A Comparison of Two MLEM2 Rule Induction Algorithms Applied to Data with Many Missing Attribute Values

Patrick G. Clark and Cheng Gao

Department of Electrical Engineering
and Computer Science,
University of Kansas
Lawrence, KS, USA
Email: patrick.g.clark@gmail.com
cheng.gao@ku.edu

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

***Abstract*—We present results of novel experiments, conducted on 18 data sets with many missing attribute values, interpreted as lost values, attribute-concept values and "do not care" conditions. The main objective was to compare two versions of the Modified Learning from Examples, version 2 (MLEM2) rule induction algorithm, emulated and true, using concept probabilistic approximations. Our secondary objective was to check which interpretation of missing attribute values provides the smallest error rate, computed as a result of ten-fold cross validation. Results of our experiments show that both versions of the MLEM2 rule induction algorithms do not differ much. On the other hand, there is some evidence that the lost value interpretation of missing attribute values is the best: in seven cases this interpretation was significantly better (with 5% of significance level, two-tailed test) than attribute-concept values, and in eight cases it was better than "do not care" conditions. Additionally, attribute-concept values and "do not care" conditions were never significantly better than lost values.**

Keywords–Probabilistic approximations; generalization of probabilistic approximations; concept probabilistic approximations; true MLEM2 algorithm; emulated MLEM2 algorithm.

I. INTRODUCTION

Lower and upper approximations are basic ideas of rough set theory. Probabilistic approximations, associated with a probability α, are a generalization of that idea. If α = 1, the probabilistic approximation is identical with the lower approximation, if α is a very small positive number, the probabilistic approximation is identical with the upper approximation. Probabilistic approximations, for completely specified data sets, were studied, e.g., in [1]–[9]. Probabilistic approximations were additionally generalized to describe incomplete data sets in [10]. Experimental research associated with such probabilistic approximations was initiated in [11][12].

In this paper, missing attribute values are interpreted as *lost values*, *attribute-concept values*, and *"do not care" conditions*. A lost value is denoted by "?", an attribute-concept value is denoted by "−", and a "do not care" condition is denoted by "*". With lost values we assume that the original attribute value was erased, and that we should induce rules from existing, specified attribute values. With attribute-concept value we assume that such missing attribute values may be replaced by any actual attribute value restricted to the concept to which the case belongs. For example, if our concept is a specific disease, an attribute is a diastolic pressure, and all patients affected by the disease have high or very high diastolic pressure, a missing attribute value of the diastolic pressure for a sick patient will be high or very-high. With the third interpretation of missing attribute values, the "do not care" condition, we assume that it does not matter what is the attribute value. Such value may be replaced by any value from the set of all possible attribute values.

For any concept X and probability α, its probabilistic approximations may be computed directly from corresponding definitions and implemented as a new program. The output of this program may be used as an input to an implementation of the MLEM2 algorithm. The respective MLEM2 rule induction algorithm will be called a *true* MLEM2 algorithm.

Another possibility is to use the existing data mining system Learning from Examples using Rough Set theory (LERS). LERS computes standard lower and upper approximations for any concept. In LERS there exists a component that implements the MLEM2 rule induction algorithm. This component may be used to compute possible rules from the probabilistic approximation of X. Some modification of the strength of induced rules is required. This approach will be called an *emulated* MLEM2 algorithm. It is easier to implement since all what we need to do is to compute the probabilistic approximation of X and to modify strengths. The main part, the MLEM2 rule induction algorithm, does not need to be separately implemented. The idea of the emulated MLEM2 algorithm was introduced in [13] and further developed in [14][15].

Experiments conducted on eight incomplete data sets with 35% of missing attribute values to compare true version of the MLEM2 rule induction algorithm with the emulated one were reported in [16]. All three interpretations of missing attribute values were used in experiments, so experiments were conducted on 24 data sets. In these experiments true and emulated versions of the MLEM2 algorithm were compared

 ISBN: 978-1-61208-486-2

using the resulting classification error rate of the induced rules against the ten-fold cross validated data set as the quality criterion. Results were inconclusive. For six data sets, for all values of the parameter α, results were identical; for other 14 data sets results did not differ significantly (we used the Wilcoxon matched-pairs signed rank test, 5% significance level, two-tailed test). For three other data sets, the true MLEM2 algorithm was better than emulated, for remaining one data set the emulated MLEM2 algorithm was better than the true one.

Usually, experiments conducted on data sets with many missing attribute values provide for more conclusive results. As a result our first objective in this paper was to conduct new experiments with data sets that contain more than 35% missing attribute values. We used three interpretations of missing attribute values, resulting in 18 combinations. Results of the same comparison of error rate of the induced rules are measurably more conclusive: in five combinations (out of 18) the emulated approach to MLEM2 algorithm was better, in one case the true approach was better (5% significance level, two-tailed test).

Our second objective was to check which interpretation of missing attribute values should be used to accomplish a lower error rate. There is some evidence that the lost value interpretation of missing attribute values is the best: in seven cases this interpretation was significantly better (with 5% of significance level, two-tailed test) than attribute-concept values, and in eight cases it was better than “do not care” conditions. Additionally, attribute-concept values and “do not care” conditions were never significantly better than lost values.

In Sections II and III background information on incomplete data sets and probabilistic approximations is covered. Section IV describes the two algorithms used in our rule induction experiments and Section V explains the experimental setup with our results. Finally we provide concluding remarks in Section VI.

II. Incomplete Data Sets

An example of incomplete data set is presented in Table I. In Table I, the set A of all attributes consists of three variables *Wind*, *Humidity* and *Temperature*. A *concept* is a set of all cases with the same decision value. There are two concepts in Table I, the first one contains cases 1, 2, 3 and 4 and is characterized by the decision value *yes* of decision *Trip*. The other concept contains cases 5, 6, 7 and 8 and is characterized by the decision value *no*.

The fact that an attribute a has the value v for the case x will be denoted by $a(x) = v$. The set of all cases will be denoted by U. In Table I, $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

For complete data sets, an attribute-value pair $(a,\ v) = t$, a *block* of t, denoted by $[t]$, is a set of all cases from U such that attribute a has a value v. An *indiscernibility relation* R on U is defined for all $x, y \in U$ by

$$xRy \text{ if and only if } a(x) = a(y) \text{ for all } a \in A.$$

For incomplete decision tables the definition of a block of an attribute-value pair must be modified in the following way [17][18]:

- If for an attribute a there exists a case x such that $a(x) =?$, i.e., the corresponding value is lost, then the case x should not be included in any blocks $[(a, v)]$ for all values v of attribute a,
- If for an attribute a there exists a case x such that the corresponding value is an attribute-concept value, i.e., $a(x) = -$, then the corresponding case x should be included in blocks $[(a, v)]$ for all specified values $v \in V(x, a)$ of attribute a, where

$$V(x,a) = \{a(y) \mid a(y) \text{ is specified}, \ y \in U, \ d(y) = d(x)\}, \quad (1)$$

and d is the decision.
- If for an attribute a there exists a case x such that the corresponding value is a “do not care” condition, i.e., $a(x) = *$, then the case x should be included in blocks $[(a, v)]$ for all specified values v of attribute a.

TABLE I. AN INCOMPLETE DATA SET

	Attributes			Decision
Case	Wind	Humidity	Temperature	Trip
1	low	*	low	yes
2	*	low	–	yes
3	high	low	low	yes
4	low	*	*	yes
5	high	high	high	no
6	?	–	high	no
7	low	low	*	no
8	high	?	low	no

For a case $x \in U$ the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where B is a subset of the set A of all attributes and 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 the set $K(x, a) = U$,
- If $a(x) = -$, then the corresponding set $K(x, a)$ is equal to the union of all blocks of attribute-value pairs (a, v), where $v \in V(x, a)$ if $V(x, a)$ is nonempty. If $V(x, a)$ is empty, $K(x, a) = U$.

The characteristic set $K_B(x)$ may be interpreted as the set of cases that are indistinguishable from x using all attributes from B and using a given interpretation of missing attribute values.

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

$[(Wind, low)] = \{1, 2, 4, 7\}$,
$[(Wind, high)] = \{2, 3, 5, 8\}$,
$[(Humidity, low)] = \{1, 2, 3, 4, 6, 7\}$,
$[(Humidity, high)] = \{1, 4, 5, 6\}$,
$[(Temperature, low)] = \{1, 2, 3, 4, 7, 8\}$,
$[(Temperature, high)] = \{4, 5, 6, 7\}$.

For Table I, $V(2, Temperature) = \{low\}$ and $V(6, Humidity) = \{low, high\}$.

The corresponding characteristic sets are

TABLE II. CONDITIONAL PROBABILITIES

Case x	Characteristic set $K_A(x)$	$Pr(\{1,2,3,4\} \mid K_A(x))$
1	{1, 2, 4, 7}	0.75
2	{1, 2, 3, 4, 7}	0.8
3	{2, 3}	1
4	{1, 2, 4, 7 }	0.75
5	{5}	0
6	{4, 5, 6, 7}	0.25
7	{1, 2, 4, 7}	0.75
8	{2, 3, 8}	0.667

$$\begin{aligned} K_A(1) &= [(Wind, low)] \cap [(Humidity, *)] \cap [(Temp, low)] \\ &= \{1,2,4,7\} \cap U \cap \{1,2,3,4,7,8\} \\ &= \{1,2,4,7\}, \end{aligned}$$

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

III. Probabilistic Approximations

For incomplete data sets there exist a number of different definitions of approximations, in this paper we will use only *concept* approximations, we will skip the word *concept*.

The B-*lower approximation* of X, denoted by $\underline{appr}(X)$, is defined as follows

$$\cup \{K_B(x) \mid x \in X, K_B(x) \subseteq X\}. \quad (2)$$

Such lower approximations were introduced in [17][19].

The B-*upper approximation* of X, denoted by $\overline{appr}(X)$, is defined as follows

$$\begin{aligned} &\cup \{K_B(x) \mid x \in X, K_B(x) \cap X \neq \emptyset\} \\ &= \cup \{K_B(x) \mid x \in X\}. \end{aligned} \quad (3)$$

These approximations were studied in [17][19][20].

For incomplete data sets there exist a few definitions of probabilistic approximations, we will use only *concept* probabilistic approximations, again, we will skip the word *concept*.

A B-probabilistic approximation of the set X with the threshold α, $0 < \alpha \leq 1$, denoted by B-$appr_\alpha(X)$, is defined as follows

$$\cup\{K_B(x) \mid x \in X,\ Pr(X|K_B(x)) \geq \alpha\}, \quad (4)$$

where $Pr(X|K_B(x)) = \frac{|X \cap K_B(x)|}{|K_B(x)|}$ is the conditional probability of X given $K_B(x)$. A-probabilistic approximations of X with the threshold α will be denoted by $appr_\alpha(X)$.

For Table I and the concept X = [(*Trip*, *yes*)] = {1, 2, 3, 4}, for any characteristic set $K_A(x)$, $x \in U$, all conditional probabilities $P(X|K_A(x))$ are presented in Table II.

There are five distinct conditional probabilities $Pr(\{1,2,3,4\} \mid K_A(x))$, $x \in U$: 0.25, 0.667, 0.75, 0.8 and 1. Therefore, there exist at most five distinct probabilistic approximations of {1, 2, 3, 4} (in our example, there are only two distinct probabilistic approximations of {1, 2, 3, 4}). A probabilistic approximation $appr_\beta(\{1,2,3,4\})$, with $\beta > 0$ and not listed below, is equal to the closest probabilistic approximation $appr_\alpha(\{1,2,3,4\})$ with α larger than or equal to β. For example, $appr_{0.7}(\{1,2,3,4\}) = appr_{0.8}(\{1,2,3,4\})$. For Table I, all distinct probabilistic approximations are

$$\begin{aligned} appr_{0.8}(\{1,2,3,4\}) &= K_B(2) \cup K_B(3) \\ &= \{1,2,3,4,7\} \cup \{2,3\} \\ &= \{1,2,3,4,7\}, \\ appr_1(\{1,2,3,4\}) &= K_B(3) = \{2,3\}, \\ appr_{0.25}(\{5,6,7,8\}) &= \{1,2,3,4,5,6,7,8\}, \\ appr_{0.333}(\{5,6,7,8\}) &= \{2,3,4,5,6,7,8\}, \\ appr_{0.75}(\{5,6,7,8\}) &= \{4,5,6,7\}, \\ appr_1(\{5,6,7,8\}) &= \{5\}. \end{aligned}$$

IV. Rule Induction

In this section we will discuss two different ways to induce rule sets using probabilistic approximations: true MLEM2 and emulated MLEM2.

A. True MLEM2

In the true MLEM2 approach, for a given concept X and parameter α, first we compute the probabilistic approximation $appr_\alpha(X)$. The set $appr_\alpha(X)$ is a union of characteristic sets, so it is globally definable [21]. Thus, we may use the MLEM2 strategy to induce rule sets [22][23] by inducing rules directly from the set $appr_\alpha(X)$. For example, for Table I, for the concept [(*Trip*, *no*)] = {5, 6, 7, 8} and for the probabilistic approximation $appr_{0.75}(\{5,6,7,8\}) = \{4,5,6,7\}$, using the true MLEM2 approach, the following single rule is induced

1, 3, 4

(Temperature, high) -> (Trip, no).

Rules are presented in the LERS format, every rule is associated with three numbers: the total number of attribute-value pairs on the left-hand side of the rule, the total number of cases correctly classified by the rule during training, and the total number of training cases matching the left-hand side of the rule, i.e., the rule domain size.

B. Emulated MLEM2

We will discuss how the existing rough set based data mining systems, such as LERS, may be used to induce rules using probabilistic approximations. All what we need to do, for every concept, is to modify the input data set, run LERS, and then edit the induced rule set [14]. We will illustrate this procedure by inducing a rule set for Table I and the concept [(*Trip*, *no*)] = {5, 6, 7, 8} using the probabilistic approximation $appr_{0.75}(\{5,6,7,8\}) = \{4,5,6,7\}$. First, a new data set should be created in which for all cases that are members of the set $appr_{0.75}(\{5,6,7,8\})$ the decision values are copied from the original data set (Table I). For all remaining cases, those not being in the set $appr_{0.75}(\{5,6,7,8\})$, a new decision value

TABLE III. A PRELIMINARY MODIFIED DATA SET

	Attributes			Decision
Case	Wind	Humidity	Temperature	Trip
1	low	*	low	SPECIAL
2	*	low	–	SPECIAL
3	high	low	low	SPECIAL
4	low	*	*	yes
5	high	high	high	no
6	?	–	high	no
7	low	low	*	no
8	high	?	low	SPECIAL

is introduced. In our experiments the new decision value was named SPECIAL. Thus a new data set is created, see Table III.

This data set is input into the LERS data mining system. The concept [(*Trip*, *no*)], computed from Table III, is {5, 6, 7}. The LERS system computes the concept upper concept approximation of the set {5, 6, 7} to be {1, 2, 4, 5, 6, 7}, and using this approximation, computes the corresponding final modified data set. The MLEM2 algorithm induces the following preliminary rule set from the final modified data sets

1, 4, 6
(Temperature, low) -> (Trip, SPECIAL)
1, 1, 4
(Wind, low) -> (Trip, yes)
1, 1, 4
(Wind, low) -> (Trip, no)
1, 2, 4
(Humidity, high) -> (Trip, no)

where the three numbers that precede every rule are computed from Table III. Because we are inducing rules for the approximation from (Trip, no) ({5, 6, 7}), only the last two rules

1, 1, 4
(Wind, low) -> (Trip, no)
1, 2, 4
(Humidity, high) -> (Trip, no)

should be saved and the remaining two rules should be deleted in computing the final rule set.

In the preliminary rule set, the three numbers that precede every rule are adjusted taking into account the preliminary modified data set. Thus during classification of unseen cases by the LERS classification system rules describe the original concept probabilistic approximation of the concept X.

V. EXPERIMENTS

In our experiments, we used six real-life data sets taken from the University of California at Irvine *Machine Learning Repository*, see Table IV. For every data set a set of templates was created. Templates were formed by replacing incrementally (with 5% increment) existing specified attribute values by *lost* values. Thus, we started each series of experiments with no *lost* values, then we added 5% of *lost* values, then we added additional 5% of *lost* values, etc., until at least one entire row of the data sets was full of *lost* values. Then, three attempts were made to change the configuration of new *lost* values and either a new data set with extra 5% of *lost* values was created or the process was terminated. Additionally, the same

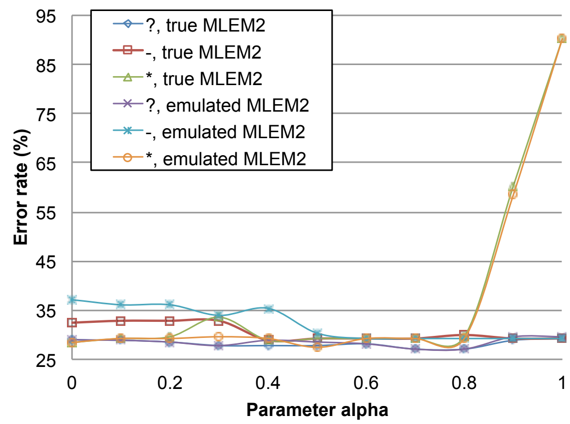

Figure 1. *Breast cancer* data set

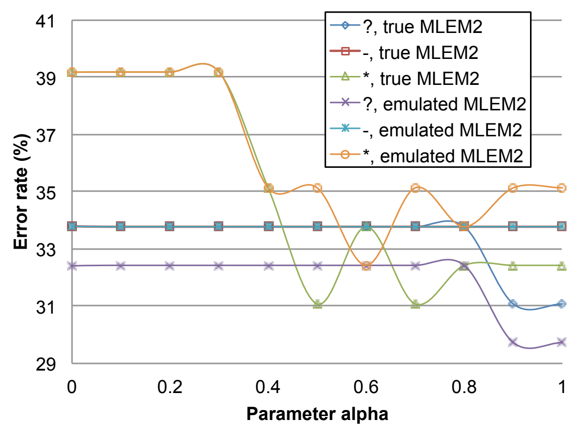

Figure 2. *Echocardiogram* data set

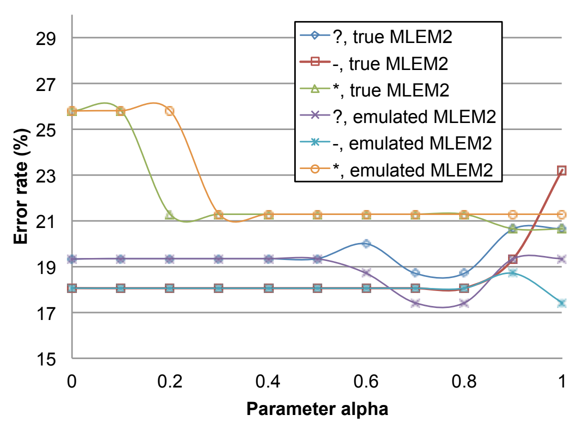

Figure 3. *Hepatitis* data set

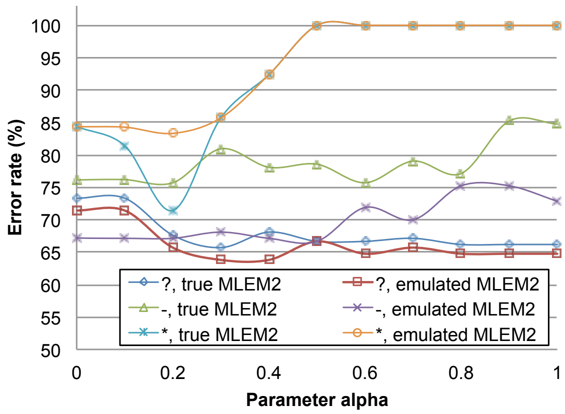

Figure 4. *Image segmentation* data set

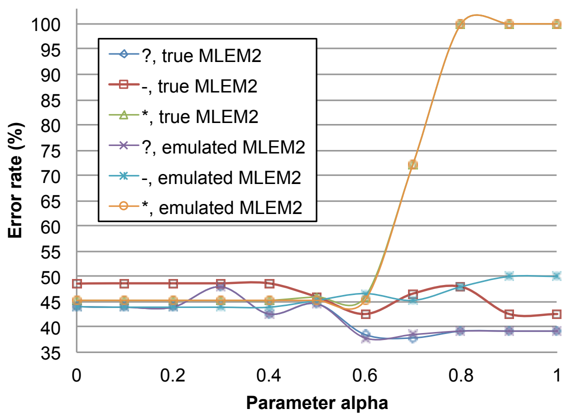

Figure 5. *Lymphography* data set

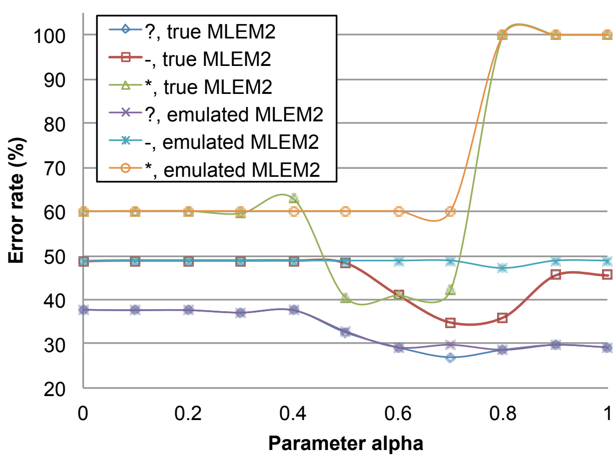

Figure 6. *Wine recognition* data set

templates were edited for further experiments by replacing question marks, representing *lost* values by "−"s, representing *attribute-concept* values, and then by "*"s, representing "do not care" conditions.

For any data set, there was some maximum for the percentage of missing attribute values. For example, for the *Breast cancer* data set, it was 44.81%. In our experiments we used only such incomplete data sets, with as many missing attribute values as possible. Note that for some data sets the maximum of the number of missing attribute values was less than 40%, we have not used such data for our experiments. Thus, for any data set from Table IV, three data sets were used for experiments, so the total number of data sets was 18.

Our first objective was to compare both approaches to rule induction, true MLEM2 and emulated MLEM2, in terms of the classification error rate of the induced rules. Results of our experiments are presented in Figures 1–6, with lost values denoted by "?", attribute-concept values denoted by "−", and "do not care" conditions denoted by "*".

For five combinations of data set and interpretation of missing attribute values the error rate was significantly smaller for the emulated version of MLEM2. The five combinations included *Breast cancer* and *Image segmentation* with "?" and "−", and *Echocardiogram* with "?". For *Wine recognition* with "−", the error rate was significantly smaller for the true version of MLEM2. In the remaining 12 combinations the difference in error rate was not significant (5% significance level, two-tailed test) and for the *Breast cancer* with "−" combination, the error rate for both versions of the MLEM2 algorithm was identical for all 11 values of α.

Our second objective was to check which interpretation of missing attribute value provides the smallest error rate, computed as a result of ten-fold cross validation. In eight combinations the error rate was significantly smaller for "?" than for "*". The eight combinations included both true and emulated MLEM2 with the *Image segmentation*, *Lymphography* and *Wine recognition* data sets, and emulated MLEM2 with the *Echocardiogram* and *Hepatitis* data sets.

For the following seven combinations the error rate was significantly smaller for "?" than for "−". The seven combinations included both true and emulated MLEM2 with the *Breast cancer* and *Wine recognition* data sets, true MLEM2 with *Image segmentation* and *Lymphography*, and emulated MLEM2 with the *Echocardiogram* data set.

In four combinations the error rate was measurably smaller for "−" than "?". The four combinations included both true and emulated MLEM2 with the *Hepatitis* data set and emulated MLEM2 with the *Image segmentation* and *Wine recognition* data sets.

For one combination the error rate was smaller for "*" than for "−", true MLEM2 and the *Breast cancer* data set. However, for the remaining combinations the difference in error rate was not significant. In addition, "−" and "*" values were never significantly better than "?".

VI. Conclusions

In our experiments we compared true and emulated versions of the MLEM2 algorithm using the error rate of the induced rule set, a result of ten-fold cross validation, as the quality criterion. Results of the same comparison of error rate of the induced rules are measurably more conclusive than previous experiments: in five combinations (out of 18) the emulated approach to MLEM2 algorithm was better, in one

 ISBN: 978-1-61208-486-2

TABLE IV. DATA SETS USED FOR EXPERIMENTS

Data set	Number of			% of
	cases	attributes	concepts	missing values
Breast cancer	277	9	2	44.81
Echocardiogram	74	7	2	40.15
Hepatitis	155	19	2	60.27
Image segmentation	210	19	7	69.85
Lymphography	148	18	4	69.89
Wine recognition	178	13	3	64.65

case the true approach was better (5% significance level, two-tailed test). In addition, there is some evidence that the lost value interpretation of missing attribute values is the best: in seven cases this interpretation was significantly better than attribute-concept values, and in eight cases it was better than "do not care" conditions. Additionally, attribute-concept values and "do not care" conditions were never significantly better than lost values.

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] 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.

[12] ——, "Rule induction using probabilistic approximations and data with missing attribute values," in Proceedings of the 15-th IASTED International Conference on Artificial Intelligence and Soft Computing ASC 2012, 2012, pp. 235–242.

[13] J. W. Grzymala-Busse, S. R. Marepally, and Y. Yao, "A comparison of positive, boundary, and possible rules using the MLEM2 rule induction algorithm," in Proceedings of the 10-th International Conference on Hybrid Intelligent Systems, 2010, pp. 7–12.

[14] J. W. Grzymala-Busse, "Generalized probabilistic approximations," Transactions on Rough Sets, vol. 16, 2013, pp. 1–16.

[15] J. W. Grzymala-Busse, S. R. Marepally, and Y. Yao, "An empirical comparison of rule sets induced by LERS and probabilistic rough classification," in Proceedings of the 7-th International Conference on Rough Sets and Current Trends in Computing, 2010, pp. 590–599.

[16] P. G. Clark and J. W. Grzymala-Busse, "A comparison of two versions of the MLEM2 rule induction algorithm extended to probabilistic approximations," in Proceedings of the 9-th International Conference on Rough Sets and Current Trends in Computing, 2014, pp. 109–119.

[17] J. W. Grzymala-Busse, "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.

[18] ——, "Three approaches to missing attribute values—a rough set perspective," in Proceedings of the Workshop on Foundation of Data Mining, in conjunction with the Fourth IEEE International Conference on Data Mining, 2004, pp. 55–62.

[19] ——, "Data with missing attribute values: Generalization of indiscernibility relation and rule induction," Transactions on Rough Sets, vol. 1, 2004, pp. 78–95.

[20] T. Y. Lin, "Topological and fuzzy rough sets," in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory, R. Slowinski, Ed. Dordrecht, Boston, London: Kluwer Academic Publishers, 1992, pp. 287–304.

[21] 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.

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

[23] ——, "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.