Automation in Knowledge Engineering Via Inductive Learning

TuBao Ho

*Japan Advanced Institute of Science and Technology – Hokuriku
Tatsunokuchi, Ishikawa 923-12, Japan.*

Phone: +81-761-51-1240. Fax: +81-761-51-1149.

Email: bao@jaist.ac.jp

Abstract

This paper presents an attempt toward the automation in knowledge engineering with the development of two inductive learning systems and their mutual use in a conceptual modeling process: systems R-CABRO and OSHAM for learning concepts from supervised and unsupervised data. We address problems, basic ideas of methods, implementation of systems, their application potential and experimental results.

Keywords

Knowledge engineering, modeling process, supervised and unsupervised learning.

1 INTRODUCTION

Much work in knowledge engineering currently focuses on systems that follow the knowledge acquisition paradigm of *expert or external resources* $\rightarrow$ *automatic acquisition* $\rightarrow$ *knowledge base* instead of the traditional paradigm from expert to a knowledge base via a knowledge engineer. According to Gaines and Shaw (1993), current knowledge engineering research is targeting tools that will provide highly integrated knowledge acquisition environments offering a wide range of complementary tools and techniques. Moreover, a new trend has emerged that knowledge acquisition is not merely an exercise in “expertise transfer” or “knowledge extraction” but a *modeling process* (Ford and Bradshaw, 1993). This paper presents an attempt toward the automation in knowledge engineering by introducing two methods and systems R-CABRO and OSHAM for the automatic construction of knowledge bases. The context in which we carried out this development is the integrated tool TESOR for knowledge-based systems that employs a model of hierarchical object knowledge bases (Ho et al., 1992). Two subsystems R-CABRO and OSHAM, which learn concepts from supervised and unsupervised data respectively, have been developed in order to automate the hierarchical object modeling process. In two next sections we briefly present R-CABRO and OSHAM, and in the last section we present the implementation of the systems, their application potential and experimental results.

2 LEARNING CONCEPTS FROM SUPERVISED DATA

The basic task of acquiring knowledge in this situation is that from a given set of labelled instances to find a classifier that correctly predicts classes of unseen instances. Among approaches to supervised concept learning, we have chosen the decision tree induction which is certainly the most active and applicable one. During the last decade, many decision tree induction methods have been developed, most notably CART (Breiman et al., 1984), ID3 and its optimized version C4.5 (Quinlan, 1993). Decision tree induction systems differ from each other in their ways to deal with two crucial problems of *attribute selection* (choosing the “best” attribute to split a decision node) and *pruning* (avoiding overfitting). To improve the performance of decision tree induction systems is to find suitable solutions of these two problems in concrete situations. Based on a novel attribute selection measure inspired from the rough set theory, decision tree system R-CABRO has been developed and used effectively.

Rough set theory is a new mathematical tool to deal with vagueness and uncertainty (Pawlak, 1991). The basic idea in this theory is to “view” approximately each subset X of an object set O by its *lower* and *upper* approximations w.r.t. an equivalence relation $E \subseteq O \times O$. These approximations of X are defined, respectively, by $E_*(X) \triangleq \{o \in O : [o]_E \subseteq X\}$ and $E^*(X) \triangleq \{o \in O : [o]_E \cap X \neq \emptyset\}$, where $[o]_E$ denotes the equivalence class of an object o in E . A key concept in the rough set theory is the *degree of dependency* $\mu_R(P)$ of an attribute set P on an attribute set R , defined as

$$\mu_R(P) = \frac{\text{card}(\bigcup_{[o]_P} R_*([o]_P))}{\text{card}(O)} = \frac{\text{card}(\{o \in O : [o]_R \subseteq [o]_P\})}{\text{card}(O)} \quad (1)$$

The measure $\mu_R(P)$ relies on one basic assumption: each equivalence class $[o]_R$ in (1) is completely included in some equivalence class $[o]_P$. However this condition is somewhat strict and $\mu_R(P)$ can not measure the dependency of P on R if the partition generated by R do not contain any cluster included in a cluster of the partition generated by P , or when two partitions generated by R and P are nearly identified but their equivalence classes do not satisfy the above assumption. It implies that $\mu_R(P)$ is not robust with noisy data and not enough sensitive in certain cases when R and P are highly mutual dependent (Ho and Nguyen, 1996). We therefore proposed a new measure $\tilde{\mu}_R(P)$, called *R-measure*, for the dependency of P on R , inspired from $\mu_R(P)$

$$\tilde{\mu}_R(P) = \sum_{[o]_R} \frac{1}{\text{card}(O)} \max_{[o]_P} \frac{\text{card}([o]_R \cap [o]_P)^2}{\text{card}([o]_R)} \quad (2)$$

Let's take a simple illustration of the advantage of $\tilde{\mu}_R(P)$ in comparison with $\mu_R(P)$. Consider instances in Table 1, and let $P = \{Flu\}$, $R = \{Headache\}$. Equivalence classes w.r.t. Flu are $\{e_1, e_4, e_5, e_7, e_8\}$, $\{e_2, e_3, e_6\}$ and equivalence classes w.r.t. $Headache$ are $\{e_1, e_2, e_3\}$, $\{e_4, e_5, e_6, e_7, e_8\}$. We can verify that $\mu_{Headache}(Flu) = 0$, and it means that according to $\mu_R(P)$, Flu is totally independent of $Headache$. However, from Table 1 we can see intuitively that the fact $Headache = yes$ probably implies $Flu = yes$ and $Headache = no$ probably implies $Flu = no$. We can verify that $\tilde{\mu}_{Headache}(Flu) = \frac{17}{30}$, and it means that according to $\tilde{\mu}_R(P)$, there is some degree of dependency of Flu on $Headache$.

Table 1 Information table

	<i>Temperature</i>	<i>Headache</i>	<i>Flu</i>
e_1	<i>normal</i>	<i>yes</i>	<i>no</i>
e_2	<i>high</i>	<i>yes</i>	<i>yes</i>
e_3	<i>very_high</i>	<i>yes</i>	<i>yes</i>
e_4	<i>normal</i>	<i>no</i>	<i>no</i>
e_5	<i>high</i>	<i>no</i>	<i>no</i>
e_6	<i>very_high</i>	<i>no</i>	<i>yes</i>
e_7	<i>high</i>	<i>no</i>	<i>no</i>
e_8	<i>very_high</i>	<i>no</i>	<i>no</i>

We have applied $\tilde{\mu}_R(P)$ to the problem of attribute selection in the common framework of decision tree induction shown in Table 2. The “best” attribute chosen at each decision node is the attribute on which the class attribute depends maximally. This selection has been adopted in the construction of the system R-CABRO.

Table 2 Framework of decision tree induction

1. Choose the “best” attribute that maximizes the chosen attribute selection measure
2. Extend tree by adding new branch for each attribute value
3. Sort training examples to leaf nodes
4. If examples unambiguously classified then stop else repeat steps 1–4 for leaf nodes.

Figure 1 Comparison of error rate averages

Our careful experimental studies, by a 10-fold cross validation on 13 data sets from the UCI repository for machine learning, have shown that R-CABRO is always with a high prediction accuracy even with noisy data. We also carried out carefully a comparative study of error rate averages, in terms of attribute selection measures, between R-CABRO (R-measure) and seven decision tree induction systems using well-known attribute se-

lection measures in the machine learning literature, included gain-ratio (GainR) in C4.5, Gini in system CART, Gini', Relief, *J*-measure (*J*-m), d_N distance, and relevance (Relev). Fig. 1 gives results of this comparison and shows that R-CABRO (R-measure) is among the most stable ones with a low error rate (Ho and Nguyen, 1996).

3 LEARNING CONCEPTS FROM UNSUPERVISED DATA

The basic task of acquiring knowledge in this situation is that from a given set of unlabelled instances to find simultaneously a hierarchical clustering that determines useful object subsets and intensional definitions for these subsets of objects. Essentially, unsupervised concept learning methods differ from each other in two factors: *views on concepts* and *constraints of categorization*. Among views on concepts, the classical, prototype and exemplar ones are widely known and used. Among categorization constraints, the similarity, feature correlation, and structure of the concept hierarchy are widely known and used. However each system based on one view on concepts is often limited in capturing the rich variety of conceptual knowledge. The learning system OSHAM, which employs the classical view on concepts is able to form effectively a concept hierarchy from unlabelled data, has been developed (Ho, 1995). Essentially, OSHAM searches to extract a good concept hierarchy by exploiting the structure of Galois lattice of concepts as the hypothesis space. Recently, OSHAM has been extended to a hybrid system that allows obtaining a higher performance by combining its original view on concepts with the prototype and exemplar views, the similarity and correlation constraints (Ho, 1996). The essential idea of the hybrid OSHAM is

Figure 2 Part of the concept hierarchy from Wisconsin breast cancer data

1. to extend the classical representation of a concept from a pair of extent, intent and its hierarchy structure information to a hybrid representation of a 10-tuple with additional components. These additional components include the probability of the concept occurrence, its homogeneity, typical instances, conditional probability for occurrence and dispersion of typical instances, the quality of concept evaluated in terms of the concept hierarchy structure.
2. to find sufficiently general and discriminant concepts at each level of the concept hierarchy while determining their typical instances based on a quality criterion that combines the similarity and structural constraints.

The following benchmark has been used for evaluating the performance of OSHAM: employing well known supervised data sets but hide the class name in forming and matching the concept hierarchy with unknown instances, and after these phases we use the class name only to estimate the predictive accuracy of the system.

Fig. 2 presents the first four levels of the concept hierarchy obtained by OSHAM from the Wisconsin breast cancer data. This data set consists of 699 observations of two classes of Benign and Malignant cases on 9 symbolic symptoms: Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses, each has 10 possible values. In Fig. 2 these attributes and their values are denoted by 1, 2, ..., 10 and each attribute-value pair is written between two parentheses.

Each concept is presented in Fig. 2 by its generated order, its non-inherited features (its inherited features are agglomerated along the branch leading to it from the root), and other five associated components in an attached small rectangle. Two first lines p and d are found from unsupervised data and indicate the probability of occurrence and the dispersion between the concept members. Next three lines, obtained by using the class name, are information about the predicted name (Benign or Malignant), the rate of the number of instances with the predicted name over the total number of instances, and the decimal value of this rate (considered as the homogeneity or estimated accuracy of the concept). Information about concepts, for example concept C_{24} , are fully represented as follows

Figure 3: Comparison with C4.5 and CART

Figure 4: Comparison with COBWEB

```

CONCEPT 24
Level = 3
Super_Concepts = { 11 }
Sub_Concepts = { 29 }
Features = (Single Epithelial Cell Size, 2) ^ (Mitoses, 1) ^ (Normal Nucleoli, 1)
Unclassified/Total instances = 19/49
Unclassified_instances = {53 122 141 185 230 238 ... 507 556 651 658 659}
Concept probability = 0.070100
Unclassified_instance_conditional_probability = 0.387755
Concept_dispersion = 0.412415
Unclassified_instance_dispersion = 0.534763
Subconcept_partition_quality = 7.164591
Concept_name = Benign (44/49, name_homogeneity = 0.897959)
    
```

Fig. 3 presents a comparison of prediction accuracy between OSHAM and the supervised methods C4.5 and CART. Fig. 4 presents a comparison of prediction accuracy between OSHAM and the unsupervised method COBWEB (Fisher, 1987). The latter is used only after the above mentioned steps have been done in order to estimate the predictive accuracy of OSHAM. OSHAM attained a prediction accuracy which is somehow lightly lower than those of C4.5 and CART and is lightly higher than those of COBWEB in these experiments.

Figure 5 Screens of interactive OSHAM

OSHAM is currently elaborated in the following directions

- an *interpreter* that allows the system to predict flexibly the membership of unknown instances by combining logical, threshold, and competitive matching approaches.
- an *incremental* INCOSHAM that allows the system to update effectively the learned knowledge when the environment changes incrementally.
- an *interactive* OSHAM in the X Window that allows the user to collaborate with the system in constructing knowledge bases (Fig. 5).
- an *approximation* model that allows OSHAM to deal better with uncertain and incomplete data. This model is based on an attempt of integrating rough set theory and formal concept analysis (referred to as rough concept analysis).

4 CONSTRUCTION OF KNOWLEDGE BASES

R-CABRO and OSHAM have been implemented in C under the X Window and used to support the construction of hierarchical object knowledge bases in knowledge-based system generator TESOR (Ho et al., 1992). The scheme for integrating R-CABRO, OSHAM and TESOR Graphic Knowledge Editor (GKE) in this knowledge acquisition modeling process is shown in Fig. 6.

Figure 6 Knowledge acquisition process in TESOR

- The expert/knowledge engineer describe a domain knowledge by the GKE if a set of rules and/or an object hierarchy is available (I).
- R-CABRO generates rules from object-attribute data sets with class information (supervised task). The output of R-CABRO is a file containing a decision tree (or rules transformed from the decision trees in the IF-THEN form of the TESOR knowledge representation language syntax). This file may be read and compiled into the internal representation form of TESOR with the help of the GKE (II). R-CABRO can effectively handle large data sets.
- From a data set without class information, OSHAM generates a concept hierarchy in terms of its quality criteria (unsupervised task). The output of OSHAM is compiled directly by TESOR GKE into the internal representation form. The object hierarchy is displayed graphically for supporting the expert to give his judgement until obtaining a satisfied hierarchical object knowledge base (III).
- The visualization of knowledge bases by the TESOR GKE offers the expert/knowledge engineer an intuitive understanding about the structure of the domain knowledge. It allows the expert/knowledge engineer, in the modeling process, to observe the results and the quality evaluations, modify parameters and adjust results at each level before the system continues to go further (IV).

Langley and Simon (1995) stated that “machine learning may never entirely replace knowledge engineering as a framework for constructing knowledge-based systems, but our examples show that significant progress toward automation has been made”. In this sense, the high performance of R-CABRO and OSHAM for both supervised and unsupervised data shows that they have a high application potential. They can be used for constructing knowledge bases with low-cost and high-quality, particularly in domains that requires one-step classification and prediction tasks, or with some forms of multi-step inference or problem solving.

REFERENCES

- Breiman, L., Friedman, J., Olshen, R., Stone, C., *Classification and regression trees*. Belmont, CA: Wadsworth, 1984.
- Fisher, D. (1987) Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2, 139–172.
- Ford, K.M., Bradshaw, J.M. (1993) *Knowledge Acquisition as Modeling*. Wiley.
- Gaines, B.R., Shaw, M.L. (1993) Eliciting Knowledge and Transferring It Effectively to a Knowledge-Based System. *IEEE Trans. on Knowledge and Data Engineering*, 5, 4–14.
- Ho, T.B., Pham, N.K., Bach, H.K., Hoang, T.M., Ngo, C.S., Nguyen, T.D., Tong, T.H., Hoang, Q.T. (1992) Development and Applications of the Expert System Generator TESOR. *Proceedings of NCSR of Vietnam*, 2, 3–14.
- Ho, T.B. (1995) An Approach to Concept Formation Based on Formal Concept Analysis. *IEICE Trans. Information and Systems*, E78-D, 553–559.
- Ho, T.B. (1996) A Hybrid Model for Concept Formation”, in *Information Modeling and Knowledge Bases VII*. IOS Publisher (in print).
- Ho, T.B., Nguyen, T.D. (1996) Evaluation of Attribute Selection Measures in Decision Tree Induction. *9th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*, Fukuoka (in print).
- Langley, P., Simon, H.A. (1995) Applications of Machine Learning and Rule Induction. *Communications of the ACM*, 38, 55–64.
- Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Pawlak, Z. (1991) *Rough Sets – Theoretical Aspects of Reasoning About Data*. Kluwer.

BIOGRAPHY

TuBao Ho is currently a visiting associate professor at the School of Information Science, Japan Advanced Institute of Science and Technology–Hokuriku. He received a B.Tech. degree in Computer Science and Applied Mathematics from Hanoi Institute of Technology (1978), M.S. and Ph.D. degrees from Pierre and Marie Curie University, Paris (1984, 1987). In 1979, he joined the Institute of Computer Science and Control, National Centre for Scientific Research of Vietnam where he is now a senior researcher and associate professor. His research interests include machine learning, knowledge-based systems, knowledge and data engineering. His is a member of IFIP Working Group 9.4.