $Pr(X|K(x))$. Additionally, if the approximation covers all cases from the concept X, we stop adding characteristic sets.

Let X be a concept and let $x \in U$. Let us compute all conditional probabilities $Pr(X|K(x))$. Then, we sort the set

$$\{Pr(X|K(x)) \mid x \in U\}. \tag{3}$$

Let us denote the sorted list of such conditional probabilities by α_1, α_2,..., α_n, where α_1 is the largest. For any i = 1, 2,..., n, the set $E_i(x)$ is defined as follows

$$\{K(x) \mid x \in U, Pr(X|K(x) = \alpha_i\}. \tag{4}$$

If we want to compute a saturated probabilistic approximation, denoted by $appr_\alpha^{saturated}(X)$, for some α, $0 < \alpha \leq 1$, we need to identify the index m such that

$$\alpha_m \geq \alpha > \alpha_{m+1}, \tag{5}$$

where $m \in \{1, 2, ..., n\}$ and α_{n+1} = 0. Then, the saturated probabilistic approximation $appr_{\alpha_m}^{saturated}(X)$ is computed using the algorithm specified in Figure 2.

For Table I, all distinct saturated probabilistic approximations are

$appr_1^{saturated}(\{1, 2, 3, 4\}) = \{3\}$,

$appr_{0.75}^{saturated}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6\}$,

$appr_1^{saturated}(\{5, 6, 7, 8\}) = \{5, 8\}$ and

$appr_{0.5}^{saturated}(\{5, 6, 7, 8\}) = \{2, 4, 5, 6, 7, 8\}$.

Note that $appr_{0.6}^{mlem2}(\{1, 2, 3, 4\})$ covers the case 7 in spite of the fact that the case 7 is not a member of the concept {1, 2, 3, 4}. The set {1, 2, 3, 4, 6, 7} is not listed among saturated probabilistic approximations of the concept {1, 2, 3, 4}.

C. Rule Induction

Once the global and saturated probabilistic approximations associated with a parameter α are constructed, rule sets are induced using the rule induction algorithm based on another parameter, also interpreted as a probability, and denoted by β. This algorithm also uses MLEM2 principles [20], with the algorithm details shown in Figure 3.

For example, for Table I and α = β = 0.5, using the saturated probabilistic approximations, the MLEM2 rule induction algorithm induces the following rules:

(Cough, no) & (Headache, no) $\rightarrow$ (Flu, no),
(Temperature, normal) $\rightarrow$ (Flu, no),
(Temperature, high) $\rightarrow$ (Flu, yes) and
(Temperature, very-high) $\rightarrow$ (Flu, yes).

IV. Experiments

The goal of this research is to select the best approach for rule set induction in data mining. The data sets were chosen to represent various types of data (symbolic or numeric) in an attempt to find the best method. For our experiments, we used eight data sets that are available in the University of California at Irvine *Machine Learning Repository*.

For every data set, a template was created. Such a template was formed by replacing randomly 35% of existing specified attribute values by *lost values*. The same templates were used for constructing data sets with "do not care" conditions, by replacing "?"s with "$*$"s. The reason that these data sets were selected is because they represent a reasonable distribution of types of data and data set sizes to measure the impacts of the experiments. Furthermore, 35% is selected as a missing attribute percentage is because it is the maximum percentage that is able to be replaced with all records having at least one value specified.

 ISBN: 978-1-61208-723-8

MLEM2 rule induction algorithm
input: a set Y (an approximation of X) and a parameter β,
output: a set $\mathcal{T}$ (a rule set),
begin
$G := Y$;
$D := Y$;
$\mathcal{T} := \emptyset$;
$\mathcal{J} := \emptyset$;
while $G \neq \emptyset$
begin
$T := \emptyset$;
$T_s := \emptyset$;
$T_n := \emptyset$;
$T(G) := \{t \mid [t] \cap G \neq \emptyset\}$;
while $(T = \emptyset$ **or** $[T] \not\subseteq D)$ **and** $T(G) \neq \emptyset$
begin
select a pair $t = (a_t, v_t) \in T(G)$ with maximum of $|[t] \cap G|$; if a tie occurs, select a pair $t \in T(G)$ with the smallest cardinality of $[t]$; if another tie occurs, select first pair;
$T := T \cup \{t\}$;
$G := [t] \cap G$;
$T(G) := \{t \mid [t] \cap G \neq \emptyset\}$;
if a_t is symbolic {let V_{a_t} be the domain of a_t}
then
$T_s := T_s \cup \{(a_t, v) | v \in V_{a_t}\}$
else {a_t is numerical, let $t = (a_t, u..v)$} and $T_n := T_n \cup \{(a_t, x..y) \mid$ disjoint $x..y$ and $u..v\} \cup \{(a_t, x..y) \mid x..y \supseteq u..v\}$;
$T(G) := T(G) - (T_s \cup T_n)$;
end {while};
if $Pr(X \mid [T]) \geq \beta$
then
begin
$D := D \cup [T]$;
$\mathcal{T} := \mathcal{T} \cup \{T\}$;
end {then}
else $\mathcal{J} := \mathcal{J} \cup \{T\}$;
$G := D - \cup_{S \in \mathcal{T} \cup \mathcal{J}}[S]$;
end {while};
for each $T \in \mathcal{T}$ **do**
for each numerical attribute a_t with $(a_t, u..v) \in T$ **do**
while (T contains at least two different pairs $(a_t, u..v)$ and $(a_t, x..y)$ with the same numerical attribute a_t)
replace these two pairs with a new pair $(a_t,$ common part of $(u..v)$ and $(x..y))$;
for each $t \in T$ **do**
if $[T - \{t\}] \subseteq D$ **then** $T := T - \{t\}$;
for each $T \in \mathcal{T}$ **do**
if $\cup_{S \in (\mathcal{T} - \{T\})}[S] = \cup_{S \in \mathcal{T}}[S]$ **then** $\mathcal{T} := \mathcal{T} - \{T\}$;
end {procedure}.

Figure 3. MLEM2 Rule Induction Algorithm

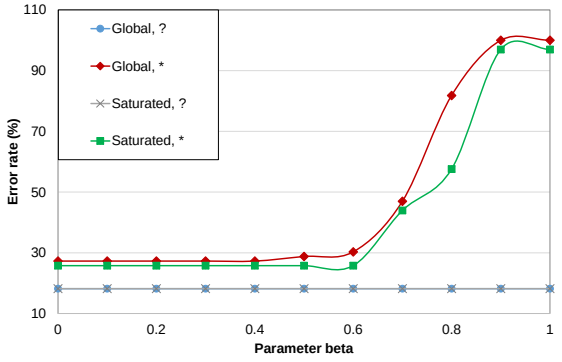

Figure 4. Error rate for the *Bankruptcy* data

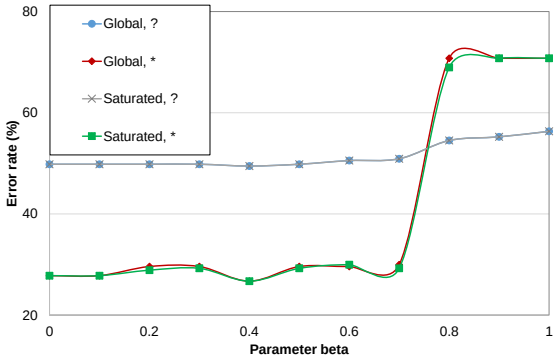

Figure 5. Error rate for the *Breast cancer* data set

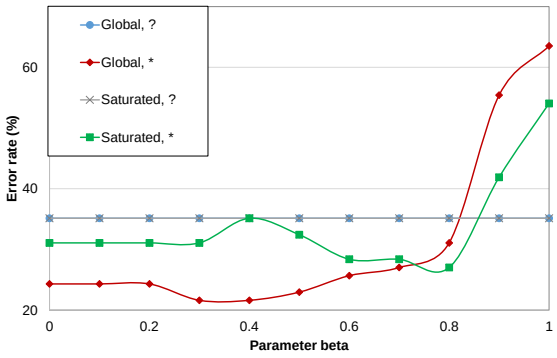

Figure 6. Error rate for the *Echocardiogram* data set

In our experiments, we used the MLEM2 rule induction algorithm. In all experiments, the parameter α was equal to 0.5. Results of our experiments are presented in Figures 4–11, where "Global" denotes a MLEM2 global probabilistic approximation, "Saturated" denotes a saturated probabilistic approximation, "?" denotes lost values and "$*$" denotes "do not care" conditions. In our experiments, four approaches for mining incomplete data sets were used, since we combined two options of probabilistic approximations: global and saturated with two interpretations of missing attribute values: lost and "do not care" conditions.

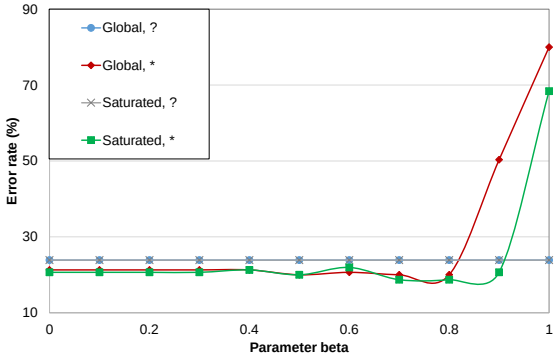

Figure 7. Error rate for the *Hepatitis* data set

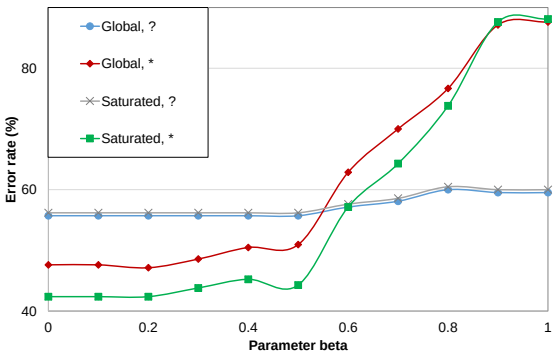

Figure 8. Error rate for the *Image segmentation* data set

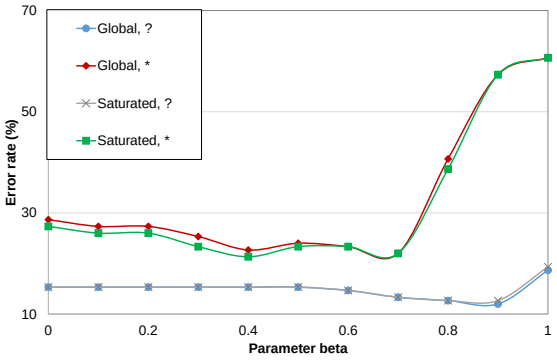

Figure 9. Error rate for the *Iris* data set

These four approaches were compared by applying the distribution free Friedman rank sum test and then by the post-hoc test (distribution-free multiple comparisons based on the Friedman rank sums), with a 5% level of significance. The null hypothesis H_0 of the Friedman test, claiming that differences between these approaches are insignificant, was rejected for *Breast cancer* and *Image recognition* as the only data sets. Results of the post-hoc distribution free all-treatments multiple comparisons Wilcoxon-Nemenyi-McDonald-Thompson test for the remaining six data sets are

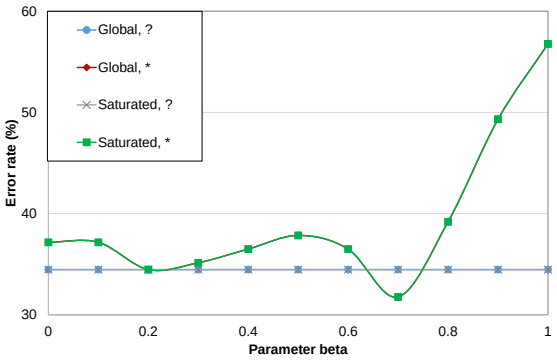

Figure 10. Error rate for the *Lymphography* data set

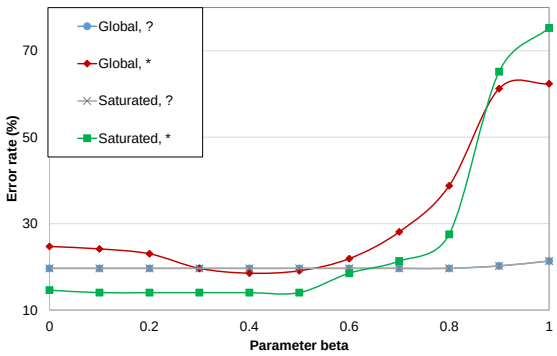

Figure 11. Error rate for the *Wine recognition* data set

TABLE II. RESULTS OF STATISTICAL ANALYSIS

Data set	Friedman test results (5% significance level)
Bankruptcy	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, *) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Iris	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, ?) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Lymphography	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, ?) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Echocardiogram	(Global, ?) is better than (Global, *) (Saturated, ?) is better than (Global, *)
Hepatitis	(Global, ?) is better than(Saturated, *) (Saturated, ?) is better than (Saturated, *)
Wine recognition	(Global, *) is better than (Saturated, *)

presented in Table II. This table is divided into three parts. In the first part, for *Bankruptcy*, *Iris* and *Lymphography* data sets, data mining approaches based on "do not care" conditions are always better than approaches based on lost values. In the second part, for *Echocardiogram* and *Hepatitis*, it is the other way around. In the third part, for *Wine recognition*, the only conclusion is that for "do not care" conditions global probabilistic approximations are better than saturated ones. Due to the varying characteristics of the data, as follows from the experimental results, no particular combination was identified as the best for all situations.

V. Conclusions and Future Work

We compared four approaches for mining incomplete data sets, combining two interpretations of missing attribute values with two types of probabilistic approximations. Our criterion of quality was an error rate computed as a result of ten-fold cross-validation. As follows from our experiments, there were significant differences between the four approaches. However, the best approach, associated with the smallest error rate, depends on a specific data set. Thus, for a given data set,

 ISBN: 978-1-61208-723-8

the base approach to mining must be selected by running experiments involved with all four approaches.

In future work, we will continue to study the experimental effects of saturation in an effort to form a theoretical basis for our ideas. Furthermore, these ideas will be extended to Maximal Consistent Blocks [21][22] to include the concepts of saturation with a thorough measurement of the experimental performance.

REFERENCES

[1] J. W. Grzymala-Busse and W. Ziarko, "Data mining based on rough sets," in Data Mining: Opportunities and Challenges, J. Wang, Ed. Hershey, PA: Idea Group Publ., 2003, pp. 142–173.

[2] Z. Pawlak and A. Skowron, "Rough sets: Some extensions," Information Sciences, vol. 177, 2007, pp. 28–40.

[3] Z. Pawlak, S. K. M. Wong, and W. Ziarko, "Rough sets: probabilistic versus deterministic approach," International Journal of Man-Machine Studies, vol. 29, 1988, pp. 81–95.

[4] D. Ślęzak and W. Ziarko, "The investigation of the bayesian rough set model," International Journal of Approximate Reasoning, vol. 40, 2005, pp. 81–91.

[5] S. K. M. Wong and W. Ziarko, "INFER—an adaptive decision support system based on the probabilistic approximate classification," in Proceedings of the 6-th International Workshop on Expert Systems and their Applications, 1986, pp. 713–726.

[6] Y. Y. Yao, "Probabilistic rough set approximations," International Journal of Approximate Reasoning, vol. 49, 2008, pp. 255–271.

[7] Y. Y. Yao and S. K. M. Wong, "A decision theoretic framework for approximate concepts," International Journal of Man-Machine Studies, vol. 37, 1992, pp. 793–809.

[8] W. Ziarko, "Variable precision rough set model," Journal of Computer and System Sciences, vol. 46, no. 1, 1993, pp. 39–59.

[9] ——, "Probabilistic approach to rough sets," International Journal of Approximate Reasoning, vol. 49, 2008, pp. 272–284.

[10] J. W. Grzymala-Busse, "Generalized parameterized approximations," in Proceedings of the 6-th International Conference on Rough Sets and Knowledge Technology, 2011, pp. 136–145.

[11] ——, "Rough set strategies to data with missing attribute values," in Notes of the Workshop on Foundations and New Directions of Data Mining, in conjunction with the Third International Conference on Data Mining, 2003, pp. 56–63.

[12] P. G. Clark, C. Gao, J. W. Grzymala-Busse, T. Mroczek, and R. Niemiec, "A comparison of concept and global probabilistic approximations based on mining incomplete data," in Proceedings of ICIST 2018, the International Conference on Information and Software Technologies, 2018, pp. 324–335.

[13] P. G. Clark and J. W. Grzymala-Busse, "Experiments on rule induction from incomplete data using three probabilistic approximations," in Proceedings of the 2012 IEEE International Conference on Granular Computing, 2012, pp. 90–95.

[14] J. W. Grzymala-Busse, "LERS—a system for learning from examples based on rough sets," in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, R. Slowinski, Ed. Dordrecht, Boston, London: Kluwer Academic Publishers, 1992, pp. 3–18.

[15] J. W. Grzymala-Busse and W. Rzasa, "Local and global approximations for incomplete data," in Proceedings of the Fifth International Conference on Rough Sets and Current Trends in Computing, 2006, pp. 244–253.

[16] ——, "Local and global approximations for incomplete data," Transactions on Rough Sets, vol. 8, 2008, pp. 21–34.

[17] J. W. Grzymala-Busse, "MLEM2: A new algorithm for rule induction from imperfect data," in Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2002, pp. 243–250.

[18] P. G. Clark and J. W. Grzymala-Busse, "Experiments on probabilistic approximations," in Proceedings of the 2011 IEEE International Conference on Granular Computing, 2011, pp. 144–149.

[19] J. W. Grzymala-Busse, "A new version of the rule induction system LERS," Fundamenta Informaticae, vol. 31, 1997, pp. 27–39.

[20] J. W. Grzymala-Busse, P. G. Clark, and M. Kuehnhausen, "Generalized probabilistic approximations of incomplete data," International Journal of Approximate Reasoning, vol. 132, 2014, pp. 180–196.

[21] Y. Leung and D. Li, "Maximal consistent block technique for rule acquisition in incomplete information systems," Information Sciences, vol. 153, 2003, pp. 85–106.

[22] P. G. Clark, C. Gao, J. W. Grzymala-Busse, and T. Mroczek, "Characteristic sets and generalized maximal consistent blocks in mining incomplete data," in Proceedings of the International Joint Conference on Rough Sets, part 1, 2017, pp. 477–486.

 ISBN: 978-1-61208-723-8