Toward A Multi-Strategy and Cooperative Discovery System

Ning Zhong
RCAST*
The University of Tokyo

Setsuo Ohsuga
Dept. of Infor. & Comp. Sci.
The Waseda University

Abstract

We have been developing a methodology/system called GLS (Global Learning Scheme) for knowledge discovery in databases. The development of GLS has two main aspects. The first is to develop a *multi-strategy* system. That is, many kinds of discovery/learning methods are cooperatively used in multiple learning phases for performing multi-aspect intelligent data analysis as well as multi-level conceptual abstraction and learning. As a multi-strategy system, GLS is implemented as a toolkit composed of several sub-systems and optional parts with a multi-level structure. We have finished main parts belonging to this aspect, and have undertaken another aspect, i.e., extending GLS into a *multi-agent, distributed* and *cooperative* discovery system. We try to increase versatility and autonomy of GLS by multi-strategy and distributed cooperation. This paper briefly discusses these two aspects of GLS.

Introduction

Increasing *versatility* and *autonomy* is one goal of research on KDD (Knowledge Discovery in Databases) (Matheus et al. 1993). Versatility can be increased by developing a *multi-strategy* system. Based on this, autonomy can be implemented by organizing dynamically the process including discovering, managing and refining knowledge hidden in databases in either the centralized or distributed cooperative mode, according to different discovery tasks and/or user requirements. We have been developing a methodology/system called GLS (Global Learning Scheme) for KDD (Zhong & Ohsuga 1992, 1994b). The development of GLS has two main aspects. The first is to develop a *multi-strategy* system. Since databases have the following features different from other learning objects such as (1) databases are not always complete but contain uncertain and incomplete data; (2) there are different kinds of data such as numerical data and symbolic data in databases; (3) databases are generally very large and complex; (4) databases for discovering knowledge are not always static but dynamic, we cannot wish a single discovery/learning algorithm for solving all problems. To meet the features of databases, we adopt the process of discovering knowledge from databases based on incipient hypothesis generation/evaluation and refinement/management in GLS as shown in Figure 1. In this process, many kinds of discovery/learning methods are cooperatively used in multiple learning phases for performing multi-aspect intelligent data analysis as well as multi-level conceptual abstraction and learning.

We have finished main parts belonging to the *multi-strategy* aspect, and have undertaken another aspect, i.e., extending GLS into a *multi-agent, distributed* and *cooperative* discovery system. We try to increase the autonomy of the discovery process by increasing the number of discovery steps in succession performed in both the centralized and distributed cooperative mode. GLS is implemented by KAUS. KAUS is a knowledge-based system developed in our laboratory which involves knowledge-bases based on Multi-Layer Logic and databases based on the Non Normal Form model (Ohsuga & Yamauchi 1985). Thanks to the useful capabilities such as meta reasoning, multiple knowledge worlds/levels and the model representation in KAUS, and the recent development of distributed KAUS (Ohsuga 1990, Suzuki et al. 1994), GLS can be easily implemented and extended by KAUS.

This paper briefly discusses these two aspects of GLS. It includes to outline several functions of GLS as a multi-strategy system, to discuss how to extend the multi-strategy system into a multi-agent, distributed and cooperative discovery system, and describe briefly an experimental application and future work.

GLS as a Multi-Strategy System

As a multi-strategy system, GLS is implemented as a toolkit composed of several sub-systems and optional parts with a multi-level structure. As shown in Figure 2, at present, two sub-systems of GLS, KOSI (Knowledge Oriented Statistic Inference) and DBI (Decomposition Based Induction), have been developed for discovering incipient hypotheses from databases, and two further sub-systems of GLS, IIBR (Inheritance

*Research Center for Advanced Science and Technology, E-mail: zhong@ai.rcast.u-tokyo.ac.jp

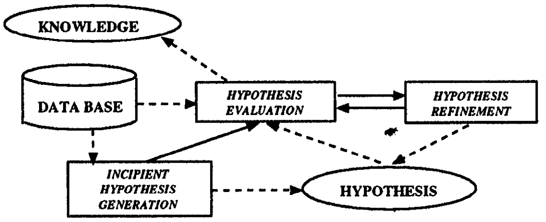

Figure 1: The process of knowledge discovery from databases

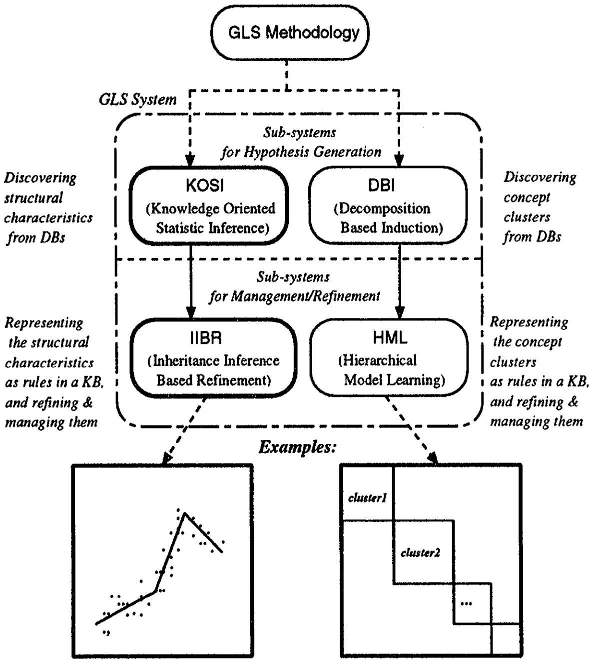

Figure 2: An overview of the GLS system

Inference Based Refinement) and HML (Hierarchical Model Learning), have also been developed for managing and refining the discovered incipient hypotheses. Since to describe each of sub-systems of GLS in detail requires much more space than we have in this paper, we only outline main functions of KOSI and IIBR as a preparation for discussing how to use them in a distributed cooperative mode. For details refer to (Zhong & Ohsuga 1993, 1994a, 1995a, 1995b).

KOSI

KOSI (Knowledge Oriented Statistic Inference) pays attention to the functional relations between any two attributes for discovering *structural characteristics* from databases (Zhong & Ohsuga 1995b). *Structural characteristics* are a kind of important regularity hidden in databases, which are denoted by regression models for describing three kinds of functional relations: the *exact*, *strong* and *weak* ones according to which method in KOSI was successfully used for discovering them and their errors. In some sense, KOSI can be regarded as an extension of BACON and its several successors for processing the data with more uncertainty (Langley & Zytkow 1989). The key point of this extension is to enhance the capability of processing uncertainty systematically by extending the heuristic search and the search control, as well as combining them with some statistical methods.

In comparison with the related systems, the most novel features of KOSI are such that

- It provides a systematic manner of discovering functional relations, and supports qualitative/quantitative discovery by using three kinds of search which are called *heuristics-type-1*, *heuristics-type-2* and *search/evaluation based on regression analysis* respectively. It also uses a model-base and many kinds of meta/domain knowledge for controlling the multi-search.

  Where, *heuristics-type-1* is mainly used for finding the *exact* functional relation which almost always holds for the collected data; *heuristics-type-2* is mainly used for finding the *strong* functional relation which holds qualitatively for the collected data. Furthermore, *search/evaluation based on regression analysis* is mainly used for further evaluating/selecting the best functional relations from the results of the two types of heuristic search, and/or finding the *weak* functional relation which presents the structure hidden in the collected data. Finally, the selected functional relations are denoted by regression models as the structural characteristics discovered, so that they can be easily managed and refined by IIBR.
- It uses several novel statistical methods. For example, AIC (Akaike Information Criterion) is used for selecting the optimal model with both better stability and smaller variance (Akaike 1974); regression analysis using the Least Square Method based on the Householder transformation is used for developing an effective, practical system; the *stepwise Chow test (SCT)*algorithm is used for clustering time-series data.
- It provides more room to the user for selectivity. That is, several methods such as the heuristic search, forming scopes and clusters, the evaluation criteria etc., can be selected by the user according to different requirements. Furthermore, the selectivity is also a basis for organizing dynamically several discovery steps as a process performed in succession in either the centralized or distributed cooperative mode.

IIBR

IIBR (Inheritance Inference Based Refinement) is another sub-system of GLS that is closely related to

KOSI. By means of IIBR, the structural characteristics denoted by regression models, which are discovered from a database by KOSI, can be represented by Multi-Layer Logic formulae and the sets of data for showing their errors in a knowledge-base, and can be managed and refined easily (Zhong & Ohsuga 1995a).

IIBR is based on inheritance inference and error analysis, as well as meta reasoning, multiple worlds/levels of KAUS and the capability of expansion of Multi-Layer Logic (Ohsuga 1990, Ohsuga & Yamauchi 1985). IIBR has been strongly influenced by Rowe's work (Rowe 1991). By means of the concept of inheritance inference on regression models developed by Rowe, IIBR is used to find matches to models for similar situations to those under study, to give a starting model for analysis. A good starting model can save a user much time, and effective inference can also save storage space by eliminating the need to save similar models. Main functions of IIBR are

- The method of the model representation is used to represent the discovered structural characteristics (regression models) in a knowledge-base for management and refinement easily;
- Inheritance relationship among regression models can be evaluated quantitatively, and meta reasoning is cooperatively used with the quantitative evaluation in refinement for acquiring the best regression models;
- The families of regression models are managed by the rule chains and the inheritance graphs of regression models;
- A suitable regression model can be selected from a family of regression models for use. And the discrimination models can be generated dynamically to select the suitable regression model when a database is decomposed into several clusters.

Optional Parts

One of the novel features of GLS is to provide more room for selectivity. That is, according to different discovery tasks, several discovery/learning methods can be selected by the user, and/or GLS itself organizes dynamically several discovery steps as a process performed in succession in either the centralized or distributed cooperative mode. Here, we briefly describe several optional methods for forming scopes/clusters, which are mainly used as a step of pre-processing in KOSI (Zhong & Ohsuga 1995b), as examples.

When a database is very large and complex, the possibility of which the functional relations are discovered from all data in this database is very little and it is also time-consuming. But if some attribute(s), in which the data were divided into scopes, are used as conditions or criteria that the functional relations should meet, then the functional relations with conditions may be discovered as fast as possible; Or a database is decomposed into several clusters, then the methods for searching functional relations can be respectively performed for every cluster in parallel. However, we cannot wish that a method for forming scopes/clusters is good for all applications since the complexity of databases and the diversification of discovery tasks. Hence, GLS provides the following methods for selectivity:

- *attribute oriented clustering using background knowledge (CBK).* Attribute oriented clustering is a kind of operation for abstracting the data in an attribute. A result of the operation is that scopes are formed. Furthermore, the formed scopes can be used as the qualitative values for further clustering other attributes (Zhong & Ohsuga 1994a, 1995b). CBK is a method of the clustering by using background knowledge. That is, background knowledge is used for conceptual abstraction (generalization) and/or the quantization of continuous values.
- *quantization by the division of ranges (QDR).* Unlike CBK, QDR is an automated method for clustering numeric attribute, in which scopes/clusters are formed by an algorithm based on a criterion of classification (Zhong & Ohsuga 1994a).
- *forming scopes/clusters by nominal or symbolic attributes (FSN).* That is, nominal or symbolic attributes are used for forming scopes/clusters. In comparison with the related methods, the most novel feature of FSN is that it can decide, by automatic search and statistics, which nominal or symbolic attributes can be used for forming scopes/clusters (Zhong & Ohsuga 1995b).
- *stepwise Chow test (SCT).* SCT is used for clustering time-series data. That is, scopes/clusters are formed by an algorithm based on a criterion called Chow-Test that was introduced by Chow to distinguish whether a structure change occurred in sample data or not (Chow 1960). SCT can discover automatically the structure changes in time-series data, cluster time-series data by discovering structure changes, and analyze/delete automatically unstable data in the area of continuous structure changes (Zhong & Ohsuga 1995b).

Distribution and Cooperation

There are two main reasons of which we need to develop a multi-agent, distributed and cooperative discovery system. The first is that when databases as discovering objects are very large and complex, we need to decompose a large database into several clusters by using clustering techniques and to process these clusters in parallel for more rational use of computer resources. Another is that since a multi-strategy discovery system as stated above has been implemented preliminarily, now we further consider how to use cooperatively these sub-systems and optional parts of GLS for more high-level, complex discovery tasks. In other words, in GLS, learning is not only to discover the knowledge according to the user requirement (we call *object-level* *learning*), but also to organize dynamically the discovery processes according to different discovery tasks and improve the performance of GLS itself (we call *meta-level learning*).

Thanks to the multiple meta-level structure of KAUS (Ohsuga 1990) and the recent development of distributed KAUS (Suzuki et al. 1994), GLS as a multi-strategy system can be easily extended into a multi-agent, distributed and cooperative discovery system. First, the agents in our sense are not static but dynamic, i.e., they are dynamically composed of Intelligent Mail Box and one or partial functions of a sub-system, or some sub-systems of GLS stated above. And their IDs are composed of names of workstation and agent. An agent can try to solve a discovery task or a sub-task, or something that serves this task but necessary to another agent's solving of the task. It can interact with other agents to help solve the task. Second, the communication protocol uses the one of distributed KAUS that is an extended and revised version of Smith's one (Suzuki et al. 1994). Third, a multiple meta-level structure is used for solving some meta-level problems and controlling the meta or object level. The meta-level problems include mainly decomposing discovery task, allocating resource, adaptive self-configuration of discovering steps, managing interaction/communication among agents, synthesizing part-results of discovery and so on. These meta-level problems are solved by a meta-meta level, i.e., the roles of the meta-meta level are to organize dynamically a discovery process and control this process as the coordinator.

Figure 3 shows briefly a sample discovery process by using both multi-strategy and distributed cooperation. Its task is to discover *structural characteristics* from databases. As the aspect of multi-strategy, it uses main functions of KOSI and IIBR; As the aspect of distributed cooperation, it shows how use some optional parts related to KOSI and IIBR in a distributed cooperative mode. In Figure 3, the parts shown in the parallel lines are main parts of parallel distributed processing. A method of organizing the discovering process is roughly described as follows:

According to the user requirement, GLS collects the useful data from a or more database(s) into the working memory, and organizes/controls the discovery processes in the meta-meta level;

If a user wants to investigate whether there are some structural characteristics among some attributes, then start KOSI;

Try to form scopes and/or clusters by using CBK, QDR, FSN, and SCT in parallel;

Execute the heuristic search by using *heuristics-type-1* and *heuristics-type-2* for every cluster in parallel;

Evaluate the results of the heuristic search by using regression analysis for every cluster in parallel;

Start IIBR, represent the discovered structural characteristics in a knowledge-base by Multi-Layer Logic formulae and the sets of data for showing their errors;

Manage/refine the discovered structural characteristics along with data change.

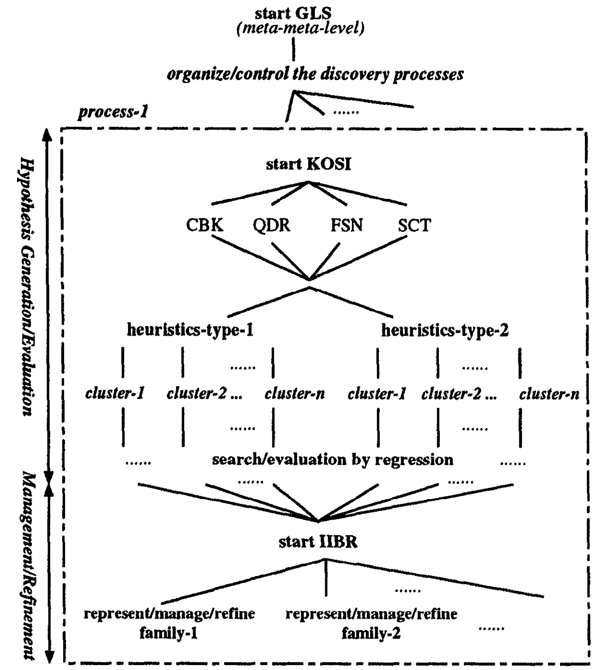

Figure 3: A sample distributed cooperative discovery process

An Experimental Application

We have tested our first experimental application which involves to organize a process for discovering, managing and refining *structural characteristics* hidden in a database called *stars* by cooperatively using KOSI and IIBR in a distributed cooperative mode. This database consists of several attributes such as the cluster designation of stars C_i, the V magnitude, the color indexes B-V and U-B, effective temperature, luminosity, mass of the stars etc. Furthermore, in order to describe data change in a database, this *stars* database is divided into two groups: *group* 1 for fundamental data and *group* 2 for its variation.

As a general knowledge in space science, we know that in comparison to field stars, members of an open cluster are more suitable objects for such calculations because of the homogeneous chemical composition in the clusters and the reliability of luminosity and temperatures determined from the cluster UBV data (Piskunov 1980). Hence, we assume that (1) a user wants to investigate if there are some structural

Table 1: The discovered structural characteristics from the DB *stars*

C_i	Polynomial regression models	σ^2	AIC
1	$Y = 4.83 + 0.08X_{lum} + \varepsilon$	0.018	-140.3
3	$Y = 5.072 - 0.183X_{lum} + 0.044X_{lum}^2 + \varepsilon$	0.013	-170.8
...	......	...	...
1	$Y = 4.987 + 0.337X_{b-v} - 0.2X_{b-v}^2 + \varepsilon$	0.014	-173.2
2	$Y = 4.673 + 0.677X_{b-v} - 0.23X_{b-v}^2 + \varepsilon$	0.029	-105.9
...	......	...	...
1	$Y = 5.099 + 0.061X_{u-b} - 0.094X_{u-b}^2 + \varepsilon$	0.017	-151.7
2	$Y = 5.0778 + 0.144X_{u-b} - 0.092X_{u-b}^2 + \varepsilon$	0.033	-81.1
...	......	...	...

characteristics between the attribute *effective temperature* and other some attributes such as *luminosity*, the color indexes *B-V* and *U-B* in the clusters; (2) the user provided the background knowledge for defining the qualitative values of *effective temperature* by CBK:

$$[effTemp]_1 = [5.0 \sim 4.5]$$
$$[effTemp]_2 = (4.5 \sim 4.0]$$
$$[effTemp]_3 = (4.0 \sim 3.5]$$
$$[effTemp]_4 = (3.5 \sim 3.0].$$

This discovery process is divided into two main stages. The **first stage** is to find structural characteristics from this *stars* database by KOSI. Firstly, according to the user requirement, collect the useful data from this *stars* database, and then try to form scopes and/or clusters by using CBK, QDR, FSN and SCT in parallel. Since there is a nominal attribute called *clusters* that can be used for forming clusters, this *stars* database is divided into several clusters by FSN. Moreover, in CBK, as a preparation for performing *heuristics-type-2*, the qualitative values of *effective temperature* stated above are used for further clustering other attributes into several sets. Secondly, try *heuristics-type-1* and *heuristics-type-2* for every cluster in parallel. Since *heuristics-type-2* is executed successfully, it is hypothesized that there are the *strong* functional relations as follows:

effTemp $\dot{\propto}_Q$ *luminosity*, and *effTemp*, *luminosity* $\in C_1, C_3, C_4, C_5, C_7, C_8, C_{10}, C_{11}$ and C_{12};

effTemp $\dot{\propto}_Q$ *B-V*, and *effTemp*, *B-V* $\in C_1, C_2, C_3, C_4, C_5, C_7$, and C_{12};

effTemp $\dot{\propto}_Q$ *U-B*, and *effTemp*, *U-B* $\in C_1, C_2, C_3, C_4, C_5, C_7, C_8, C_{10}, C_{11}$ and C_{12}.

Third, evaluate these hypothesized *strong* functional relations by using regression analysis in parallel, and select the best ones as structural characteristics discovered. Table 1 shows a part of the results. In Table 1, Y is *effective temperature*, X_{lum} is *luminosity*, X_{b-v} is *B-V* and X_{u-b} is *U-B*.

Based on the results stated above, the **second stage** is to represent the discovered structural characteristics by Multi-Layer Logic formulae and the sets of data for showing their errors in a knowledge-base, as well as to manage and refine them by IIBR. The following *Rule-1* is an example of the representation:

```
Rule-1: /* the rule for inferring the effective temperature of stars
from the luminosity of stars */
!ins_e clusters 1, 3; /* use clusters 1, 3 */
!ins_e variance 0.018, 0.013;
   /* the variance of the reg-models belonging to a family */
!ins_e ai-pr-1-0 4.83, 5.072; /* the coefficient A0 */
!ins_e ai-pr-1-1 0.08, -0.183; /* the coefficient A1 */
!ins_e ai-pr-1-2 0, 0.044; /* the coefficient A2 */
[∀ X-luminosity,Y-effTemp/float] [∀ Mode,Check-N/int]
[∀ A0#/ai-pr-1-0] [∀ A1#/ai-pr-1-1] [∀ A2#/ai-pr-1-2]
   /* declare the domains of variables */
( | (p-stars Mode Check-N Y-effTemp X-luminosity)
   /* infer the effective temperature from the luminosity */
   ~($pr 2 Y-effTemp A0 X-luminosity A1 A2)
   /* infer the eff-temp by the polynomial regression model */
   ~($scope_kb rule-set3) /* transfer to the world: rule-set3 */
   ~(storeInfor Mode Check-N pr Y-effTemp X-luminosity)
   /* store the inferred result and the variable */
).
```

Managing/refining the regression models are important issues when several regression models were generated along with data change (e.g., *group* 2 of data is added to this *stars* database). In particular, the contents of most databases are ever changing; and erroneous data can be a significant problem in real-world databases. Hence, the process of discovering knowledge from databases is a process based on incipient hypothesis generation/evaluation and refinement/management as shown in Figure 1. As stated above, in IIBR, the rule chains and the inheritance graphs corresponding to the families of regression models are used for management. By means of them, the following jobs can be done:

- Regression models discovered from databases are first stored in the rule chains, and then are refined by evaluating quantitatively the inheritance relationship among regression models;
- The time and history of regression models are represented and managed. That is, the rule chains of storing regression models are dynamically generated as time goes on for recording the evolution process of regression models;
- The inheritance graphs of regression models are dynamically generated for describing the relationship among regression models.

The rule chains and the inheritance graphs of regression models are defined by the Multi-Layer Logic formulae and the set-elements relations, and are managed by a meta knowledge level as shown in Figure 4. Figure 4 also shows the structure of the inheritance graphs and some operations for them in IIBR.

Discussion

In comparison, GLS is mostly similar to INLEN in related systems (Michalski 1992). In INLEN, a database, a knowledge-base and several existing methods of machine learning are integrated as several operators. These operators can generate diverse kinds of knowledge about the properties and regularities existing in

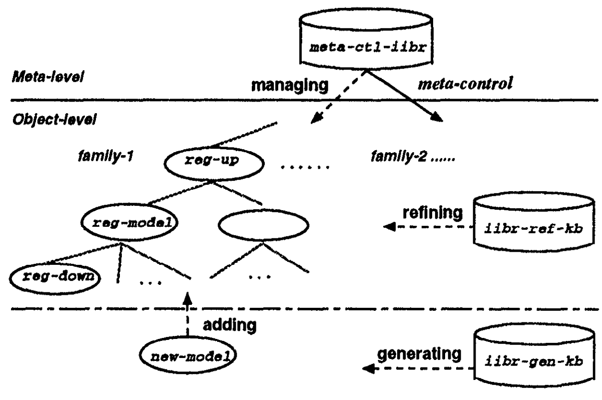

Figure 4: An inheritance graph for a regression models family and operations for it in IIBR

the data. INLEN was implemented as a toolkit like GLS. However, GLS can organize dynamically the discovery processes performed in either the centralized or distributed cooperative mode. Moreover, the refinement for knowledge is one of important capabilities of GLS that was not developed in INLEN.

Since the GLS system to be finished by us is very large and complex, we have only finished main parts belonging to the *multi-strategy* aspect and have undertaken to extend it into a *multi-agent, distributed* and *cooperative* discovery system. That is, the work that we are doing takes but one step toward a multistrategy and cooperative discovery system. Other future work mainly involves to perfect current system, to develop an intelligent user interface, to combine autonomous discovery with interactive discovery, to support the process from discovery to invention, and apply our system to more real problems and application fields for further testing and demonstration.

Acknowledgements

The authors would like to thank Dr. Maria Zemankova and Prof. Jan Zytkow for their valuable comments in the ISMIS'94 conference. Further we would like to thank the referees of the paper for their valuable comments.

References

Akaike, H. 1974. A New Look at The Statistical Model Identification. *IEEE Trans. Autom. Contr.*, AC-19.

Chow, G.C. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3).

Langley, P. & Zytkow, J.M. 1989. Data-Driven Approaches to Empirical Discovery. *Artificial Intelligence*, 40(1-3):283-312.

Matheus, C.J., Chan, P.K. & Piatetsky-Shapiro, G. 1993. Systems for Knowledge Discovery in Databases. *IEEE Trans. Knowl. Data Eng.*, 5(6):904-913.

Michalski, R.S. et al. 1992. Mining for Knowledge in Databases: The INLEN Architecture, Initial Implementation and First Results. *J. of Intell. Infor. Sys.*, KAP, 1(1):85-113.

Michalski, R.S. et al. 1983/86/90/94. *Machine Learning - An Artificial Intelligence Approach*, Vols.1,2,3,4.

Ohsuga, S. & Yamauchi, H. 1985. Multi-Layer Logic - A Predicate Logic Including Data Structure as Knowledge Representation Language. *New Generation Computing*, 3(4):403-439.

Ohsuga, S. 1990. Framework of Knowledge Based Systems. *Knowl. Based Sys.*, 3(4):204-214.

Ohsuga, S. 1993. How can Knowledge Based Systems Solve Large Scale Problems?: Model Based Decomposition and Problem Solving. *Knowledge Based Systems*, 6(1):38-62.

Piatetsky-Shapiro, G. & Frawley, W.J. (eds.). 1991. *Knowledge Discovery in Databases.* AAAI/MIT Press.

Piatetsky-Shapiro, G. & Matheus, C.J. 1992. Knowledge Discovery Workbench for Exploring Business Databases. *Int. J. of Intell. Sys.*, 7(7):675-686.

Piskunov A.E. 1980. Bull. Inf. CDS, 19, 67.

Rowe, N.C. 1991. Managemnet of Regression-model Data. *Data & Knowl. Eng.*, North-Holland, 6(4):349-363.

Shavlik, J.W. & Dietterich, T.G. (eds.). 1990. *Readings in Machine Learning.* Morgan Kaufmann Publishers.

Suzuki, E., Hori, K., Ohsuga, S. & Morizet-Mahoudeaux, P. 1994. Problem Solving by Negotiation among Autonomous Knowledge Processing Systems. *J. of Japanese Society for Artif. Intell.*, 9(1):109-118.

Vittal, J. et al. 1992. Intelligent and Cooperative Information Systems Meet Machine Learning. *Int. J. of Intelligent and Cooperative Information Systems*, 1(2):347-361.

Zhong, N. & Ohsuga, S. 1992. GLS - A Methodology for Discovering Knowledge from Databases. *Proc. 13th Int. CODATA Conf. entitled "New Data Challenges in Our Information Age"*, A20-A30.

Zhong, N. & Ohsuga, S. 1993. HML - An Approach for Managing/Refining Knowledge Discovered from Databases. *Proc. 5th IEEE Int. Conf. on Tools with Artif. Intell. (TAI'93)*, 418-426.

Zhong, N. & Ohsuga, S. 1994a. Discovering Concept Clusters by Decomposing Databases. *Data & Knowl. Eng.*, North-Holland, 12(2):223-244.

Zhong, N. & Ohsuga, S. 1994b. The GLS Discovery System: Its Goal, Architecture and Current Results. *Proc. 8th Inter. Symp. on Methodologies for Intell. Sys. (ISMIS'94).* LNAI 869, Springer, 233-244.

Zhong, N. & Ohsuga, S. 1995a. Managing/Refining Structural Characteristics Discovered from Databases. *Proc. 28th Hawaii Int. Conf. on Sys. Sciences (HICSS-28)*, Vol.3:283-292.

Zhong, N. & Ohsuga, S. 1995b. KOSI - An Integrated System for Discovering Functional Relations from Databases. (to appear in) *J. of Intell. Infor. Sys.*, KAP, 5(1).

Zytkow, J.M. & Zembowicz, R. 1993. Database Exploration in Search of Regularities. *J. of Intell. Infor. Sys.*, KAP, 2(1):39-81.

Zytkow, J.M. 1993. Introduction: Cognitive Autonomy in Machine Discovery. *Machine Learning*, KAP, 12(1-3):7-16.