Evaluating a Rule Evaluation Support Method Based on Objective Rule Evaluation Indices

Hidenao Abe[1], Shusaku Tsumoto[1], Miho Ohsaki[2], and Takahira Yamaguchi[3]

[1] Department of Medical Informatics, Shimane University, School of Medicine
89-1 Enya-cho, Izumo, Shimane 693-8501, Japan
abe@med.shimane-u.ac.jp, tsumoto@computer.org
[2] Faculty of Engineering, Doshisha University
1-3 Tataramiyakodani, Kyo-Tanabe, Kyoto 610-0321, Japan
mohsaki@mail.doshisha.ac.jp
[3] Faculty of Science and Technology, Keio University
3-14-1 Hiyoshi, Kohoku Yokohama, Kanagawa 223-8522, Japan
yamaguti@ae.keio.ac.jp

Abstract. In this paper, we present an evaluation of novel rule evaluation support method for post-processing of mined results with rule evaluation models based on objective indices. Post-processing of mined results is one of the key issues in a data mining process. However, it is difficult for human experts to evaluate many thousands of rules from a large dataset with noises completely. To reduce the costs of rule evaluation task, we have developed the rule evaluation support method with rule evaluation models, which are obtained with objective indices of mined classification rules and evaluations of a human expert for each rule. To evaluate performances of learning algorithms for constructing rule evaluation models, we have done a case study on the meningitis data mining as an actual problem. Furthermore, we have also evaluated our method on four rulesets from the four kinds of UCI datasets.

1 Introduction

In recent years, huge data are easily stored on information systems in natural science, social science and business domains, developing information technologies. With these huge data, people hope to utilize them for their purposes. Besides, data mining techniques have been widely known as a process for utilizing stored data on database systems, combining different kinds of technologies such as database technologies, statistical methods and machine learning methods. Especially, IF-THEN rules, which are produced by rule induction algorithms, are discussed as one of highly usable and readable output of data mining. However, to large dataset with hundreds attributes including noises, the process often obtains many thousands of rules. From such huge rule set, it is difficult for human experts to find out valuable knowledge which are rarely included in the rule set.

To support such a rule selection, many efforts have done using objective rule evaluation indices such as recall, precision, and other interestingness measurements (we call them 'objective indices' later). However, it is also difficult to

estimate a criterion of a human expert with single objective rule evaluation index, because his/her subjective criterion such as interestingness and importance for his/her purpose is influenced by the amount of his/her knowledge and/or a passage of time.

To above issues, we have been developed an adaptive rule evaluation support method for human experts with rule evaluation models, which predict experts' criteria based on objective indices, re-using results of evaluations of human experts. In Section 3, we describe the rule evaluation model construction method based on objective indices. Then we present a performance comparison of learning algorithms for constructing rule evaluation models in Section 4. With the results of the comparison, we discuss about the availability of our rule evaluation model construction approach.

2 Related Work

Many efforts have done to select valuable rules from mined large rule set based on objective rule evaluation indexes. Some of these works suggest the indexes to discover interesting rules from such a large amount of rules.

Focusing on interesting rule selection with objective indexes, researchers have developed more than forty objective indexes based on number of instances, probability, statistics values, information quantity, distance of rules or their attributes, and complexity of a rule[11, 21, 23]. Most of these indexes are used to remove meaningless rules rather than to discover really interesting ones for a human expert, because they can not include domain knowledge. In contrast, a dozen of subjective indexes estimate how a rule fits with a belief, a bias or a rule template formulated beforehand by a human expert. Although these subjective indexes are useful to discover really interesting rules to some extent due to their built-in domain knowledge, they depend on the precondition that a human expert is able to clearly formulate his/her interest. Although interestingness indexes were verified their availabilities on each suggested domain, nobody has validated their availabilities on the other domains or/and characteristics related to the background of a given dataset.

Ohsaki et. al[15] investigated the relation between objective indexes and real human interests, taking real data mining results and their human evaluations. In this work, the comparison shows that it is difficult to predict real human interests with a single objective index. Based on the result, they indicated the possibility of logical combination of the objective indexes to predict real human interests more exactly.

3 Rule Evaluation Support with Rule Evaluation Model based on Objective Indices

We considered the process of modeling rule evaluations of human experts as the process to clear up relationships between the human evaluations and features
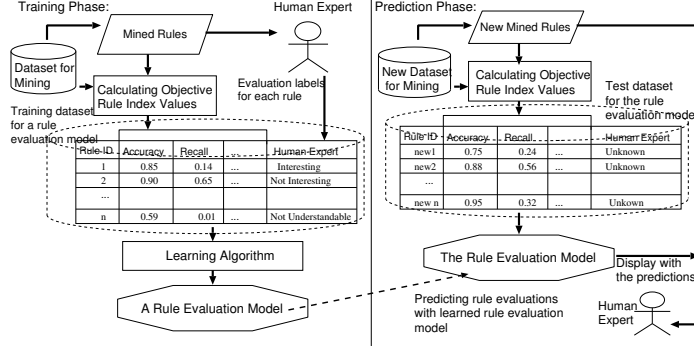
Training Phase: Mined Rules   Human Expert

Dataset for Mining

Training dataset for a rule evaluation model

Calculating Objective Rule Index Values   Evaluation labels for each rule

Rule ID	Accuracy	Recall	...	Human Expert
1	0.85	0.14	...	Interesting
2	0.90	0.65	...	Not Interesting
...				
n	0.59	0.01	...	Not Understandable

Learning Algorithm

A Rule Evaluation Model

Prediction Phase: New Mined Rules

New Dataset for Mining

Calculating Objective Rule Index Values   Test dataset for the rule evaluation model

Rule ID	Accuracy	Recall	...	Human Expert
new1	0.75	0.24	...	Unknown
new2	0.88	0.56	...	Unknown
...				
new n	0.95	0.32	...	Unknown

The Rule Evaluation Model   Display with the predictions

Predicting rule evaluations with learned rule evaluation model   Human Expert

Fig. 1. Overview of the construction method of rule evaluation models.

of input if-then rules. With this consideration, we decided that the process of rule evaluation model construction can be implemented as a learning task. Fig.1 shows the process of rule evaluation model construction based on re-use of human evaluations and objective indices for each mined rule.

At the training phase, attributes of a meta-level training data set is obtained by objective indices such as recall, precision and other rule evaluation values. The human evaluations for each rule are joined as class of each instance. To obtain this data set, a human expert has to evaluate the whole or part of input rules at least once. After obtaining the training data set, its rule evaluation model is constructed by a learning algorithm. At the prediction phase, a human expert receives predictions for new rules based on their values of the objective indices. Since the task of rule evaluation models is a prediction, we need to choose a learning algorithm with higher accuracy as same as current classification problems.

4 Performance Comparisons of Learning Algorithms for Rule Model Construction

To predict human evaluation labels of a new rule based on objective indices more exactly, we have to construct a rule evaluation model, which has higher predictive accuracy.

In this section, we firstly present the result of an empirical evaluation with the dataset from the result of a meningitis data mining[9]. Then to confirm the performance of our approach on the other datasets, we evaluated the five algorithms on four rule sets from four kinds of UCI benchmark datasets [10]. With the experimental results, we discuss about the following three view points: accuracies of rule evaluation models, analyzing learning curves of the learning algorithms, and contents of learned rule evaluation models.

As an evaluation of accuracies of rule evaluation models, we have compared predictive accuracies on the whole dataset and Leave-One-Out. The accuracy of a validation dataset D is calculated with correctly predicted instances $Correct(D)$

Table 1. The objective rule evaluation indices for classification rules used in this research. **P:** Probability of the antecedent and/or consequent of a rule. **S:** Statistical variable based on P. **I:** Information of the antecedent and/or consequent of a rule. **N:** Number of instances included in the antecedent and/or consequent of a rule. **D:** Distance of a rule from the others based on rule attributes.

Theory	Index Name (**Abbreviation**) [Reference Number of Literature]
P	Coverrage(**Coverage**), Prevalence(**Prevalence**)
	Precision(**Precision**), Recall(**Recall**)
	Suppurt(**Support**), Specificity(**Specificity**)
	Accuracy(**Accuracy**), Lift(**Lift**)
	Leverage(**Leverage**), Added Value(**Added Value**)[21]
	Klösgen's Interestingness(**KI**)[14], Relative Risk(**RR**)[1]
	Brin's Interest(**BI**)[2], Brin's Conviction(**BC**)[2]
	Certainty Factor(**CF**)[21], Jaccard Coefficient(**Jaccard**)[21]
	F-Measure(**F-M**)[19], Odds Ratio(**OR**)[21]
	Yule's Q(**YuleQ**)[21], Yule's Y(**YuleY**)[21]
	Kappa(**Kappa**)[21], Collective Strength(**CST**)[21]
	Gray andOrlowska's Interestingness weighting Dependency(**GOI**)[7]
	Gini Gain(**Gini**)[21], Credibility(**Credibility**)[8]
S	χ^2 Measure for One Quadrant(χ^2-**M1**)[6]
	χ^2 Measure for Four Quadrant(χ^2-**M4**)[6]
I	J-Measure(**J-M**)[20], K-Measure(**K-M**)[15]
	Mutual Information(**MI**)[21]
	Yao and Liu's Interestingness 1 based on one-way support(**YLI1**)[23]
	Yao and Liu's Interestingness 2 based on two-way support(**YLI2**)[23]
	Yao and Zhong's Interestingness(**YZI**)[23]
N	Cosine Similarity(**CSI**)[21], Laplace Correction(**LC**)[21]
	ϕ Coefficient(ϕ)[21], Piatetsky-Shapiro's Interestingness(**PSI**)[16]
D	Gago and Bento's Interestingness(**GBI**)[5]
	Peculiarity(**Peculiarity**)[24]

as $Acc(D) = (Correct(D)/|D|) \times 100$, where $|D|$ means the size of the dataset. Recalls of class i on a validation dataset is calculated with correctly predicted instances about the class $Correct(D_i)$ as $Recall(D_i) = (Correct(D_i)/|D_i|) \times 100$, where $|D_i|$ means the size of instances with class i. Also the precision of class i is calculated with the size of instances predicted i as $Precision(D_i) = (Correct(D_i)/Predicted(D_i)) \times 100$.

As for learning curves, we obtained learning curves about accuracies to the whole training dataset to evaluate whether each learning algorithm can perform in early stage of a process of rule evaluations. Accuracies from randomly sub-sampled training datasets are averaged with 10 times trials on each percentage of subset.

Looking at elements of the rule evaluation models on the meningitis data mining result, we consider the characteristics of objective indices, which are used in these rule evaluation models.

To construct a dataset to learn a rule evaluation model, values of objective indices have been calculated for each rule, taking 39 objective indices as shown in Table1. Thus each dataset for each rule set has the same number of instances as the rule set. Each instance consists of 40 attributes including the class attribute.

To these dataset, we applied five learning algorithms to compare their performance as a rule evaluation model construction method. We used the following learning algorithms from Weka[22]: C4.5 decision tree learner[18] called J4.8,

neural network learner with back propagation (BPNN)[12], support vector machines (SVM)[4][17], classification via linear regressions (CLR)[5][3], and OneR[13].

4.1 Constructing Rule Evaluation Models on an Actual Datamining Result

In this case study, we have taken 244 rules, which are mined from six dataset about six kinds of diagnostic problems as shown in Table2. These datasets are consisted of appearances of meningitis patients as attributes and diagnoses for each patient as class. Each rule set was mined with each proper rule induction algorithm composed by a constructive meta-learning system called CAMLET[9]. For each rule, we labeled three evaluations (I:Interesting, NI:Not-Interesting, NU:Not-Understandable), according to evaluation comments from a medical expert.

Table 2. Description of the meningitis datasets and their datamining results

Dataset	#Attributes	#Class	#Mined rules	#'I' rules	#'NI' rules	#'NU' rules
Diag	29	6	53	15	38	0
C_Cource	40	12	22	3	18	1
Culture+diag	31	12	57	7	48	2
Diag2	29	2	35	8	27	0
Course	40	2	53	12	38	3
Cult_find	29	2	24	3	18	3
TOTAL	—	—	244	48	187	9

Comparison on Classification Performances In this section, we show the result of the comparisons of accuracies on the whole dataset, recall of each class label, and precisions of each class label. Since Leave-One-Out holds just one test instance and remains as the training dataset repeatedly for each instance of a given dataset, we can evaluate the performance of a learning algorithm to a new dataset without any ambiguity.

The results of the performances of the five learning algorithms to the whole training dataset and the results of Leave-One-Out are also shown in Table3. All of the accuracies, Recalls of I and NI, and Precisions of I and NI are higher than predicting default labels.

Table 3. Accuracies(%), Recalls(%) and Precisions(%) of the five learning algorithms.

| | On the whole training dataset | | | | | | Leave-One-Out | | | | | | |
| | | Recall of | | | Precision of | | | | Recall of | | | Precision of | |
	Acc.	I	NI	NU	I	NI	NU	Acc.	I	NI	NU	I	NI	NU
J4.8	85.7	41.7	97.9	66.7	80.0	86.3	85.7	79.1	29.2	95.7	0.0	63.6	82.5	0.0
BPNN	86.9	81.3	89.8	55.6	65.0	94.9	71.4	77.5	39.6	90.9	0.0	50.0	85.9	0.0
SVM	81.6	35.4	97.3	0.0	68.0	83.5	0.0	81.6	35.4	97.3	0.0	68.0	83.5	0.0
CLR	82.8	41.7	97.3	0.0	71.4	84.3	0.0	80.3	35.4	95.7	0.0	60.7	82.9	0.0
OneR	82.0	56.3	92.5	0.0	57.4	87.8	0.0	75.8	27.1	92.0	0.0	37.1	82.3	0.0

Comparing with the accuracy of OneR, the other learning algorithms achieve equal or higher performance with combination of multiple objective indices than

---

[4] The kernel function was set up polynomial kernel.

[5] We set up the elimination of collinear attributes and the model selection with greedy search based on Akaike Information Metric.

sorting with single objective index. Looking at Recall values on class I, BPNN have achieved the highest performance. As for the other algorithms, they show lower performance than OneR, because they have tended to be learned classification patterns for the major class NI.

The accuracies of Leave-One-Out shows robustness of each learning algorithm. These learning algorithms have achieved from 75.8% to 81.9%. However, these learning algorithms have not been able to classify the instances with class NU, because it is difficult to predict a minor class label in this dataset.

Learning Curves of the Learning Algorithms Since the rule evaluation model construction method needs evaluations of mined rules by a human expert, we have investigated learning curves of each learning algorithm to estimate minimum training subset to construct a valid rule evaluation model. The upper table in Fig.2 shows accuracies to the whole training dataset with each subset of training dataset. The percentages of achievements for each learning algorithm, comparing with the accuracy with the whole dataset, are shown in the lower chart of Fig.2.

As shown in these results, SVM and CLR, which learn hype-planes, achieves grater than 95% with only less than 10% of training subset. Although decision tree learner and BPNN could learn better classifier to the whole dataset than these hyper-plane learners, they need more training instances to learn accurate classifiers.
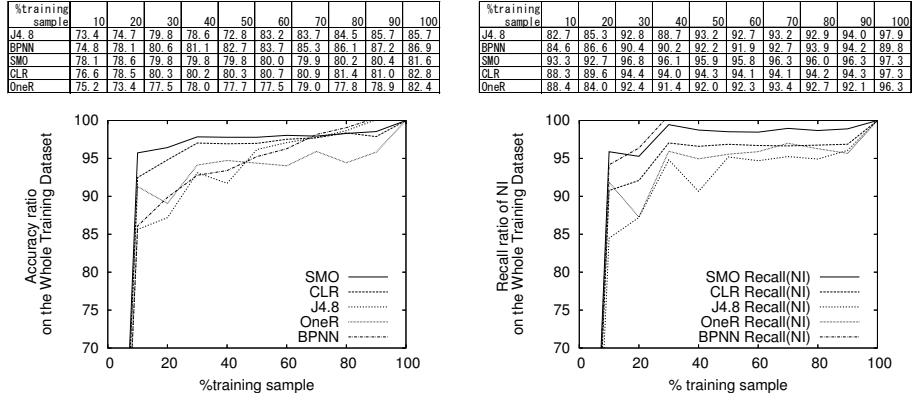
%training sample	10	20	30	40	50	60	70	80	90	100
J4.8	73.4	74.7	79.8	78.6	72.8	83.2	83.7	84.5	85.7	85.7
BPNN	74.8	78.1	80.6	81.1	82.7	83.7	85.3	86.1	87.2	86.9
SMO	78.1	78.6	79.8	79.8	79.8	80.0	79.9	80.2	80.4	81.6
CLR	76.6	78.5	80.3	80.2	80.3	80.7	80.9	81.4	81.0	82.8
OneR	75.2	73.4	77.5	78.0	77.7	77.5	79.0	77.8	78.9	82.4

%training sample	10	20	30	40	50	60	70	80	90	100
J4.8	82.7	85.3	92.8	88.7	93.2	92.7	93.2	92.9	94.0	97.9
BPNN	84.6	86.6	90.4	90.2	92.2	91.9	92.7	93.9	94.2	89.8
SMO	93.3	92.7	96.8	96.1	95.9	95.8	96.3	96.0	96.3	97.3
CLR	88.3	89.6	94.4	94.0	94.3	94.1	94.1	94.2	94.3	97.3
OneR	88.4	84.0	92.4	91.4	92.0	92.0	93.4	92.7	92.1	96.3

Fig. 2. Learning Curves of accuracies(%) on the learning algorithms with sub-sampled training dataset: The left table shows accuracies(%) on each training dataset to the whole dataset. The left graph shows their achievement ratio(%). Also the right table shows recalls(%), and the graph shows their achievement ratio(%).

To eliminate known ordinary knowledge from large rule set, it is needed to classify non-interesting rules correctly. The right upper table in Fig.2 shows percentages of recalls on NI. The right lower chart in Fig.2 also shows the percentages of achievements on recall of NI, comparing with the recall of NI on the whole training dataset. Looking at this result, we can eliminate NI rules with rule evaluation models from SVM and BPNN even if there is only 10% of
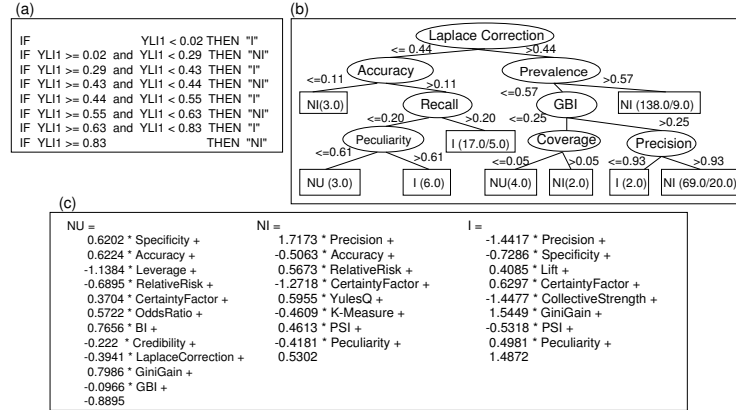
Fig. 3. Learned models to the meningitis data mining result dataset.

rule evaluations by a human expert. This is guaranteed with no less than 80% precisions of all learning algorithms.

Rule Evaluation Models on the Actual Datamining Result Dataset In this section, we present rule evaluation models to the whole dataset learned with OneR, J4.8 and CLR, because they are represented as explicit models such as a rule set, a decision tree, and a set of linear models.

Fig.3 shows rule evaluation models on the actual data mining result: The rule set of OneR is shown in Fig.3(a), Fig.3(b) shows the decision tree learned with J4.8, and Fig.3(c) shows linear models to classify each class.

Looking at indices used in learned rule evaluation models, they are not only the group of indices increasing with a correctness of a rule, but also they are used some different groups of indices on different models. Almost indices such as YLI1, Laplace Correction, Accuracy, Precision, Recall, Coverage, PSI and Gini Gain are the former type of indices on the models. The later indices are GBI and Peculiality, which sums up difference of antecedents between one rule and the other rules in the same ruleset. This corresponds to the comment from the human expert. He said that he evaluated these rules not only correctness but also his interest based on his expertise

4.2 Constructing Rule Evaluation Models on Artificial Evaluation Labels

We have also evaluated our rule evaluation model construction method with rule sets from four datasets of UCI Machine Learning Repository to confirm the lower limit performances on probabilistic class distributions.

We selected the following four datasets: Mushroom, Heart, Internet Advertisement Identification (called InternetAd later) and Letter. With these datasets, we obtained rule sets with bagged PART, which repeatedly executes PART[4] to bootstrapped training sub-sample datasets.

To these rule sets, we calculated the 39 objective indices as attributes of each rule. As for the class of these datasets, we set up three class distributions with multinomial distribution. Table4 shows us the datasets with three different class distributions. The class distribution for 'Distribution I' is $P = (0.35, 0.3, 0.3)$ where p_i is the probability for class i. Thus the number of class i in each instance D_j become $p_i D_j$. As the same way, the probability vector of 'Distribution II' is $P = (0.3, 0.5, 0.2)$, and 'Distribution III' is $P = (0.3, 0.65, 0.05)$.

Table 4. Datasets of the rule sets learned from the UCI benchmark datasets

	#Mined Rules	#Class labels			%Def. class
		L1	L2	L3	
Distribution I		(0.30)	(0.35)	(0.35)	
Mushroom	30	8	14	8	46.7
InternetAd	107	26	39	42	39.3
Heart	318	97	128	93	40.3
Letter	6340	1908	2163	2269	35.8
Distribution II		(0.30)	(0.50)	(0.20)	
Mushroom	30	11	16	3	53.3
InternetAd	107	30	53	24	49.5
Heart	318	99	140	79	44.0
Letter	6340	1890	3198	1252	50.4
Distribution III		(0.30)	(0.65)	(0.05)	
Mushroom	30	7	21	2	70.0
InternetAd	107	24	79	9	73.8
Heart	318	98	205	15	64.5
Letter	6340	1947	4062	331	64.1

Accuracy Comparison on Classification Performances To above datasets, we have attempted the five learning algorithms to estimate whether their classification results can go to or beyond the accuracies of just predicting each default class. The left table of Table5 shows the accuracies of the five learning algorithms to each class distribution of the three datasets. As shown in Table5, J48 and BPNN always work better than just predicting a default class. However, their performances are suffered from probabilistic class distributions to larger datasets such as Heart and Letter.

Table 5. Accuracies(%) on whole training datasets labeled with three different distributions(The left table). Number of minimum training sub-samples to outperform %Def. class(The right table).

	J48	BPNN	SVM	CLR	OneR
Distribution I					
Mushroom	80.0	93.3	56.7	66.7	53.3
InternetAd	84.1	82.2	29.9	53.3	60.7
Heart	78.0	75.8	40.3	42.5	54.7
Letter	36.8	36.4	30.1	36.6	52.1
Distribution II					
Mushroom	93.3	93.3	80.0	80.0	76.7
InternetAd	73.8	79.4	49.5	59.8	60.7
Heart	72.3	69.2	35.9	47.8	55.7
Letter	51.0	51.0	50.4	50.4	57.0
Distribution III					
Mushroom	93.3	96.7	70.0	70.0	76.7
InternetAd	86.0	90.7	70.1	69.2	72.0
Heart	78.0	77.7	64.5	65.7	71.4
Letter	64.1	64.3	64.1	64.1	68.3

	J48	BPNN	SVM	CLR	OneR
Distribution I					
Mushroom	8	8	12	18	14
InternetAd	14	14	–	30	14
Heart	42	31	66	114	98
Letter	189	217	–	955	305
Distribution II					
Mushroom	6	4	4	6	12
InternetAd	24	24	52	42	70
Heart	52	40	–	104	92
Letter	897	>1000	451	–	>1000
Distribution III					
Mushroom	22	14	22	28	22
InternetAd	80	66	–	–	–
Heart	114	94	142	318	182
Letter	>1000	>1000	998	>1000	>1000

Evaluation on Learning Curves As same as evaluations of learning curves on the meningitis rule set, we have estimated the minimum training subsets for a valid model, which works better than just predicting a default class.

The right table in Table5 shows sizes of minimum training subsets, which can be constructed more accurate rule evaluation models than percentages of a

default class by each learning algorithm. To smaller dataset, such as Mushroom and InternetAd, they can construct valid models with less than 20% of given training datasets. However, to larger dataset, they need more training subsets to construct valid models, because their performances with whole training dataset fall to the percentages of default class of each dataset as shown in the left table in Table5.

5 Conclusion

In this paper, we have described rule evaluation support method with rule evaluation models to predict evaluations for an IF-THEN rule based on objective indices, re-using evaluations of a human expert.

As the result of the performance comparison with the five learning algorithms, rule evaluation models have achieved higher accuracies than just predicting each default class. Considering the difference between the actual evaluation labeling and the artificial evaluation labeling, it is shown that the medical expert evaluated with noticing particular relations between an antecedent and a class/another antecedent in each rule. In the estimation of robustness to a new rule with Leave-One-Out, we have achieved more than 75.8% with these learning algorithms. On the evaluation with learning curves to the dataset of the actual datamining result, SVM and CLR have achieved more than 95% of achievement ratio compared to the accuracy of the whole training dataset with less than 10% of subset of the training dataset with certain human evaluations. These results indicate the availability of this rule evaluation support method for a human expert.

As future work, we will introduce a selection method of learning algorithms to construct a proper rule evaluation model according to each situation. We also apply this rule evaluation support method to estimate other data mining result such as decision tree, rule set, and committee of them with objective indices, which evaluate whole mining results.

References

1. Ali, K., Manganaris,S., Srikant, R.: Partial Classification Using Association Rules. Proc. of Int. Conf. on Knowledge Discovery and Data Mining KDD-1997 (1997) 115–118
2. Brin, S., Motwani, R., Ullman, J., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. Proc. of ACM SIGMOD Int. Conf. on Management of Data (1997) 255–264
3. Frank, E., Wang, Y., Inglis, S., Holmes, G., and Witten, I. H.: Using model trees for classification, Machine Learning, Vol.32, No.1 (1998) 63–76
4. Frank, E, Witten, I. H., Generating accurate rule sets without global optimization, in Proc. of the Fifteenth International Conference on Machine Learning, (1998) 144–151
5. Gago, P., Bento, C.: A Metric for Selection of the Most Promising Rules. Proc. of Euro. Conf. on the Principles of Data Mining and Knowledge Discovery PKDD-1998 (1998) 19–27

6. Goodman, L. A., Kruskal, W. H.: Measures of association for cross classifications. Springer Series in Statistics, 1, Springer-Verlag (1979)
7. Gray, B., Orlowska, M. E.: CCAIIA: Clustering Categorical Attributes into Interesting Association Rules. Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining PAKDD-1998 (1998) 132–143
8. Hamilton, H. J., Shan, N., Ziarko, W.: Machine Learning of Credible Classifications. Proc. of Australian Conf. on Artificial Intelligence AI-1997 (1997) 330–339
9. Hatazawa, H., Negishi, N., Suyama, A., Tsumoto, S., and Yamaguchi, T.: Knowledge Discovery Support from a Meningoencephalitis Database Using an Automatic Composition Tool for Inductive Applications, in Proc. of KDD Challenge 2000 in conjunction with PAKDD2000 (2000) 28–33
10. Hettich, S., Blake, C. L., and Merz, C. J.: UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html], Irvine, CA: University of California, Department of Information and Computer Science, (1998).
11. Hilderman, R. J. and Hamilton, H. J.: Knowledge Discovery and Measure of Interest, Kluwe Academic Publishers (2001)
12. Hinton, G. E.: "Learning distributed representations of concepts", *Proceedings of 8th Annual Conference of the Cognitive Science Society*, Amherest, MA. REprinted in R.G.M.Morris (ed.) (1986)
13. Holte, R. C.: Very simple classification rules perform well on most commonly used datasets, Machine Learning, Vol. 11 (1993) 63–91
14. Klösgen, W.: Explora: A Multipattern and Multistrategy Discovery Assistant. in Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy R. (Eds.): Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, California (1996) 249–271
15. Ohsaki, M., Kitaguchi, S., Kume, S., Yokoi, H., and Yamaguchi, T.: Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis, in Proc. of ECML/PKDD 2004, LNAI3202 (2004) 362–373
16. Piatetsky-Shapiro, G.: Discovery, Analysis and Presentation of Strong Rules. in Piatetsky-Shapiro, G., Frawley, W. J. (eds.): Knowledge Discovery in Databases. AAAI/MIT Press (1991) 229–248
17. Platt, J.: Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, eds., MIT Press (1999) 185–208
18. Quinlan, R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993)
19. Rijsbergen, C.: Information Retrieval, Chapter 7, Butterworths, London, (1979) http://www.dcs.gla.ac.uk/Keith/Chapter.7/Ch.7.html
20. Smyth, P., Goodman, R. M.: Rule Induction using Information Theory. in Piatetsky-Shapiro, G., Frawley, W. J. (eds.): Knowledge Discovery in Databases. AAAI/MIT Press (1991) 159–176
21. Tan, P. N., Kumar V., Srivastava, J.: Selecting the Right Interestingness Measure for Association Patterns. Proc. of Int. Conf. on Knowledge Discovery and Data Mining KDD-2002 (2002) 32–41
22. Witten, I. H and Frank, E.: DataMining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, (2000)
23. Yao, Y. Y. Zhong, N.: An Analysis of Quantitative Measures Associated with Rules. Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining PAKDD-1999 (1999) 479–488
24. Zhong, N., Yao, Y. Y., Ohshima, M.: Peculiarity Oriented Multi-Database Mining. IEEE Trans. on Knowledge and Data Engineering, 15, 4, (2003) 952–960