International Journal of Hybrid Information Technology
Vol. 5, No. 2, April, 2012

Rough-Set based Criteria for Incremental Rule Induction

Shusaku Tsumoto and Shoji Hirano
Department of Medical Informatics, School of Medicine, Faculty of Medicine
Shimane University
89-1 Enya-cho Izumo 693-8501 Japan
E-mail: {tsumoto,hirano}@med.shimane-u.ac.jp

Abstract

This paper proposes a new framework for incremental learning based on accuracy and coverage. Classified addition of example into four cases, two inequalities for accuracy and coverage are obtained. The proposed method classifies a set of formulae into three layers: rule layer, subrule layer and non-rule layer by using the inequalities obtained. Then, subrule layer plays a central role in updating rules.

1: Introduction

There have been proposed several symbolic inductive learning methods, such as induction of decision trees [1, 2, 7], and AQ family [3, 4, 5]. These methods are applied to discover meaningful knowledge from large databases, and their usefulness is in some aspects ensured. However, most of the approaches induces rules from all the data in databases, and cannot induce incrementally when new samples are derived. Thus, we have to apply rule induction methods again to the databases when such new samples are given, which causes the computational complexity to be expensive even if the complexity is n^2.

Thus, it is important to develop incremental learning systems in order to manage large databases [9, 12]. However, most of the previously introduced learning systems have the following two problems: first, those systems do not outperform ordinary learning systems, such as AQ15 [5], C4.5 [8] and CN2 [3]. Secondly, those incremental learning systems mainly induce deterministic rules. Therefore, it is indispensable to develop incremental learning systems which induce probabilistic rules to solve the above two problems.

Extending concepts of rule induction methods based on rough set theory, we introduce a new approach to knowledge acquistion, which induces probabilistic rules incrementally, called PRIMEROSE-INC2 (Probabilistic Rule Induction Method based on Rough Sets for Incremental Learning Methods).

This method first calculates all the accuracy and coverage values of attributes and induces rules. Then, it classifies attribute into rule layer and subrule layer. Then when an additional example is given, the method is classified into one of the four cases. Then, it updates rule layer and subrule layer and induce rules. The method repeats this process.

The paper is organized as follows: Section 2 makes a brief description about rough set theory and the definition of probabilistic rules based on this theory. Section 4 sec; discusses problems in incremental learning of probabilistic rules and gives some interesting formal results. Then, Section 5 presents an induction algorithm for incremental learning. Finally, Section 6 concludes this paper.

2: Rough Sets and Probabilistic Rules

2.1: Rough Set Theory

Rough set theory clarifies set-theoretic characteristics of the classes over combinatorial patterns of the attributes, which are precisely discussed by Pawlak [6, 13]. This theory can be used to acquire some sets of attributes for classification and can also evaluate how precisely the attributes of database are able to classify data. In the framework of rough set theory, the classification of training samples D can be viewed as a search for the best set $[x]_R$ which is supported by the relation R. In this way, we can define the characteristics of classification in the set-theoretic framework. For example, accuracy and coverage, or true positive rate can be defined as:

$$\alpha_R(D) = \frac{|[x]_R \cap D|}{|[x]_R|} \quad and \tag{1}$$

$$\kappa_R(D) = \frac{|[x]_R \cap D|}{|D|}, \tag{2}$$

where $|A|$ denotes the cardinality of a set A, $\alpha_R(D)$ denotes an accuracy of R as to classification of D, and $\kappa_R(D)$ denotes a coverage, or a true positive rate of R to D, respectively.

For further information on rough set theory, readers could refer to [6, 13].

2.2: Probabilistic Rules

The simplest probabilistic model is that which only uses classification rules which have high accuracy and high coverage. [1] This model is applicable when rules of high accuracy can be derived. Such rules can be defined as:

$$R \stackrel{\alpha,\kappa}{\to} d \quad \text{s.t.} \quad R = \vee_i R_i = \vee \wedge_j [a_j = v_k], \quad \alpha_{R_i}(D) > \delta_\alpha \text{ and } \kappa_{R_i}(D) > \delta_\kappa,$$

where δ_α and δ_κ denote given thresholds for accuracy and coverage, respectively.

It is notable that this rule is a kind of probabilistic proposition with two statistical measures, which is one kind of an extension of Ziarko's variable precision model(VPRS) [13]. [2]

3: Problems in Incremental Rule Induction

The most important problem in incremental learning is that it does not always induce the same rules as those induced by ordinary learning systems [3], although an applied domain is deterministic. Furthermore, since induced results are strongly dependent on the former training samples, the tendency of overfitting is larger than the ordinary learning systems.

[1] In this model, we assume that accuracy is dominant over coverage.

[2] In VPRS model, the two kinds of precision of accuracy is given, and the probabilistic proposition with accuracy and two precision conserves the characteristics of the ordinary proposition. Thus, our model is to introduce the probabilistic proposition not only with accuracy, but also with coverage.

[3] Here, ordinary learning systems denote methods that induce all rules by using all the samples.

The most important factor of this tendency is that the revision of rules is based on the formerly induced rules, which is the best way to suppress the exhaustive use of computational resources. However, when induction of the same rules as ordinary learning methods is required, computational resources will be needed, because all the candidates of rules should be considered.

Thus, for each step, computational space for deletion of candidates and addition of candidates should be needed, which causes the computational speed of incremental learning to be slow. Moreover, in case when probabilistic rules should be induced, the situation becomes much severer, since the candidates for probabilistic rules become much larger than those for deterministic rules.

4: Theory for Incremental Learning

[Question] How accuracy and coverage will change when a new sample is added to the dataset ?

Usually, datasets will monotonically increase. Let $[x]_R(t)$ and $D(t)$ denote a supporting set of a formula R in given data an a target concept d at time t.

$$[x]_R(t+1) = \begin{cases} [x]_R(t)+1 & \textit{an additional example satifies} \\ & R \\ [x]_R(t) & \textit{otherwise} \end{cases}$$

$$D(t+1) = \begin{cases} D(t)+1 & \textit{an additional example belongs} \\ & \textit{to a target concept } d. \\ D(t) & \textit{otherwise} \end{cases}$$

Thus, from the definition of accuracy and coverage(Eqn(1) and (2), accuracy and coverage may nonmonotonically change. Since the above classification gives four additional patterns, we will consider accuracy and coverage for each case as shown in Table 1. in which $|[x]_R(t)|$, $|D(t)|$ and $|[x]_R \cap D(t)|$ are denoted by n_R, n_D andn_{RD}. In summary, Table 2 gives the

Table 1. Four patterns for an additional Example

t:	$[x]_R(t)$	$D(t)$	$[x]_R \cap D(t)$
	n_R	n_D	n_{RD}

t+1	$[x]_R(t+1)$	$D(t+1)$	$[x]_R \cap D(t+1)$
	n_R+1	n_D+1	$n_{RD}+1$
	n_R+1	n_D	n_{RD}
	n_R	n_D+1	n_{RD}
	n_R	n_D	n_{RD}

classification of four cases of an additional example.

Table 2. Summary of Change of Accuracy and Coverage

			$\alpha(t+1)$	$\kappa(t+1)$
n_R	n_D	n_{RD}	$\alpha(t)$	$\kappa(t)$
n_R+1	n_D	n_{RD}	$\frac{\alpha(t)n_R}{n_R+1}$	$\kappa(t)$
n_R	n_D+1	n_{RD}	$\alpha(t)$	$\frac{\kappa(t)n_D}{n_D+1}$
n_R+1	n_D+1	$n_{RD}+1$	$\frac{\alpha(t)n_R+1}{n_R+1}$	$\frac{\kappa(t)n_D+1}{n_D+1}$

4.1: Updates of Accuracy and Coverage

From Table 2, updates of Accuracy and Coverage can be calculated from the original datasets for each possible case. Since rules is defined as a probabilistic proposition with two inequalities, supporting sets should satisfy the following constraints:

$$\alpha(t+1) > \delta_\alpha \quad \kappa(t+1) > \delta\kappa$$

Then, the conditions for updating can be calculated from the original datasets: when accuracy or coverage does not satisfy the constraint, the corresponding formula should be removed from the candidate of rules. On the other hand, both accuracy and coverage satisfy both constraints, the formula should be included into the candidate. Thus, the following inequalities are important:

$$\begin{aligned}
\alpha(t+1) &= \frac{\alpha(t)n_R+1}{n_R+1} > \delta_\alpha, \\
\kappa(t+1) &= \frac{\kappa(t)n_D+1}{n_D+1} > \delta\kappa, \\
\alpha(t+1) &= \frac{\alpha(t)n_R}{n_R+1} < \delta_\alpha, \mathit{and} \\
\kappa(t+1) &= \frac{\kappa(t)n_D}{n_D+1} < \delta\kappa.
\end{aligned}$$

Thus, the following inequalities are obtained for accuracy and coverage.

Theorem 1 *If accuracy and coverage of a formula R to d saisfies one of the following inequalities, then R may exclude or include into the candidates of formulae for probabilistic rules.*

$$\frac{\delta_\alpha(n_R+1)-1}{n_R} < \alpha_R(D)(t+1) < \frac{\delta_\alpha(n_R+1)}{n_R}, \quad (3)$$

$$\frac{\delta_\kappa(n_D+1)-1}{n_D} < \kappa_R(D)(t+1) < \frac{\delta_\kappa(n_D+1)}{n_D}. \quad (4)$$

It is notable that the lower and upper bounds can be calculated from the original datasets.

Select all the formula whose accuracy and coverge satisfy the above inequalities They will be a candidate for updates. A set of formulae which satisfies the inequalities for probabilistic rules is called *rule layer* and one which satisfies Eqn (3) and (4) is called *subrule layer*. Figure 1 illustrates the relations between rule layer and sublayer.

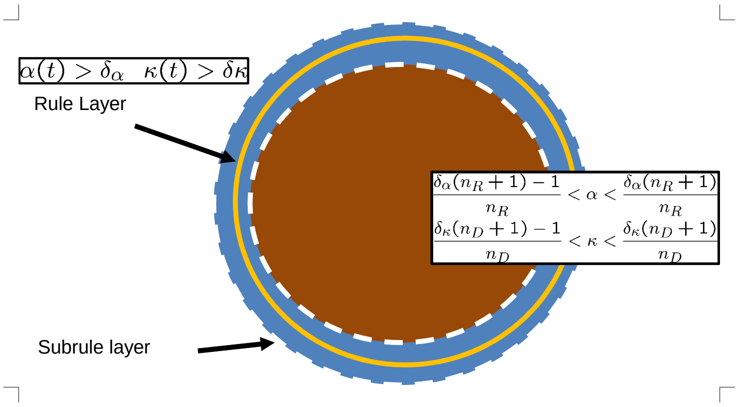

Figure 1. Intuitive Diagram of Rule and Subrule Layers

5: An Algorithm for Incremental Learning

5.1: Algorithm

In order to provide the same classificatory power to incremental learning methods as ordinary learning algorithms, we introduce an incremental learning method PRIMEROSE-INC2 (Probabilistic Rule Induction Method based on Rough Sets for Incremental Learning Methods)[4].

From the results in the above section, a selection algorithm is defined as follows, where the following four lists are used. $List_c$ stores a formula which satisfies the above inequalities shown in Eqn (3) nad (4). This $list_c$ is called *subrule layer*. $List_a$ is a list of formulae probabilistic rules which satisfy the condition on the thresholds of accuracy and coverage, which called *rule layer*. Finally, $List_r$ stores a list of formulae which do not satisfy the above condition.

1. Apply rule induction to the initial data.
2. Store formulae in rules into $List_a$ and others in $List_r$.
3. Calculate upper and lower bounds for accuracy and coverage from accuracy, coverage and given thresholds for rules.
4. Check the above inequalities for attirbutes in $List_a$ and $List_r$. If an attribute satisfies one of the inequalities, then it includes into $List_c$.
5. From an additional example, classify addition from four cases.
   (a) An additional example satisfies a target and a formula in $List_c$. Move a formula from $List_r$ to $List_a$.
   (b) An additional example only satisfies a formula in $List_c$. Move a formula from $List_a$ to $List_r$.

[4]This is an extended version of PRIMEROSE-INC[10]

(c) An additional example only satisfies a target in $List_c$. Move a formula from $List_a$ to $List_r$.

(d) An additional example neither satisfies a target nor a formula in $List_c$. No movement.

6. Generate rules from $List_a$.
7. Return to (3).

6: Conclusion

Extending concepts of rule induction methods based on rough set theory, we have introduced a new approach to knowledge acquistion, which induces probabilistic rules incrementally, called PRIMEROSE-INC2 (Probabilistic Rule Induction Method based on Rough Sets for Incremental Learning Methods).

The method classifies elementary attribute-value pairs into three categories: rule layer, subrule layer and non-rule layer by using the obtained by inequalities obtained from a proposed framework.

Acknowledgements

This research is supported by Grant-in-Aid for Scientific Research (B) 24300058 from Japan Society for the Promotion of Science(JSPS).

References

[1] L. Breiman, J. Freidman, R. Olshen, and C. Stone. *Classification And Regression Trees.* Wadsworth International Group, Belmont , CA, 1984.

[2] Bojan Cestnik, Igor Kononenko, and Ivan Bratko. Assistant 86: A knowledge-elicitation tool for sophisticated users. In *EWSL*, pages 31–45, 1987.

[3] P. Clark and T. Niblett. The cn2 induction algorithm. *Machine Learning*, 3, 1989.

[4] Ryszard S. Michalski. A theory and methodology of inductive learning. *Artif. Intell.*, 20(2):111–161, 1983.

[5] Ryszard S. Michalski, Igor Mozetic, Jiarong Hong, and Nada Lavrac. The multi-purpose incremental learning system aq15 and its testing application to three medical domains. In *AAAI*, pages 1041–1047, 1986.

[6] Z. Pawlak. *Rough Sets.* Kluwer Academic Publishers, Dordrecht, 1991.

[7] J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[8] J.R. Quinlan. *C4.5 - Programs for Machine Learning.* Morgan Kaufmann, Palo Alto, 1993.

[9] Ning Shan and Wojciech Ziarko. Data-based acqusition and incremental modification of classification rules. *Computational Intelligence*, 11:357–370, 1995.

[10] Shusaku Tsumoto. Incremental rule induction based on rough set theory. In Marzena Kryszkiewicz, Henryk Rybinski, Andrzej Skowron, and Zbigniew W. Ras, editors, *ISMIS*, volume 6804 of *Lecture Notes in Computer Science*, pages 70–79. Springer, 2011.

[11] Shusaku Tsumoto and Hiroshi Tanaka. Primerose: Probabilistic rule induction method based on rough sets and resampling methods. *Computational Intelligence*, 11:389–405, 1995.

[12] Paul E. Utgoff. Incremental induction of decision trees. *Machine Learning*, 4:161–186, 1989.

[13] W. Ziarko. Variable precision rough set model. *Journal of Computer and System Sciences*, 46:39–59, 1993.