Investigation of Rule Interestingness in Medical Data Mining

Miho OHSAKI†, Yoshinori SATO†, Shinya KITAGUCHI†, Hideto YOKOI††, and Takahira YAMAGUCHI†

† Faculty of Information, Shizuoka University, Graduate School of Information, Sizuoka University
Johoku 3–5–1, Hamamatsu-shi, Shizuoka, 432–8011 Japan
†† Department of Medical Informatics, Chiba University Hospital
Inohana 1–8–1, Chuo-ku, Chiba-shi, Chiba, 260–0856 Japan
E-mail: †{miho,cs8037,cs9026,yamaguti}@cs.inf.shizuoka.ac.jp, ††yokoih@telemed.ho.chiba-u.ac.jp

Abstract This research experimentally investigates the performance of conventional rule interestingness measures and discusses their availability to supporting KDD through system-human interaction. We compared the evaluation results by a medical expert and that by several selected measures for the rules discovered from the medical test data on chronic hepatitis. The measure on antecedent-consequent dependency using all instances showed the highest performance, and ones on the both of dependency and generality the lowest under this experimental condition. The whole trend of the experimental results indicated that the measures detected really interesting rules at a certain level and offered us the rough guideline to apply them to system-human interaction.

Key words Medical Data Mining, Time Series, Rule Interestingness, System-Human Interaction, Chronic Hepatitis

医療データマイニングにおけるルールの興味深さの検討

大崎　美穂†　佐藤　芳紀†　北口　真也†　横井　英人††　山口　高平†

† 静岡大学情報学部，静岡大学大学院情報学研究科　〒432–8011 静岡県浜松市城北 3–5–1
†† 千葉大学医学部附属病院医療情報部　〒260–8677 千葉県千葉市中央区亥鼻 1–8–1
E-mail: †{miho,cs8037,cs9026,yamaguti}@cs.inf.shizuoka.ac.jp, ††yokoih@telemed.ho.chiba-u.ac.jp

あらまし 本研究では，実証実験によってルールの興味深さ指標の性能を調べ，システムと人間のインタラクションを通した知識発見支援に利用できるかを検討した．実験では，慢性肝炎の検査データから得たルールに対して，専門家，および，今回選定した指標が与えた評価を比較した．その結果，本実験条件では，全事例を用いて条件部と結論部の依存性を算出する指標の性能が高く，依存性と一般性の両方を考慮する指標の性能が低かった．全体傾向としては，今回の指標は専門家が求める興味深いルールをある程度検出できると分かり，システムと人間のインタラクションへの応用の指針が得られた．

キーワード 医療データマイニング，時系列，ルールの興味深さ，システムと人間のインタラクション，慢性肝炎

1. Introduction

The concern with the contribution of data mining to Evidence-Based Medicine (EBM) has been growing for the last several years, and there have been many medical data mining studies [18], [24]. It is experientially known as the important factor influencing on discovered knowledge quality how to pre-/post-process real ill-defined clinical data and how to polish up rules through the interaction between a system and its human user. However, it has not been discussed enough [1].

Thus, to discuss the pre-/post-processing and the system-human interaction in medical domain, we have been conducting case studies using medical test data on chronic hepatitis. We estimated that the temporal patterns of medical test results would be useful for a medical expert to grasp diagnosis and predict prognosis. We then obtained the graph-based rules predicting the future pattern of GPT, one of major medical tests. We iterated the rule generation by our mining system and the rule evaluation by a medical expert two times [25].

As the results, we obtained the knowledge on the rule interestingness for a medical expert and learned the lessons on the pre-/post-processing and the system-human interaction in medical domain. We then focused on system-human interaction and proposed its concept model and its semi-automatic system framework [26]. These case studies made us recognize the significance of clarifying the rule interestingness really required by a human user and that of feedbacking such information to a mining system.

Therefore, this research has the following two purposes: **(1)** investigating the conventional interestingness measures in Knowledge Discovery in Databases (KDD) and comparing them with the rule evaluation results by a medical expert, and **(2)** discussing whether they are available to support system-human interaction in medical domain.

In this paper, Section 2. introduces conventional interestingness measures and shows the selected several measures suitable to our purpose. Section 3. notes the experimental conditions and results to evaluate the rules on chronic hepatitis with the measures and to compare them with the evaluation results by a medical expert. In addition, it discusses the availability of the measures for system-human interaction support. Finally, Section 4. concludes the paper and comments on the future work.

2. Related Work

2.1 Outcome of Our Previous Research

We have conducted the case studies to discover the rules on diagnosis and prognosis from a chronic hepatitis data set. The set of the rule generation by our mining system and the rule evaluation by a medical expert was iterated two times and leaded us to discover the rules valued as interesting ones by the medical expert.

We used the data set of the medical test results on viral chronic hepatitis [12]. Before mining, we finely pre-processed it based on medical expert's advice since such a real medical data set is ill-defined and has many noises and missing values. We then extracted the representative temporal patterns from the data set by clustering and generated the rules consisting of the patterns and predicting the prognosis by a decision tree [2].

Figure 1 shows one of the rules, which the medical expert focused on, obtained in the first mining. It estimates the future trend of GPT, one of major medical tests to grasp chronic hepatitis symptom, in the future one year by using the change of several medical test results in the past two years. The medical expert commented on it as follows: the rule offers a hypothesis that GPT value changes with about a three-years cyclic, and the hypothesis is interesting since it differs from the conventional common sense of medical experts that GPT value basically decreases in a monotone.

We then improved our mining system, extended the observation term, and generated new rules. Figure 2 shows two of the rules, which the medical expert valued, obtained in our second mining. The medical expert commented on them that they imply GPT value globally changes two times in the past five years and more strongly support the hypothesis of GPT's cyclic change.

The literature [25] explains the details of our previous research, namely the pre-processing methods, the developed system, and the process of rule generation and evaluation. Refer it if you need to know the details.

As the next research, we tried to systematize the knowledge obtained in the previous research, on the pre-/post-processing suitable to medical data and the system-human interaction to polish up rules. Especially on system-human interaction, we formulated its concept model that describes

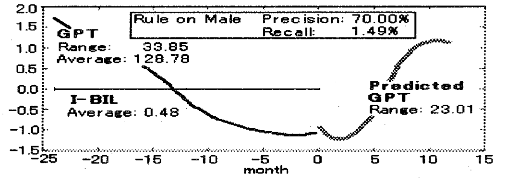

図 1 第一回目のマイニングにおいて高く評価されたルールの一例.
Fig. 1 Rule valued by a medical expert in the first mining.

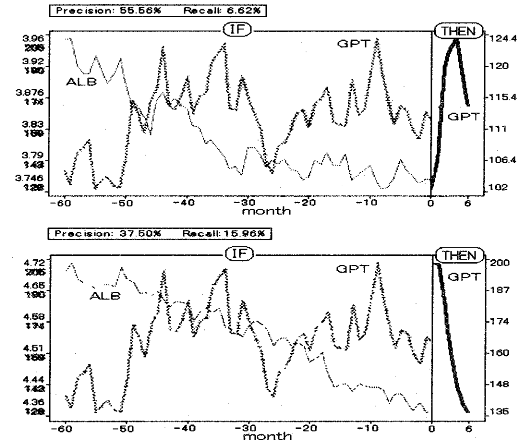

図 2 第二回目のマイニングにおいて高く評価されたルールの一例.
Fig. 2 Rules valued by a medical expert in the second mining.

the roles and the functions of a system and a human user (See Figure 3) and the framework to support semi-automatic system-human interaction based on the model (See Figure 4).

As shown in Figure 3, a mining system discovers the rules faithfully to the data and offers them to a medical expert as the materials for hypothesis generation and justification. While, the medical expert generates and justifies a hypothesis, a seed of new knowledge, by evaluating the rules based on his/her domain knowledge. A system to support such interaction requires the function to generate and present rules to a medical expert based on their validity at the viewpoints of objective data structure and subjective human evaluation criteria. The flow of "System Evaluation" and "Human Evaluation" in Figure 4 means that.

Note that the word 'justification' in this paper does not mean the highly reliable proof of a hypothesis by additional medical experiments under strictly controlled conditions. It means the additional information extraction to enhance the reliability of an initial hypothesis from the same data.

The literature [26] explains the details of the concept model of system-human interaction and the framework for semi-automatic system-human interaction. Refer it to know the details.

These our researches notified us that it is required for realizing the framework in Figure 4 to investigate the rule interestingness measures available for "System Evaluation" and the relation between "System Evaluation" and "Human Evaluation". Therefore, this research selects several conventional measures and compares the rule evaluation results by them with "Human Evaluation", namely that by a medical expert.

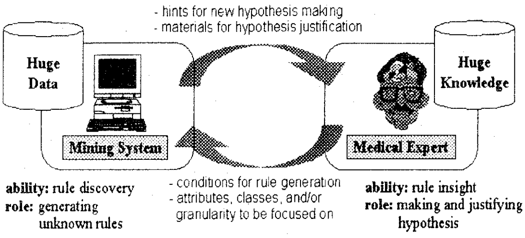

図 3 概念レベルでのシステムと専門家のインタラクションモデル.

Fig. 3 Interaction model between a system and a human expert at a concept level.

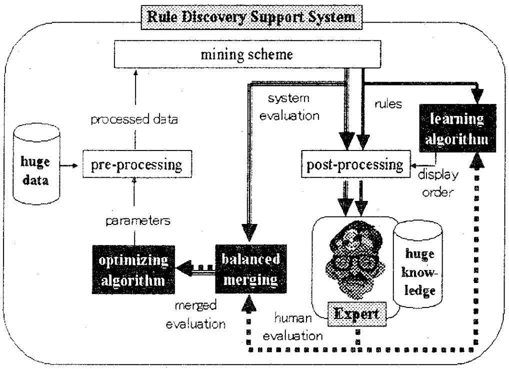

図 4 インタラクションの半自動化フレームワーク.

Fig. 4 Framework for semi-automatic system-human interaction.

2.2 Rule Interestingness Measures

Rule interestingness is one of active research fields in KDD. There have been many studies to formulate interestingness measures and to evaluate rules with them instead of humans. Interestingness measures are categorized into objective and subjective ones. Objective measures mean how a rule is mathematically meaningful based on the distribution structure of the instances related to the rule. Subjective measures mean how a rule fit with a belief, a bias, or a rule template formulated beforehand by a human user [13].

Objective measures are mainly used to remove meaningless rules at the viewpoint of data structure rather than to discover really interesting ones for a human user, since they do not include domain knowledge [3], [5], [6], [9], [15], [22], [23], [28], [31], [33]. On the other hand, subjective measures are available to discover really interesting rules to some extend due to their built-in domain knowledge. However, they depend on the precondition that a human user can clearly formulate his/her own interest [16], [17], [20], [21], [27], [29], [30]. Few subjective measures adaptively learn real human interest through system-human interaction [7].

The conventional interestingness measures, not only objective ones but also subjective ones, do not directly reflect the interest that a human user really has. To avoid the confusion of real human interest and the interestingness measures, we define them as follows (Note that while we define "Real Human Interest" by ourselves, the definitions of the other terms are based on many conventional studies on interestingness measures):

Objective Measure: The feature such as correctness, uniqueness, etc. of a rule or a set of rules, mathematically calculated using data structure. It does not include human evaluation criteria. **Subjective Measure:** The similarity or the difference between the information on interestingness beforehand given by a human user and that obtained from a rule or a set of rules. Although it includes human evaluation criteria in its initial state, its calculation of similarity or difference is mainly based on data structure. **Real Human Interest:** The interest in a rule, which a human user really feels in his/her mind. It is formed from the synthesis of human natural cognition, individual domain knowledge and experiences, and the influences of the rules that the human user evaluated before.

2.2.1 Objective Measures

Objective measures are the mathematical analysis results of data distribution structure. There are many objective measures, and they can be categorized into some groups using evaluation criterion, evaluation target, and theory for analysis. The evaluation target means whether an objective measure evaluates a rule or a set of rules. This research deals with not the objective measures for a set of rules [10], [13] but ones for a rule, because it focuses on the quality of each rule.

Table 1 shows some major objective measures. They assume one of the following evaluation criteria and examine how a rule matches with the criteria by calculating the instance distribution difference between the data and the rule or between the antecedent and the consequent of the rule.

Correctness: How many instances the antecedent and/or the consequent of a rule supports, or how strong their dependence is [15], [22], [23], [28], [31]. **Information Plentifulness:** How much information a rule possesses [9]. **Generality:** How similar the trend of a rule is to that of all data [6]. **Uniqueness:** How different the trend of a rule is from that of all data [3], [5], [33] or the other rules [6], [23].

Although objective measures are useful to automatically remove obviously meaningless rules, some evaluation criteria have the contradiction to each other such as generality and uniqueness. In addition, the evaluation criterion of an objective measure may not match with or may contradict real human interest. For example, a rule with a plenty of information may be too complex for a human user to understand. Many of the objective measure proposers showed the validity of their measures with mathematical proofs or the experimental results using benchmark data. However, they hardly conducted the comparison between their measures and the other ones or the investigation of the relation between their measures and real human interest for a concrete application.

2.2.2 Subjective Measures

Subjective measures are the similarity or the difference between the information given by a human user and that by a rule. There are several subjective measures, and they can be categorized into some groups using human evaluation criterion, method to give information from a human user to a mining system, theory for calculating similarity or difference [4], [16], [17], [19]~[21], [27], [30]. We do not mention the details of subjective measures, because this research mainly focuses on the relation between objective measures and real human interest.

2.2.3 Selection of Interestingness Measures

This research experimentally investigates the availability of objective measures by comparing them with real human interest in medical domain. As mentioned in Section 2.2.1, the evaluation criteria of objective measures are obviously not the same of humans, since objective measures do not include the knowledge on rule semantics. However, they may be available to support the KDD through system-human in-

表 1 ルールの興味深さの客観的指標の一覧. 本研究で用いたものには*を付けた. Calculation の列における以下の記号は, 指標の算出に何を使うかを意味する. **N:** ルールの条件部, 結論部に含まれる事例数. **P:** ルールの条件部, 結論部の生起確率. **S:** ルールの条件部, 結論部の生起確率に基づく統計量. **I:** ルールの条件部, 結論部の情報量. **D:** ルールの属性値に基づくルール間の距離. **C:** ルールの木構造の複雑さ.

Table 1 List of the objective measures of rule interestingness. The measures used in this research have the symbol '*'. The following symbols in the column 'Calculation' mean what is used to calculate the measure. **N:** Number of instances included in the antecedent and/or the consequent. **P:** Probability of the antecedent and/or the consequent. **S:** Statistical variable based on **P**. **I:** Information of the antecedent and/or the consequent. **D:** Distance of the rule from the other rules based on rule attributes. **C:** Complexity of the tree structure of the rule.

Name	Calculation	Evaluation Criterion
Rule Interest [28]	N	Dependency between the antecedent and the consequent
Support *	P	Generality of the rule
Precision (Confidence) *	P	Performance of the rule to predict the consequent
Recall *	P	Performance of the rule not to leak the consequent
Accuracy	P	Summation of the precision and its converse of contrapositive
Lift *	P	Dependency between the antecedent and the consequent
Leverage *	P	Dependency between the antecedent and the consequent
Reliable Exceptions [22]	P	Rule with small support and high precision
Gray and Orlowska's measure (GOI) * [8]	P	Multiplication of the support and the antecedent-consequent dependency
Surprisingness [5]	P	Rule occurring Sympson's paradox
χ^2 measure 1 * [23]	S	Dependency between the antecedent and the consequent
χ^2 measure 2 [23]	S	Similarity between two rules
J-Measure [31] *	I	Dependency between the antecedent and the consequent
General Measure [15]	S & I	Fusion of the χ^2 measure 1 and the information gain measure
Distance Metric [6]	D	Distance of the rule from the rule with the highest coverage
Dong and Li's measure [3]	D	Distance of the rule from the other rules
Peculiarity [33]	D	Distance of the attribute value from frequent attribute values
I-Measure [9]	C	Complexity of the rule

teraction if they possess a certain level of performance to detect really interesting rules. That is the motivation of this research. The investigation of subjective measures will be our future work.

From the objective measures shown in Table 1, we selected the followings as the investigation targets: the most popular ones (Support, Precision, Recall, Lift, and Leverage), probability-based one (GOI [8]), statistics-based one (χ^2 measure 1 [23]), and information-based one (J-Measure [31]).

3. Comparison between Objective Measures and Real Human Interest

3.1 Experimental Conditions

In our previous researches, we repeated the data mining process two times using a dataset of chronic hepatitis and generated a set of rules for each mining (Refer Section 2.1). After each mining, a medical expert evaluated the rules and gave each rule one of the following rule quality labels: 'Especially-Interesting', 'Interesting', 'Not-Understandable', and 'Not-Interesting'. 'Especially-Interesting' means that the rule was a key factor to generate the hypothesis of GPT's cyclic change in the first mining or to justify it in the second mining. As the results, we obtained 12 and 8 'Interesting' rules in the first and the second mining, respectively.

In this research, we applied the objective measures selected in Section 2.2.3 to the same rules and sorted them in the descending order of their evaluation values. We then regarded the rules from top to 12-th in the first mining and that to 8-th in the second mining as 'Interesting' ones judged by the objective measures.

Note that there are two types of GOI [8], GOI-D (GOI emphasizing Dependency) and GOI-G (GOI emphasizing Generality). GOI is the multiplication of antecedent-consequent dependency and generality factors and possesses a parameter to balance them. Therefore, we used GOI-D in which the weight of the dependency factor was twice that of the generality one and GOI-G with the adverse condition.

3.2 Results and Discussion

The upper and the lower tables in Table 2 show the evaluation results in the first and the second mining, respectively. The caption of Table 2 explains the contents of these tables in detail. The tables describe how the evaluation results of an objective measure matches with that of the medical expert. The white cells in the square on the left side of a table mean the evaluation concordance on 'Interesting'. Similarly, that in the gray-colored columns means that on 'Especially-Interesting'. Therefore, the number of the former and the latter describes the detection performance of an objective measure on 'Interesting' and 'Especially-Interesting', respectively.

To grasp the whole trend of the experimental results, we define the comprehensive criteria to evaluate the objective measure's performance as follows: **#1** Performance on 'Interesting' (the number of 'Interesting' rules judged by an objective measure per that by the medical expert), **#2** Performance on 'Especially-Interesting' (the number of 'Especially-Interesting' rules judged by an objective measure per that by the medical expert), **#3** Count-based performance on all evaluation (the number of rules with the same evaluation results by an objective measure and the medical expert per that of all rules), and **#4** Correlation-based performance on all evaluation (the correlation coefficient between the evalua-

Rule ID	2	3	11	4	5	8	12	13	22	23	24	27	6	17	21	1	7	9	10	14	15	16	18	19	20	25	26	28	29	30	#1	#2	#3	#4
Human Expert	EI	EI	EI	I	I	I	I	I	I	I	I	I	NU	NU	NU	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI				
Support																															5/12	1/3	16/30	0.13
Precision																															6/12	0/3	18/30	0.23
Recall																															8/12	2/3	22/30	0.48
Lift																															6/12	1/3	18/30	0.15
Leverage																															6/12	0/3	18/30	0.21
GOI-D																															5/12	0/3	16/30	-0.03
GOI-G																															5/12	1/3	16/30	0.22
χ^2																															8/12	1/3	22/30	0.38
J-Measure																															4/12	1/3	14/30	0.12

Rule ID	13	21	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	#1	#2	#3	#4
Human Expert	EI	EI	I	I	I	I	I	I	NU	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI				
Support																						2/8	1/2	9/21	-0.24
Precision																						0/8	0/2	5/21	-0.51
Recall																						4/8	2/2	13/21	0.27
Lift																						6/8	0/2	17/21	0.36
Leverage																						5/8	0/2	15/21	0.11
GOI-D																						4/8	0/2	13/21	0.07
GOI-G																						2/8	1/2	9/21	-0.39
χ^2																						6/8	0/2	19/21	0.36
J-Measure																						2/8	1/2	9/21	-0.36

表 2 第 1 回目 (上側), および, 第 2 回目 (下側) のマイニングで得たルール群に対して, 専門家と客観的指標が与えた評価の結果. 各行は, 専門家, 客観的指標がルールに与えた評価結果を意味し, 各列は 1 つ 1 つのルールを意味する. ルールは, 専門家の評価が高いものから降順にソートした.「興味あり」と評価されたルールをボックスで囲い,「特に興味あり」と評価されたルールには灰色を付けた. 専門家の評価において, **EI:** 特に興味あり, bf I: 興味あり, **NU:** 理解不能, **NI:** 興味なし, である. 客観的指標の評価において, □: 専門家と一致, ■: 専門家と不一致, である. 表の右側 4 つの列において, **#1:** 「興味あり」の抽出精度, **#2:** 「特に興味あり」の抽出精度, **#3:** 専門家の評価の推定精度 (一致数), **#4:** 専門家の評価傾向の推定精度 (相関係数), である.

Table 2 Evaluation results by a medical expert and the selected objective measures, for the rules obtained in the first mining (the upper table) and that in the second one (the lower table). Each line means a set of evaluation results of the medical expert or the objective measure, and each column means each rule. The rules are sorted in the descending order of the evaluation values given by the medical expert. The rules judged 'Interesting' by the medical expert are surrounded by a square, and ones judged 'Especially-Interesting' are colored in gray. In the line of medical expert's evaluation, **EI:** 'Especially-Interesting', **I:** 'Interesting', **NU:** 'Not-Understandable', and **NI:** 'Not-Interesting'. In the lines of objective measure's evaluation, **white cell:** "Same as medical expert's evaluation", and **black cell:** "Different from that". In the four columns in the right side, **#1:** Performance on 'Interesting', **#2:** Performance on 'Especially-Interesting', **#3:** Count-based performance on all evaluation, and **#4:** Correlation-based performance on all evaluation.

tion results by an objective measure and that by the medical expert).

At first, we discuss on the results in each mining. As shown in the upper table of Table 2, χ^2 measure 1 and Recall demonstrated the highest performance, and J-Measure, GOI-D, GOI-G, and Support the lowest in the first mining. While, in the lower table of Table 2, χ^2 measure 1 and Lift demonstrated the highest performance, and J-Measure, GOI-D, and Support the lowest in the second mining. Although the objective measures with the highest performance failed to detect some of 'Especially-Interesting' and 'Interesting' rules, their availability to supporting system-human interaction was confirmed at a certain level.

Next, we discuss on the whole trend of the results through the first and the second mining. χ^2 measure 1 maintained the highest performance, and J-Measure, GOI-D, and Support the lowest. Although the performance of Recall and Lift slightly changed, there was no objective measure with dramatic performance change.

We then consider why such trend appeared comparing it with the analysis of medical expert's comments on evaluation. The analysis illustrated the following points of medical expert's observation: **(1)** the medical expert focused on the shape of temporal patterns in a rule rather than the rule performance to predict prognosis, **(2)** he evaluated a rule considering the reliability, the unexpectedness, and the other factors, and **(3)** although the reliability was one of important evaluation factor, many reliable rules were not interesting due to their well-knownness.

The highest performance of χ^2 measure 1 may be caused by **(1)**. Only χ^2 measure 1 uses the instances for the all combination of supporting the antecedent, not supporting the antecedent, supporting the consequent, and not supporting the consequent [23]. Accordingly, it valued the rules in which the temporal patterns in the antecedent and that in the consequent were smoothly connected. This feature of χ^2 measure 1 possibly met the medical expert's needs. While, the lowest performance of Support seems to be deserved by considering **(3)**. The reason of the lowest performance of J-Measure and GOI-D can be estimated base on **(2)**. J-Measure and GOI-D consists of the generality and dependency factors [8], [31], and the balance of these factors was not the same in medical expert's mind in this experiment.

We then summarize the results and the discussions so far: χ^2 measure 1 [23] showed the highest performance, and while J-Measure [31], GOI-D [8], and Support the lowest under this experimental conditions. The objective measures used here possessed not enough but a certain level of performance to detect really interesting rules. The results indicated that the availability of the objective measures for supporting KDD through system-human interaction.

4. Conclusions and Future Work

This research discussed how objective measures can contribute to detect the interesting rules for a medical expert through the experiment using a real chronic hepatitis dataset. The objective measures used here possessed a certain level of detection performance, and then their availability for system-human interaction was indicated. In our future work, we will design the system-human interaction function using objective measures based on this research outcome and equip it to our rule discovery support system. In addition, we will continue the case studies on objective and subjective measures from the viewpoints of not only post-processing but also rule quality management.

文 献

[1] K. J. Cios, and G. W. Moore, "Uniqueness of medical data mining," Artificial Intelligence in Medicine, vol.26, no.1–2, pp.1–24, 2002.

[2] G. Das, L. King-Ip, M. Heikki, et al., "Rule Discovery from Time Series," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD-98), New York, USA, pp.16–22, 1998.

[3] G. Dong and J. Li, "Interestingness of Discovered Association Rules in Terms of Neighborhood-Based Unexpectedness," Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'98), Melbourne, Australia, pp.72–86, 1998.

[4] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, et al. (eds.), "Advances in Knowledge Discovery and Data Mining," AAAI Press/MIT Press, pp.401–419, 1995.

[5] A. A. Freitas, "On Rule Interestingness Measures," Knowledge-Based Systems, vol.12, no.5–6, pp.309–315, 1999.

[6] P. Gago and C. Bento, "A Metric for Selection of the Most Promising Rules," Proc. of European Conference on the Principles of Data Mining and Knowledge Discovery (PKDD'98), Nantes, France, pp.19–27, 1998.

[7] T. Terano and M. Inada, A. Ghosh and S. Tsutsui (eds.), "Advances in Evolutionary Computing," Springer, pp.847–862, 2003.

[8] B. Gray and M. E. Orlowska, "CCAIIA: Clustering Categorical Attributes into Interesting Association Rules," Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'98), Melbourne, Australia, pp.132–143, 1998.

[9] H. J. Hamilton and D. F. Fudger, "Estimating DBLearn's Potential for Knowledge Discovery in Databases," Computational Intelligence, vol.11, no.2, pp280–296, 1995.

[10] H. J. Hamilton, N. Shan, and W. Ziarko, "Machine Learning of Credible Classifications," Proc. of Australian Conf. on Artificial Intelligence (AI'97), Perth, Australia, pp.330–339, 1997.

[11] C. Hausdorf and C. Muller, "A Theory of Interestingness for Knowledge Discovery in Databases Exemplified in Medicine," Int'l Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP'96), Budapest, Hungary, pp.31–36, 1996.

[12] "Hepatitis Dataset for Discovery Challenge," European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'02), Helsinki, Finland, http://lisp.vse.cz/challenge/ecmlpkdd2002/, 2002.

[13] R. J. Hilderman and H. J. Hamilton, Knowledge Discovery and Measure of Interest, Kluwer Academic Publishers, 2001.

[14] O. Hogl, H. Stoyan, and W. Stuhlinger, "On Supporting Medical Quality with Intelligent Data Mining," Proc. of Hawaii Int'l Conf. on System Sciences (HICSS'01), no. HCDAM03, 2001.

[15] S. Jaroszewicz and D. A. Simovici, "A General Measure of Rule Interestingness," Proc. of European Conf. on Principles of Data Mining and Knowledge Discovery (PKDD'01), Freiburg, Germany, pp.253–265, 2001.

[16] M. Kamber and R. Shinghal, "Evaluating the Interestingness of Characteristic Rules," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'96), Portland, Oregon, USA, pp.263–266, 1996.

[17] M. Klementtinen, H. Mannila, P. Ronkainen, et al., "Finding Interesting Rules from Large Sets of Discovered Association Rules," Proc. of Int'l Conf. on Information and Knowledge Management (CIKM'94), Gaitersburg, Maryland, pp.401–407, 1994.

[18] N. Lavrač, "Selected Techniques for Data Mining in Medicine," Artificial Intelligence in Medicine, vol.16, pp.3–23, 1999.

[19] B. Liu, W. Hsu, and S. Chen, "Using General Impressions to Analyze Discovered Classification Rules," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'97), Newport Beach, California, USA, pp.31–36, 1997.

[20] B. Liu, W. Hsu, S. Chen, and Y. Mia, "Analyzing the Subjective Interestingness of Association Rules," Intelligent Systems, vol.15, no.5, pp.47–55, 2000.

[21] B. Liu, W. Hsu, and Y. Mia, "Identifying Non-Actionable Association Rules," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'01), San Francisco, USA, pp.329–334, 2001.

[22] H. Liu, H. Lu, L. Feng, and F. Hussain, "Efficient Search of Reliable Exceptions," Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'99), Beijing, China, pp.194–203, 1999.

[23] Y. Morimoto, T. Fukuda, H. Matsuzawa, et al., "Algorithms for Mining Association Rules for Binary Segmentations of Huge Categorical Databases," Proc. of Int'l Conf. on Very Large Databases (VLDB'98), New York, USA, pp.380–391, 1998.

[24] H. Motoda (eds.), "Active Mining," ISO Press, 2002.

[25] M. Ohsaki, Y. Sato, H. Yokoi, et al., "A Rule Discovery Support System for Sequential Medical Data, – In the Case Study of a Chronic Hepatitis Dataset –," Int'l Workshop on Active Mining (AM-2002) in IEEE Int'l Conf. on Data Mining (ICDM'02), Maebashi, Japan, pp.97–102, 2002.

[26] M. Ohsaki, Y. Sato, S. Kitaguchi, et al., "A Rule Discovery Support System for Sequential Medical Data, – In the Case Study of a Chronic Hepatitis Dataset –," Technical Report of the Institute of Electronics, Information, and Communication Engineers (IEICE), AI2002-81, 2003 (in Japanese).

[27] B. Padmanabhan and A. Tuzhilin, "A Belief-Driven Method for Discovering Unexpected Patterns," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'98), New York, USA, pp.94–100, 1998.

[28] G. Piatetsky-Shapiro, "Discovery, Analysis and Presentation of Strong Rules," Knowledge Discovery in Databases, pp.229–248, AAAI/MIT Press, 1991.

[29] S. Sahara, "On Incorporating Subjective Interestingness Into the Mining Process," Proc. of IEEE Int'l Conf. on Data Mining (ICDM'02), Maebashi, Japan, pp.681–684, 2002.

[30] A. Silberschatz and A. Tuzhilin, "On Subjective Measures of Interestingness in Knowledge Discovery," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'95), Montreal, Canada, pp.275–281, 1995.

[31] P. Smyth and R. M. Goodman, "Rule Indcution using Information Theory," Knowledge Discovery in Databases, pp.159–176, AAAI/MIT Press, 1991.

[32] P. Tan, V. Kumar, and J. Srivastava, "Selecting the Right Interestingness Measure for Association Patterns," Proc. of Int'l Conf. on Knowledge Discovery and Data Mining (KDD'02), Edmonton, Alberta, Canada, pp.32–41, 2002.

[33] N. Zhong, Y. Y. Yao, and M. Ohshima, "Peculiarity Oriented Multi-Database Mining," IEEE Transactions on Knowledge and Data Engineering, vol.15, no.4, pp.952–960, 2003.