Local Probabilistic Approximations for Incomplete Data

Patrick G. Clark[1], Jerzy W. Grzymala-Busse[1,2], and Martin Kuehnhausen[1]

[1] Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA
{pclark,jerzy,mkuehnha}@ku.edu
[2] Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, 35-225 Rzeszow, Poland

Abstract. In this paper we introduce a generalization of the local approximation called a local probabilistic approximation. Our novel idea is associated with a parameter (probability) α. If $\alpha = 1$, the local probabilistic approximation becomes a local lower approximation; for small α, it becomes a local upper approximation. The main objective of this paper is to test whether proper local probabilistic approximations (different from local lower and upper approximations) are better than ordinary local lower and upper approximations. Our experimental results, based on ten-fold cross validation, show that all depends on a data set: for some data sets proper local probabilistic approximations are better than local lower and upper approximations; for some data sets there is no difference, for yet other data sets proper local probabilistic approximations are worse than local lower and upper approximations.

1 Introduction

Lower and upper approximations are fundamental concepts of rough set theory [1, 2]. An idea of the local and global approximations was introduced in [3]. Later on, local and global approximations were discussed in [4]. Local approximations are unions of complex blocks which, in turn, are intersections of attribute-value pair blocks. Global approximations are unions of characteristic sets, where the characteristic set is a generalization of the elementary set, well known in rough set theory.

In this paper we introduce a novel idea of the local probabilistic approximation, a generalization of the local lower and upper approximations. The local probabilistic approximation is defined using a parameter α that has an interpretation as a conditional probability of the concept given complex block. If $\alpha = 1$, the local probabilistic approximation is the local lower approximation; if α is small (in this paper 0.001), the local probabilistic approximation is the local upper approximation.

Theoretical properties of global probabilistic approximations, based on an equivalence relation, were studied for many years in variable precision rough set theory, Bayesian rough sets, etc. [5–10]. Global probabilistic approximations

L. Chen et al. (Eds.): ISMIS 2012, LNAI 7661, pp. 93–98, 2012.

based on an arbitrary binary relation were defined in [11]. First results of their practical usefulness were published in [12].

In this paper we will distinguish two interpretations of a missing attribute value: lost values and "do not care" conditions. If an attribute value was originally given but now is not accessible (e.g., was erased or forgotten) we will call it *lost*. If a data set consists lost values, we will try to induce rules from existing data. Another interpretation of a missing attribute value is based on a refusal to answer a question, e.g., some people may refuse to tell their citizenship status, such a value will be called a "do not care" condition. In data sets with "do not care" conditions we will replace such a missing attribute value with all possible attribute values.

2 Attribute-Value Pair Blocks

We assume that the input data sets are presented in the form of a *decision table*. Rows of the decision table represent *cases* and columns are labeled by *variables*. The set of all cases will be denoted by U. Some variables are called *attributes* while one selected variable is called a *decision* and is denoted by d. The set of all attributes will be denoted by A.

An important tool to analyze data sets is a *block of an attribute-value pair*. Let (a, v) be an attribute-value pair. For *complete* decision tables, i.e., decision tables in which every attribute value is specified, a block of (a, v), denoted by $[(a, v)]$, is the set of all cases x for which $a(x) = v$, where $a(x)$ denotes the value of the attribute a for the case x. For incomplete decision tables the definition of a block of an attribute-value pair is modified.

- 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 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.

A special block of a decision-value pair is called a *concept*. 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 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 $K(x, a) = U$.

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.

3 Local Probabilistic Approximations

For incomplete data sets, a set X will be called *B-globally definable* if it is a union of some characteristic sets $K_B(x)$, $x \in U$. A set T of attribute-value pairs, where all attributes belong to set B and are distinct, will be called a *B-complex*. We will discuss only *nontrivial complexes*, i.e., such complexes that the intersection of all attribute-value blocks from a given complex is not the empty set. A block of B-complex T, denoted by $[T]$, is defined as the set $\cap\{[t] \mid t \in T\}$.

For an incomplete decision table and a subset B of A, a union of intersections of attribute-value pair blocks of attribute-value pairs from some B-complexes, will be called a *B-locally definable* set. Any set X that is B-globally definable is B-locally definable, the converse is not true.

Let X be any subset of the set U of all cases. Let $B \subseteq A$. In general, X is not a B-definable set, locally or globally. A *B-local probabilistic approximation* of the set X with the parameter α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{local}B(X)$, is defined as follows

$$\cup\{[T] \mid \exists \textit{ a family } \mathcal{T} \textit{ of } B\textit{-complexes } T \textit{ of } X, T \in \mathcal{T}, \ Pr(X|[T]) \geq \alpha\}.$$

Due to computational complexity, in our experiments we used a heuristic approach to computing another local probabilistic approximation, denoted by $appr_{\alpha}^{mlem2}(X)$, since it is inspired by the MLEM2 rule induction algorithm [4]. Using this approach, $appr_{\alpha}^{mlem2}(X)$ is constructed from A-complexes Y that are the most relevant to X, i.e., with $|X \cap Y|$ as large as possible, if there is more than one A-complex that satisfies this criterion, the largest conditional probability of X given Y is the next criterion to select an A-complex. Note that if two A-complexes are equally relevant, then the second criterion selects an A-complex with the smaller block cardinality.

4 Experiments

In our experiments we used eight real-life data sets taken from the University of California at Irvine *Machine learning Repository*. These data sets were enhanced by replacing 35% of existing attribute values by missing attribute values, separately by *lost* values and by *"do not care" conditions*. Thus, for any data set, two data sets were created for experiments, one with missing attribute values interpreted as lost values and the other one as "do not care" conditions.

The main objective of our research was to test whether local probabilistic approximations, different from local lower and upper approximations, are better than local lower and upper approximations in terms of an error rate. Therefore, we conducted experiments of a ten-fold cross validation increasing the parameter α in local probabilistic approximations inspired by MLEM2, with increments equal to 0.1, from 0 to 1.0. For a given data set, in our experiments we used ten-fold cross validation with a random re-ordering of all cases, but during all eleven experiments this order was constant, i.e., all ten pairs of training and testing data subsets were the same. Results of our experiments are presented in Figures 1–4.

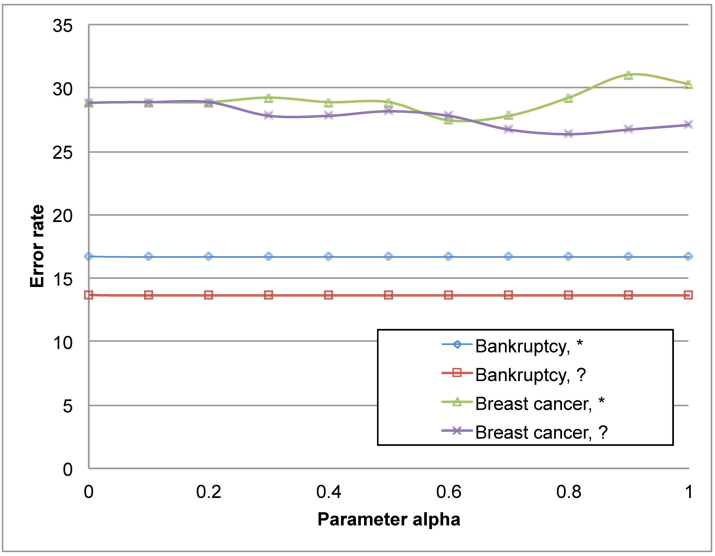

Fig. 1. Error rates for data sets *Bankruptcy* and *Breast cancer* with lost values, denoted by ? and "do not care" conditions denoted by *

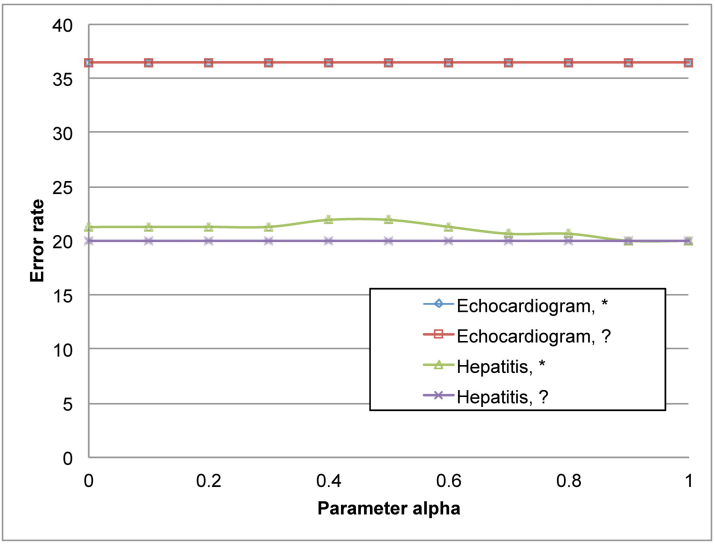

Fig. 2. Error rates for data sets *Echocardiogram* and *Hepatitis* with lost values, denoted by ? and "do not care" conditions denoted by *

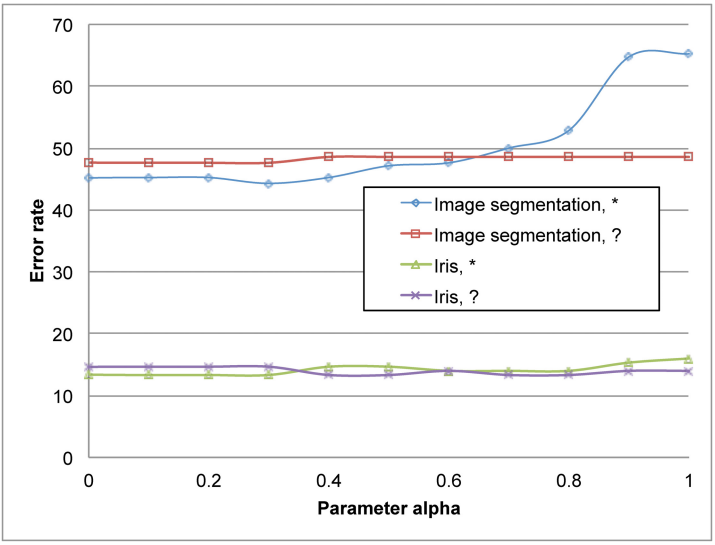

Fig. 3. Error rates for data sets *Image segmentation* and *Iris* with lost values, denoted by ? and "do not care" conditions denoted by *

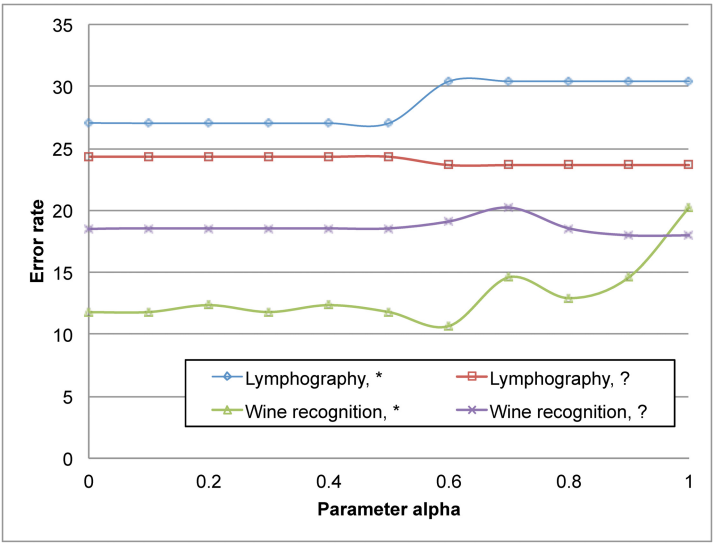

Fig. 4. Error rates for data sets *Lymphography* and *Wine recognition* with lost values, denoted by ? and "do not care" conditions denoted by *

5 Conclusions

As follows from our experiments, for three out of 16 possibilities (the *Breast cancer* data set with missing attribute values interpreted as lost values with α =0.8, the *Breast cancer* data set with missing attribute values interpreted as "do not care" conditions with $\alpha = 0.6$, and *Wine recognition* data set with missing attribute values interpreted as "do not care" conditions with $\alpha = 0.6$) local probabilistic approximations are better than ordinary local lower and upper approximations. On the other hand, for other three possibilities (*Breast cancer* data set with missing attribute values interpreted as "do not care" conditions with $\alpha = 0.9$, the *Hepatitis* data set with missing attribute values interpreted as "do not care" conditions with $\alpha = 0.5$ and for *Wine recognition* data set with missing attribute values interpreted as lost values conditions with $\alpha = 0.7$) local probabilistic approximations are worse than ordinary local lower and upper approximations. Therefore, it is obvious that usefulness of proper local probabilistic approximations depends on a data sets and an interpretation of missing attribute values.

References

1. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
2. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
3. Grzymala-Busse, J.W., Rzasa, W.: Local and Global Approximations for Incomplete Data. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H.S., Słowiński, R. (eds.) RSCTC 2006. LNCS (LNAI), vol. 4259, pp. 244–253. Springer, Heidelberg (2006)
4. Grzymala-Busse, J.W., Rzasa, W.: A local version of the MLEM2 algorithm for rule induction. Fundamenta Informaticae 100, 99–116 (2010)
5. Grzymala-Busse, J.W., Ziarko, W.: Data mining based on rough sets. In: Wang, J. (ed.) Data Mining: Opportunities and Challenges, pp. 142–173. Idea Group Publ., Hershey (2003)
6. Pawlak, Z., Skowron, A.: Rough sets: Some extensions. Information Sciences 177, 28–40 (2007)
7. Ślęzak, D., Ziarko, W.: The investigation of the bayesian rough set model. International Journal of Approximate Reasoning 40, 81–91 (2005)
8. Yao, Y.Y.: Probabilistic rough set approximations. International Journal of Approximate Reasoning 49, 255–271 (2008)
9. Ziarko, W.: Variable precision rough set model. Journal of Computer and System Sciences 46(1), 39–59 (1993)
10. Ziarko, W.: Probabilistic approach to rough sets. International Journal of Approximate Reasoning 49, 272–284 (2008)
11. Grzymała-Busse, J.W.: Generalized Parameterized Approximations. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 136–145. Springer, Heidelberg (2011)
12. Clark, P.G., Grzymala-Busse, J.W.: Experiments on probabilistic approximations. In: Proceedings of the 2011 IEEE International Conference on Granular Computing, pp. 144–149 (2011)