A Rule Evaluation Support Method with Learning Models Based on Objective Rule Evaluation Indexes

Hidenao Abe¹, Shusaku Tsumoto¹, Miho Ohsaki², and Takahira Yamguchi³ ¹Department 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 ²Faculty of Engineering, Doshisha University ³Faculty of Science and Technology, Keio University

Abstract

In this paper, we present a novel rule evaluation support method for post-processing of mined results with rule evaluation models based on objective indexes. Post-processing of mined results is one of the key issues to make a data mining process successfully. 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 procedures, we have developed the rule evaluation support method with rule evaluation models, which are obtained with objective rule evaluation indexes and evaluations of a human expert for each rule. Since the method is needed more accurate rule evaluation models, we have compared learning algorithms to construct rule evaluation models with the actual meningitis data mining result and actual rule sets from UCI datasets. Then we show the availability of our adaptive rule evaluation support method.

1. Introduction

In recent years, data mining techniques have been widely known as a hopeful process to utilize huge data which are easily stored on information systems in various fields. Especially, IF-THEN rule extraction methods 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 crucial rule evaluation procedure, many efforts have done to select valuable ones using an objective rule evaluation index such as recall, precision, and other interestingness measurements[6, 12, 14] (we call them 'objective indexes' later). However, it is also difficult to estimate a criterion of a human expert with single objective index, because subjective criterion of the expert is influenced by the amount of his/her knowledge and/or a passage of time.

To above issues, we have developed an adaptive rule evaluation support method for human experts with rule evaluation models, which predict criteria of a human expert based on multiple objective indexes. In Section 2, we explain the rule evaluation model construction method concretely. Then we present a performance comparison of learning algorithms for constructing rule evaluation models in Section 3. Finally, we discuss about the availability of our rule evaluation model construction approach.

2. Rule Evaluation Support with Rule Evaluation Model based on Objective Indexes

We considered the process of modeling rule evaluations of human experts as the process to clear up relationships between the human evaluations and features of input ifthen rules. With this consideration, we have decided to design such rule evaluation process as a learning task. Figure1 shows the process of rule evaluation model construction based on re-use of human evaluations and objective indexes for each mined rule. At the training phase, attributes of a meta-level training data set is obtained by objective indexes 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 indexes. Then the evaluation for the new rules became new evaluation labels for another rule evaluation model training phase.



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

3. Performance Comparison of Learning Algorithms for Rule Model Construction

To predict human evaluation labels of a new rule based on objective indexes more exactly, the method have to construct a better rule evaluation model. So we evaluate our method from the following three viewpoints: accuracies of rule evaluation models, learning curves of learning algorithms, and contents of learned rule evaluation models.

To evaluate our approach, we have taken the dataset from the mining result of a meningitis data mining[4]. Then to confirm the performance of our approach on the other datasets, we have also taken the rule sets from three UCI benchmark datasets[5]. To construct a dataset to learn a rule evaluation model, values of objective indexes have been calculated for each rule, taking 39 objective indexes investigated by Ohsaki et. al [9].

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[13]: C4.5 decision tree learner[11] called J4.8, neural network learner with back propagation (BPNN)[7], support vector machines (SVM)¹[10], classification via linear regressions (CLR)²[1], and OneR[8].

3.1. Constructing Rule Evaluation Models on an Actual Datamining Result

In this case study, we have taken 244 rules, which are mined from a dataset consisted of appearances of meningitis patients and six kinds of diagnosis as shown in Table1. For each rule, we labeled three evaluations (I:Interesting, NI:Not-Interesting, NU:Not-Understandable), according to comments from a medical expert.

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

Table 1. Number of rules obtained by the meningitis datamining result and distribution of their class.

	On the whole training dataset									
		Recall of			Precision of					
	Acc.	Ι	NI	NU	Ι	NI	NU			
J4.8	85.7	41.7	97.9	66.7	80.0	86.3	85.7			
BPNN	86.9	81.3	89.8	55.6	65.0	94.9	71.4			
SVM	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			
OneR	82.0	56.3	92.5	0.0	57.4	87.8	0.0			
	Leave-One-Out									
		Recall of			Precision of					
	Acc.	Ι	NI	NU	Ι	NI	NU			
J4.8	79.1	29.2	95.7	0.0	63.6	82.5	0.0			
BPNN	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			
CLR	80.3	35.4	95.7	0.0	60.7	82.9	0.0			

Table 2. Accuracies(%), Recalls(%) and Precisions(%) of the five learning algorithms.

3.1.1. Comparison on Classification Performance The performances of the five learning algorithms to the whole training dataset and the results of Leave-One-Out(LOO) are also shown in Table2. LOO is a deterministic evaluation method to measure a robustness of learning algorithms to another unknown dataset.

As shown in Table2, all of the accuracies, Recalls of I and NI, and Precisions of I and NI are higher than predicting default labels. In addition, the learning algorithms, which combine multiple objective indexes, achieve equal or higher performance than that of OneR. Looking at Recalls on class I, BPNN have achieved the highest performance. As for the other algorithms, they have tended to be learned classification patterns for the major class NI. On the accuracies of LOO, these learning algorithms have achieved from 75.8% to 81.9%, excepting that of minor class NU. Although OneR selects an adequate objective index to sort and classify 244 training datasets, it is not always work well to select just one objective index.

3.1.2. Learning Curves of the Learning Algorithms We have investigated learning curves of each learning algorithm to estimate how many evaluations are needed to construct an adequate rule evaluation model, because our method requires evaluations of mined rules by a human expert at least once. The upper table in Figure2 shows accuracies to the

¹ The kernel funcution was set up polynomial kernel.

² We set up the elimination of collinear attributes and the model selection with greedy search based on Akaike Informatio Metric.



Figure 2. Learning Curves of accuracies(%) on the learning algorithms with sub-sampled training dataset.

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 Figure2. As shown in these results, SVM and CLR, which learn hype-planes, achieves grater than 95% with only less than 10% of training subset. Besides, J4.8 and BPNN need more training data to learn accurate classifiers than SVM and CLR.

3.1.3. Rule Evaluation Models on the Actual Datamining Result Dataset Figure3 shows the readable rule evaluation models of the meningitis data mining result learned with OneR, J4.8 and CLR. The rule set of OneR is shown



Figure 3. Learned models to the meningitis data mining result dataset.

in Figure3(a). OneR has selected YLI1[14] to classify the evaluation labels. Although YLI1 was suggested to predict interestingness of a human expert, YLI1 estimates a correctness of each rule on a validation dataset. As shown in Figure3(b), J4.8 leaned the decision tree. At upper level includ-

ing the root node, this model takes Laplace Correction[12], Accuracy, Precision and Recall, which represent correctness of rules. Coverage and Prevalence represent a generality of the antecedent and the consequent of a rule. At the lower level of the tree takes GOI[3] and Peculiarity[15], which calculates specificity of each rule compared to other rules. Figure3(c) shows linear models to classify each class. As for models to class NI and I, they have the same indexes such as Precision, Certainty Factor, PSI, and Peculiarity with opposite coefficients. The strongest factors on these models are Precision and Gini Gain, which increase their values with the correctness of a rule. To class NU, the strongest factor is Leverage based on Precision with a correction using a generality of a rule.

3.2. Constructing Rule Evaluation Models on Artificial Evaluation Labels

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

We selected the following three datasets: Mushroom, Heart, and Internet Advertisement Identification (called InternetAd later). With these datasets, we obtained rule sets with bagged PART, which repeatedly executes PART[2] to bootstrapped training sub-sample datasets. To these rule sets, we calculated the 39 objective indexes as attributes of each rule. As for the class of these datasets, we set up three class distributions with multinomial distribution. 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). Table3(A) shows us the datasets with three different class distributions.

3.2.1. 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. Table3(B) shows the accuracies of the five learning algorithms to each class distribution of the three datasets. J48 and BPNN always work better than just predicting a default class.

3.2.2. Evaluation on Learning Curves We also executed learning curve analysis on these artificial data sets as shown in Section 3.1.2. Table3(C) shows sizes of minimum training subset, which needs to go beyond percentages of a default class by each learning algorithm. Although the performances SVM, CLR and OneR suffered from probabilistic class distributions of the datasets, the trend of that SVM and CLR can achieve as good performance as its of whole training dataset earlier is not suffered.



Table 3. (A)Datasets of the rule sets learned from the UCI benchmark datasets, (B)Accuracies(%) on whole training datasets labeled with three different distributions, (C)Number of minimum training sub-samples to outperform %Def. class.

4. 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 indexes, 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 on both of the actual and artificial dataset. We also investigated the amount of training sub-samples to construct more accurate rule evaluation models with less samples. The result shows that we can construct enough models with only 10%. These results indicate that our method can support human rule evaluations with adaptive rule evaluation models based on objective indexes.

As future works, 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 indexes, which evaluate whole mining results.

Acknowledgement

This research is supported by the Grant-in-Aid for Young Scientists(B), 17700152, by Ministry of Education, Science, and Culture for Japan.

References

- 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
- [2] 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
- [3] 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

- [4] 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
- [5] 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).
- [6] Hilderman, R. J. and Hamilton, H. J.: Knowledge Discovery and Measure of Interest, Kluwe Academic Publishers (2001)
- [7] 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)
- [8] Holte, R. C.: Very simple classification rules perform well on most commonly used datasets, Machine Learning, Vol. 11 (1993) 63–91
- [9] 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
- [10] 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
- [11] Quinlan, R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993)
- [12] 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
- [13] Witten, I. H and Frank, E.: DataMining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, (2000)
- [14] 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
- [15] Zhong, N., Yao, Y. Y., Ohshima, M.: Peculiarity Oriented Multi-Database Mining. IEEE Trans. on Knowledge and Data Engineering, 15, 4, (2003) 952–960