Evaluating a Rule Evaluation Support Method with Learning Models Based on Objective Rule Evaluation Indices - A Case Study with a Meningitis Data Mining Result -

Hidenao Abe
Department of Medical Informatics, Shimane University, School of Medicine
abe@med.shimane-u.ac.jp

Shusaku Tsumoto
Department of Medical Informatics, Shimane University, School of Medicine
tsumoto@computer.org

Miho Ohsaki
Faculty of Engineering, Doshisha University
mohsaki@mail.doshisha.ac.jp

Takahira Yamaguchi
Faculty of Science and Technology, Keio University
yamaguti@ae.keio.ac.jp

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 indices. 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 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. Then we discuss the availability of our rule evaluation support method.

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. In addition, rule selection methods have been never explicitly re-used the history of each rule evaluation such as focused items and relationships between items, which is only stored in his/her mind of the human expert.

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 with the actual meningitis dataset [9] in Section 4. With this result of the comparison, we discuss about the availability of our rule evaluation model construction approach.

2. Related Work

To avoid the confusion of real human interests, objective indices, and subjective indices, we clearly define them as follows: **Objective Indices:** The feature such as the correctness, uniqueness, and strength of a rule, calculated by

the mathematical analysis. **Subjective Indices:** The similarity or difference between the information on interestingness given beforehand by a human expert and those obtained from a rule. **Real Human Interests:** The interest felt by a human expert for a rule in his/her mind.

Focusing on interesting rule selection with objective indices, researchers have developed more than forty objective indices based on number of instances, probability, statistics, information quantity, distance of rules or their attributes, and complexity of a rule [10, 20, 22]. Most of these indices 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 indices estimate how a rule fits with a belief, a bias or a rule template formulated beforehand by a human expert. Although these subjective indices 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 indices 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[14] investigated the relation between objective indices 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 indices 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 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. Figure1 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.

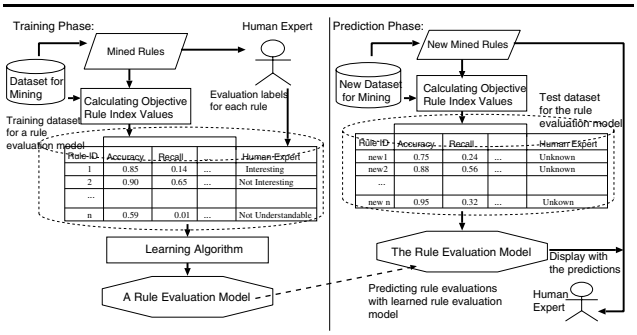


Figure 1. Overview of the construction method of rule evaluation models.

4. Performance Comparison of Learning Algorithms to Construct Rule Evaluation Models

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 mining result of a meningitis data mining. Then we discuss about the availability of our method from the following three viewpoints: accuracies of rule evaluation models, learning curves of 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. 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. As for elements of the rule evaluation models to the whole dataset, we consider the characteristics of each learning algorithm on the attribute space consisted of the objective indices.

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. The 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.

In this case study, we have taken 244 rules, which are

Theory	Index Name (**Abbreviation**)
P	Coverrage(**Coverage**), Prevalence(**Prevalence**) Precision(**Precision**), Recall(**Recall**) Suppurt(**Support**), Specificity(**Specificity**), Accuracy(**Accuracy**), Lift(**Lift**) Leverage(**Leverage**), Added Value(**Added Value**)[20], Klösgen's Interestingness(**KI**)[13] Relative Risk(**RR**)[1], Brin's Interest(**BI**)[2], Brin's Conviction(**BC**)[2], Certainty Factor(**CF**)[20] Jaccard Coefficient(**Jaccard**)[20], F-Measure(**F-M**)[18], Odds Ratio(**OR**)[20], Credibility(**Credibility**)[8] Yule's Q(**YuleQ**)[20], Yule's Y(**YuleY**)[20], Kappa(**Kappa**)[20], Collective Strength(**CST**)[20] Gray andOrlowska's Interestingness weighting Dependency(**GOI**)[7], Gini Gain(**Gini**)[20]
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**)[19], K-Measure(**K-M**)[14], Mutual Information(**MI**)[20] Yao and Liu's Interestingness 1 based on one-way support(**YLI1**)[22] Yao and Liu's Interestingness 2 based on two-way support(**YLI2**)[22] Yao and Zhong's Interestingness(**YZI**)[22]
N	Cosine Similarity(**CSI**)[20], ϕ Coefficient(ϕ)[20] Laplace Correction(**LC**)[20], Piatetsky-Shapiro's Interestingness(**PSI**)[15]
D	Gago and Bento's Interestingness(**GBI**)[5], Peculiarity(**Peculiarity**)[23]

Table 1. The objective rule evaluation indices for classification rules. 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.

mined from a dataset consisted of appearances of meningitis patients and six kinds of diagnosis as shown in Table2. Although the rules evaluated with displaying recall and precision of each rule, the medical expert was not supported any method such as sorting with an objective index and cutting off some rules with an objective index. For each rule, we labeled three evaluations (I:Interesting, NI:Not-Interesting, NU:Not-Understandable), according to evaluation comments from the 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 2. Number of rules obtained by the meningitis datamining result and distribution of their class.

To this 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 [21]: C4.5 decision tree learner [17] called J4.8, neural network learner with back propagation (BPNN) [11], support vector machines (SVM)[1] [16], classification via linear regressions (CLR)[2] [3], and One R[12].

1 The kernel function 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.

4.1. Comparison on Classification Performance

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 instance as the test data and remains as the training data repeatedly for each instance of a given dataset, we can evaluate the performance of a learning algorithm to a new dataset in deterministic.

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

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.

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

	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

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

Robustness with Leave-One-Out Estimation Each value of Leave-One-Out estimation shows robustness of each learning algorithm to an unknown test dataset. On the accuracies, 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, which is a minor class label in this dataset.

Looking at each learning algorithm, the values of BPNN show the trend of over fitting, comparing with its values of training dataset and its values of Leave-One-Out. Although OneR selects an adequate objective index to sort and classify 244 training datasets, it shows that the selection of just one objective index limits the prediction performance to a new dataset.

4.2. 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 how many evaluations are needed to construct an adequate rule evaluation model. The upper table in Figure2 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 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. Although decision tree learner and BPNN could learn better classifier to the whole dataset than these hyper-plane learners, they need more training data to learn accurate classifiers.

To eliminate known ordinary knowledge from large rule set, it is needed to classify non-interesting rules correctly. The upper table in Figure3 shows percentages of recalls on NI. The lower chart in Figure3 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 rule evaluations by a human expert. This is guaranteed with no less than 80% precisions of all learning algorithms.

% Accuracy to the whole training dataset

%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

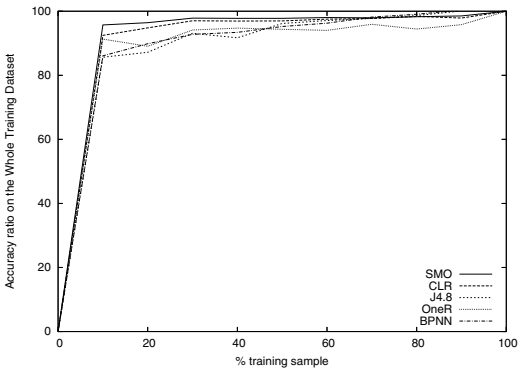


Figure 2. Learning Curves of accuracies(%) on the learning algorithms with sub-sampled training dataset.

4.3. 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.

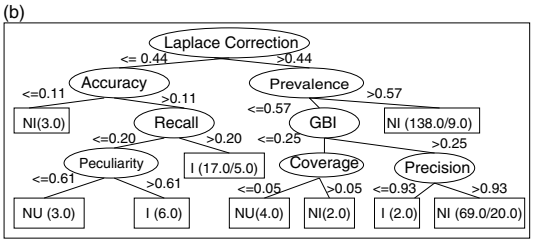
The rule set of OneR is shown in Figure4(a). OneR has selected YLI1 [22] to classify the evaluation labels. Although YLI1 corrects support to predict interestingness of a human expert, YLI1 estimates a correctness of each rule on a validation dataset.

As shown in Figure4(b), J4.8 learned the decision tree. At the root node, this model takes Laplace Correction [20], which is a corrected Precision with constant values. At the other levels, it takes indices evaluating a correctness of a rule such as Accuracy, Precision and Recall. Coverage and Prevalence are indices to evaluate a generality of the antecedent and the consequent of a rule. GOI [7] calculate index values with the classification result of a rule. Peculiarity [23] sums up differences of antecedents between one

(a)

```
IF                                YLI1 < 0.02 THEN "I"
IF YLI1 >= 0.02 and YLI1 < 0.29 THEN "NI"
IF YLI1 >= 0.29 and YLI1 < 0.43 THEN "I"
IF YLI1 >= 0.43 and YLI1 < 0.44 THEN "NI"
IF YLI1 >= 0.44 and YLI1 < 0.55 THEN "I"
IF YLI1 >= 0.55 and YLI1 < 0.63 THEN "NI"
IF YLI1 >= 0.63 and YLI1 < 0.83 THEN "I"
IF YLI1 >= 0.83                 THEN "NI"
```

(b)



(c)

```
NU =
     0.6202 * Specificity +
     0.6224 * Accuracy +
    -1.1384 * Leverage +
    -0.6895 * RelativeRisk +
     0.3704 * CertaintyFactor +
     0.5722 * OddsRatio +
     0.7656 * BI +
    -0.222  * Credibility +
    -0.3941 * LaplaceCorrection +
     0.7986 * GiniGain +
    -0.0966 * GBI +
    -0.8895

NI =
     1.7173 * Precision +
    -0.5063 * Accuracy +
     0.5673 * RelativeRisk +
    -1.2718 * CertaintyFactor +
     0.5955 * YulesQ +
    -0.4609 * K-Measure +
     0.4613 * PSI +
    -0.4181 * Peculiarity +
     0.5302

I =
    -1.4417 * Precision +
    -0.7286 * Specificity +
     0.4085 * Lift +
     0.6297 * CertaintyFactor +
    -1.4477 * CollectiveStrength +
     1.5449 * GiniGain +
    -0.5318 * PSI +
     0.4981 * Peculiarity +
     1.4872
```

Figure 4. Learned models to the meningitis data mining result dataset: (a) rule set learnd from OneR, (b) decision tree learned from J4.8, (c) linear regression models learned from CLR.

% Recall of NI to the whole training dataset

% 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	92.3	93.4	92.7	92.1	96.3

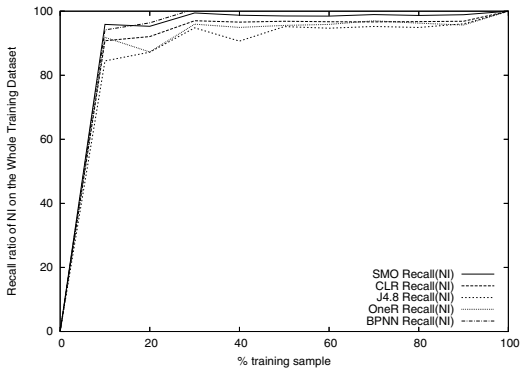


Figure 3. Learning Curves of recalls(%) for NI on the learning algorithms with sub-sampled training dataset.

rule and the other rules in the same rule set.

Figure4(c) shows linear models to classify each class. The prediction has done with integrating the responses of these linear models. As for models to class NI and I, they have the same indices 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.

4.4. Discussion

On the Classification Performances As shown in Table3, J4.8 decision tree learner and BPNN neural network learner work better than the other algorithms on both of the actual problems. The classification result about class I indicates that these instances are difficult to separate with liner expressions in this attribute space based on the 39 objective indices. To predict such labels correctly, we should apply nonlinear classifier learned from nonlinear learners.

Although these five learning algorithms have achieved 81.6% of the highest accuracy in the Leave-One-out estimation, we need to obtain more accurate rule evaluation models with meta-learning algorithms such as boosting, bagging and so forth.

On the Learning Curves With this analysis of the learning curves about each amount of training samples, we consider the following guideline: At early stage of rule evaluation support, the system should select hyper-plane learners to construct better rule evaluation models rapidly. Then closing stage of evaluations, the system should select more accurate learning algorithm to predict minor but valuable rules.

On the Learned Rule Evaluation Models 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. This indicates that the rule model construction method needs to select prior algorithms on data pre-processing procedures and rule evaluation model learning algorithms.

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 on 39 objective indices, rule evaluation models have achieved higher accuracies than random predictions. In the estimation of robustness for a new rule with Leave-One-Out, we have achieved more than 75.8% of accuracies with these learning algorithms. On the evaluation with learning curves to the whole training dataset, SVM and CLR have achieved more than 95% of achievement ratio compared to the accuracy of the whole training dataset with just 10% of subset as their training dataset. These result related to performances of rule evaluation models indicate the availability of this rule evaluation support for a human expert. The rule evaluation models of the learning algorithms to the dataset include different objective indices for different learning models. We will be needed to select a proper attribute set for each learning algorithm to construct better rule evaluation model.

As future works, we will introduce a selection method of learning algorithms to construct a proper rule evaluation model according to each situation. At the same time, we will apply this rule evaluation support method to other data mining results from different kind of domains not only medical domains but also business domain and so forth.

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] Hilderman, R. J. and Hamilton, H. J.: Knowledge Discovery and Measure of Interest, Kluwe Academic Publishers (2001)

[11] 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)

[12] Holte, R. C.: Very simple classification rules perform well on most commonly used datasets, Machine Learning, Vol. 11 (1993) 63–91

[13] 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

[14] 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

[15] 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

[16] 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

[17] Quinlan, R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993)

[18] C. Rijsbergen: Information Retrieval, Chapter 7, Butterworths, London, (1979)

[19] 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

[20] 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

[21] Witten, I. H and Frank, E.: DataMining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, (2000)

[22] 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

[23] Zhong, N., Yao, Y. Y., Ohshima, M.: Peculiarity Oriented Multi-Database Mining. IEEE Trans. on Knowledge and Data Engineering, 15, 4, (2003) 952–960