Ubiquitous Mining with Interactive Data Mining Agents

Xin-Dong Wu[1,2] (吴信东), *Senior Member, IEEE*, Xing-Quan Zhu[3] (朱兴全), *Member, ACM, IEEE*
Qi-Jun Chen[2] (陈琪君), and Fei-Yue Wang[4] (王飞跃), *Member, ACM, Fellow, IEEE*

[1] *School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China*

[2] *Department of Computer Science, University of Vermont, Burlington, VT 05405, U.S.A.*

[3] *Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia*

[4] *Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*

E-mail: xwu@cs.uvm.edu; xqzhu@it.uts.edu.au; feiyue.wang@mail.ia.ac.cn

Received February 27, 2009; revised July 9, 2009.

Abstract   Due to the increasing availability and sophistication of data recording techniques, multiple information sources and distributed computing are becoming the important trends of modern information systems. Many applications such as security informatics and social computing require a ubiquitous data analysis platform so that decisions can be made rapidly under distributed and dynamic system environments. Although data mining has now been popularly used to achieve such goals, building a data mining system is, however, a nontrivial task, which may require a complete understanding on numerous data mining techniques as well as solid programming skills. Employing agent techniques for data analysis thus becomes increasingly important, especially for users not familiar with engineering and computational sciences, to implement an effective ubiquitous mining platform. Such data mining agents should, in practice, be intelligent, complete, and compact. In this paper, we present an interactive data mining agent — OIDM (online interactive data mining), which provides three categories (classification, association analysis, and clustering) of data mining tools, and interacts with the user to facilitate the mining process. The interactive mining is accomplished through interviewing the user about the data mining task to gain efficient and intelligent data mining control. OIDM can help users find appropriate mining algorithms, refine and compare the mining process, and finally achieve the best mining results. Such interactive data mining agent techniques provide alternative solutions to rapidly deploy data mining techniques to broader areas of data intelligence and knowledge informatics.

Keywords   information systems, human-centered computing, data mining, intelligent agents

1 Introduction

Advances in database technologies and data collection techniques have incurred the collection of huge amounts of data. Generally, the information in a database can be divided into two categories: explicit information and implicit information. Explicit information is the information represented by the data while implicit information is the information contained (or hidden) in the data. For example, in a relational database, a tuple in a "student" table represents explicitly the basic information about a student. Other information items, such as the relationships between tables and the dependencies between attributes are also documented in the database, which are regarded as explicit information and can be recovered by using traditional database retrieval techniques. There is also implicit information. For example, the associations between the birth months of the students and their exam scores are information items that are implicit but useful. Those information items can be discovered (or mined) but cannot be retrieved. Data mining techniques are developed for this purpose. Such an operation is referred to as *data mining* or *knowledge discovery in databases* (KDD[1−3]). Data mining can be defined as the *discovery of interesting, implicit, and previously unknown knowledge from large databases*[2]. It involves techniques from machine learning, database systems, statistics, and pattern discovery.

The ability of turning data into actionable knowledge has rendered data mining a central tool for applications varying from computational sciences, business intelligence, life science, ecology and geology, to recently emerged social computing and security

informatics[4−5]. For example, in the battle of fighting terrorism[6−7], project planning[8], and E-commerce[9], data mining is identified as an essential tool where clustering, classification, and association rule analysis are commonly used to discover useful patterns from the data. Despite the above facts, the main hurdle of applying data mining to domain specific applications is twofold. 1) Data mining itself represents a large body of sophisticated algorithms with specific input and output requirements. As a result, the whole implementation process for data mining is generally tedious and expensive. 2) Many applications are facing dynamic environments with distributed data sources or computing environments. As a result, they are in need of a ubiquitous data mining platform that is easy to adjust and provides a flexible scalability whereas major data mining algorithms require the mining environments to be static and cannot be easily adjusted to meet the scalability requirements.

Three questions need to be answered in order to perform a data mining task effectively and efficiently. 1) For a specific data set, what is the most suitable data mining algorithm? Nowadays, various algorithms have been developed to deal with different problems (classification, clustering, and association mining). Even classification could imply very many different algorithms, such as C4.5[10], CN2[11], and HCV[1]. This question becomes more difficult for a data mining novice. 2) How can the data mining process be intelligently and autonomously applied to the underlying data. 3) How could the user be actively and interactively involved in the mining process? Among the above three concerns, the last two appear to be contradictive, yet they state the reality of data mining in broader domains as a whole — while users require an autonomous data mining agent with minimum user involvement, the complex and diverse data mining techniques often make the mining results meaningless unless users' participation is involved. Unfortunately, even though the research in data mining and agent technologies have made substantial progresses in recent years, rare efforts have been made to solve the above critical issues.

In this paper, we provide a unique view on using Online Interactive Data Mining, OIDM, to form a ubiquitous mining platform for domain specific applications. OIDM combines normal functions of an agent system: asking questions, integrating evidence, algorithm recommendation, and summarizing the results. The interactive property is accomplished through interviewing the user in an intelligent way and integrating the feedback from the user. As an online data mining agent, OIDM can iteratively and progressively help the user find the best mining results for his/her data mining

tasks. Instead of developing a new mining algorithm, OIDM is constructed on a large number of available data mining algorithms, as our goal is to free the user from the programming and involve the user into an active mining process. In short, OIDM has the following features:

• Each OIDM agent is an autonomous entity capable of a complete set of data mining functionalities. In addition, multiple OIDM agents can be stacked and deployed in a distributed environment to support parallel and distributed data mining.

• Each OIDM is a programming-free agent, thus no programming work and no technique details are needed from the user side.

• The interactive mechanism involves the user into an intelligent conversation process during the mining process, from which the mining results can be automatically collected to meet users' requirements.

• Multi-layer result summarization presents the mining results in a progressive way, which helps the user in interpreting the mining results.

2 Related Work

Data mining is a complicated process which usually involves data collection, preprocessing and data enhancement, mining, and interpreting and comparing the mining results[5] (as shown in Fig.1), and such a complicated process raises a huge burden on the end users especially when they are not familiar with the data mining algorithms. Consequently, agent techniques[13−16] have recently been employed to alleviate the users from being heavily involved in the complicated mining process. Such mining agents are now becoming rapidly available with dedicated mining functionalities from clustering, classification, to stream data mining. In addition, because each data mining agent is assumed to perform the mining tasks in an autonomous way, the employment of the agent techniques also provides inherent solutions to support distributed data mining where mining is simultaneously triggered at multiple distributed sites. Although data mining agents have been used for different applications, existing solutions in the area merely focused on developing autonomous mining agents and providing effective communications between the mining agents[13−16]. Due to
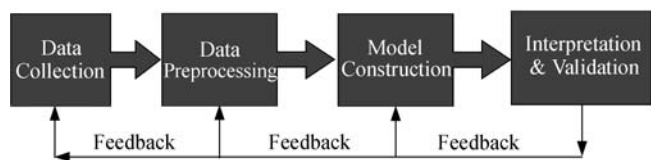

Fig.1. General data mining process (revised from Wu et al.[12]).

1020

J. Comput. Sci. & Technol., Nov. 2009, Vol.24, No.6

the inherent complexity of the data mining process and the large number of algorithms available for selection, devising fully autonomous mining agents without user interference is neither feasible nor practical. For example, given an input dataset, the mining process must know the mining objective (clustering, classification, etc.) and a target field must be clearly specified for building classification models. Meanwhile, even for classification only, a large number of learning algorithms are available and each of them may provide different results, in terms of the model interpretability and the prediction accuracies, on the same input data. As a result, it is very possible that the blind-mined results are useless unless the users are actively involved in the mining process. This, however, is inadequately addressed in the agent-based data mining research.

From the data mining perspective, the interaction between users and the mining algorithms has shown to be very helpful in customizing the mining process and confining the mining results[17−19]. To select an optimal learning algorithm for a certain task, two popular mechanisms exist: 1) one approach is to learn a decision tree for the applicability of the algorithms based on the data characteristics, and 2) another approach is a user-centered mechanism used in the consultant part of the MLT-project[18]. The survey[19], which dealt with the question of how companies can apply inductive learning techniques, concluded that the process of machine learning should primarily be user-driven, instead of data- or technology-driven. Such a conclusion can also be found among many other papers.

In the early 1990s, several researchers at the University of Aberdeen conducted a research project — CONSULTANT[18]. CONSULTANT is employed to help the user find the best classification tool for a specific dataset. CONSULTANT questions the user about the task to be solved, gathers data and background knowledge, and recommends one or more learning tools. However, with an interactive mechanism, CONSULTANT can only deal with the classification problems.

To facilitate knowledge acquisition, a model needs to be predefined in the toolbox. This is generally acknowledged in the knowledge elicitation community[29]: "The main theories of knowledge acquisition are model-based to a certain extent. The model-based approach covers the idea that abstract models of the tasks that expert systems have to perform can highly facilitate knowledge acquisition." To enhance the flexibility of a CONSULTANT-like mechanism in model construction, White and Sleeman[20] introduced MUSKRAT (Multistrategy Knowledge Refinement and Acquisition Toolbox), which includes an advisory system coupled with several knowledge acquisition tools and problem

solvers. MUSKRAT compares the requirements of the selected problem solver with the available sources of information (knowledge, data, and human experts). As a result, it may recommend either reusing the existing knowledge base, or applying one or more knowledge acquisition tools, based on their knowledge-level descriptions.

Although helpful in involving the users into the mining process, the above techniques only address data mining problems through machine learning techniques. To broaden the meaning of interactive mining, other research efforts have been made, in which interactive mining can be facilitated by visualization techniques and active data mining[21−23], or decomposing a problem into subtasks where different mining mechanisms could be involved.

Ware[22] proposed a graphical interactive approach to machine learning that makes the learning process explicit by visualizing the data and letting the user "draw" decision boundaries in a simple but flexible manner. A similar research effort can be found in Hellerstein[23]. However, even though these visualization techniques could make data mining more intuitive, it may decrease the mining efficiency in handling realistic problems where data mining could be very complicated and involve different mining mechanisms. A similar problem in statistics has been conducted by Hand[17].

In the OIDM project, we adopt a CONSULTANT-like mechanism to facilitate interactive data mining, and an interaction model is defined in advance. The reasons for using a predefined model are as follows:

1) It is simple for system management. Adding a new data mining algorithm can be accomplished by minor modifications in the system model.

2) It can help generate a compact solution for interactive mining. Though interactive, an efficient system should not require the user to answer dozens of questions before he/she can get the results. Being cooperative is the users' willingness but not his/her responsibility. To be practical, an interactive system should be as compact as possible, which means it can guide the user to achieve what they want in a few steps.

Although OIDM is similar to CONSULTANT, there are three differences between them. First, OIDM is an agent-based autonomous mining system with complete functionalities from data preprocessing, and model selection, to result interpretation and comparison. In addition OIDM agents can stack together to form a large mining system for distributed mining. In comparison, CONSULTANT is more like a single classification tool. Second, OIDM provides a broader range of data mining functionalities, which cover classification, clustering

and association analysis. Third, the goal of OIDM is to assist the user find the best data mining result not just the best classification tool.

3 Data Mining Agents and OIDM System Design

The key of using agent technologies to boost the data mining process is to rely on a number of agents to form a self-organized and self-efficient ubiquitous mining platform[24−26]. In Fig.2, we illustrate the proposed agent-based design, where the whole ubiquitous mining platform consists of a number of OIDM agents organized in three tiers: OIDM Agents, Tier 2 Agents, and Tier 1 Agents.

In short, a tier 1 OIDM agent is a collective data mining agent that dispatches data mining tasks and collects the mining results from the tier 2 OIDM agents. The main task of the tier 1 agent is to initiate the connection of all OIDM agents to form a network, and further arrange the mining jobs from distributed sites if necessary, and collect and report (e.g., visually demonstrate) the final results to the users. A tier 2 OIDM agent is mainly in charge of the parallel mining activities at each individual site, which includes initiating a number of OIDM agents in a single site to perform parallel mining. For example, for a large data repository, each OIDM agent can work on a portion of the data (scheduled by a dedicated tier 2 OIDM agent), or all OIDM agents can work on the whole data repository but each of them works towards one special data mining functionality, such as clustering or classification. A tier 2 agent does not, however, perform any particular mining activities but rather work in bridging the connection between the OIDM agents and the tier 1 OIDM agent. At the lowest level, an OIDM agent is a self-efficient data mining entity which performs all the detailed tasks. This includes interaction with end users, loading the data for preprocessing, interactive

mining, and result comparisons. Despite its comprehensive mining activities, each OIDM agent performs the mining procedure without any synchronization with other OIDM agents, other than listening to the dispatching control information from a tier 2 agent. In other words, all OIDM agents are loosely connected with no communication between the OIDM agents. As a result, a system based on the above three tier mining agents has simple yet efficient command control mechanism and minimum communication overhead.

Although the system in Fig.2 is shown to have a number of three tier OIDM agents, it is, however, not always necessary to deploy a system with such a layered structure. For example, if the mining is carried out at one individual site without any parallel mining requirement, a single OIDM agent can thus be used to fulfill the objective. Such a design makes it very flexible to construct an agent-based mining system, and also provides the simplest scenario for users to rely on a single OIDM agent to gain preliminary mining results before they can move ahead.

Among all three tiers of OIDM agents, tiers 1 and 2 agents are mainly designed for communication and information collection purposes, whereas an OIDM agent plays a critical role in the whole system design. In the following subsections, we will ignore details on tiers 1 and 2 agents but primarily focus on the design and implementation issues of the OIDM agent.

3.1 Data Mining Agent

To design an interactive data mining agent in handling realistic problems, four goals need to be achieved:

• *Autonomous* — Each data mining agent is an autonomous mining agent capable of fulfilling dedicated data mining functions, as well as communicating with other agents in the case that multiple agents are forming a networked system. This also includes collective and cooperative functionalities, such that a higher tier
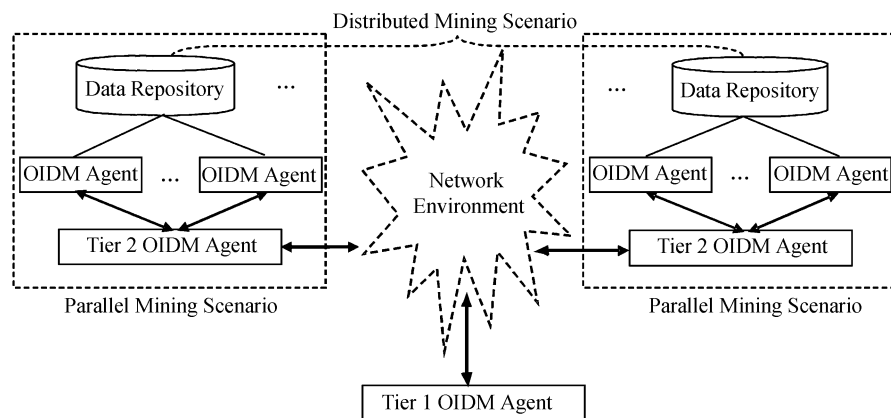

Fig.2. Agent-based ubiquitous mining platform.

agent can dispatch the mining tasks and collect results from its lower tier dependent agents.

• *Interactive* — For each individual mining agent, the interaction between the user and the system is the channel through which the system can collect information from the user. Furthermore, it is a good way to make the user know more about the mining task and the underlying data mining algorithms.

• *Complete* — To provide complete data mining solutions, each mining agent must be configured with a number of basic data mining algorithms. In addition, the agent should also be able to collect all the necessary information from the users before the algorithm selection. The user should be provided with as many choices as possible for each question.

• *Compact* — To be a compact solution and reveal the autonomous character, each mining agent should only post the indispensable questions to the user. The interviewing between the mining agent and the user must be carefully designed with intelligent progress control. A compact design should make the data mining process as intuitive as possible.

3.2 OIDM Agent System Workflow

The system framework of OIDM is shown in Fig.3. It runs by following a predefined model. First, OIDM recommends to the user one specific mining algorithm through the Algorithm Selection Module (Subsection 3.3). Once the algorithm is selected, OIDM asks the user to provide input data. The user can choose to upload data files or paste the data in the given text areas
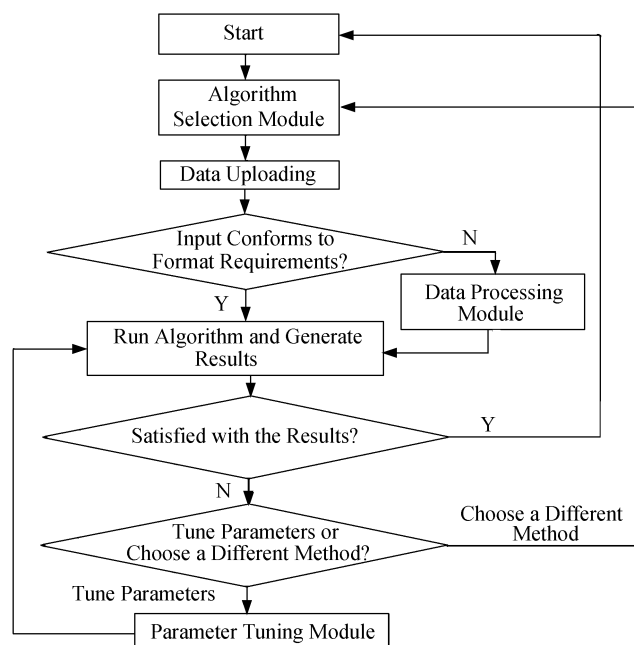
on OIDM. Based on the input data, OIDM constructs the input files, which conforms to the selected algorithm through the Data Processing Module (Subsection 3.4). After the data processing stage, OIDM runs the mining algorithm on the input data and provides the user with the results through a Multi-Level Summarization Mechanism (Subsection 3.5). The user may find the results not satisfactory. To refine the results, the user can choose to tune the parameters through the Parameter Tuning Module (Subsection 3.6) or select a different algorithm. By doing this, OIDM can not only guide the user to select the right mining tool, but also provide the experimental result comparisons between different mining mechanisms or different parameter settings of the same algorithm. OIDM will follow this iterative workflow until useful mining results are found.

3.3 OIDM Data Mining Modules

OIDM consists of the following eight typical mining algorithms, which cover three most popular categories of data mining problems: classification, clustering and association analysis. In this subsection, we introduce the interaction model that is used to guide the user in selecting a mining algorithm. The functionalities of the system can be easily extended through adding more mining algorithms.

• C4.5[10]: a decision tree construction program.

• C4.5 Rules[10]: a program that generates production rules from unpruned decision trees.

• HCV[1]: an extension matrix based rule induction algorithm.

• OneR: a program that constructs one-level rules that test one particular attribute only.

• Prism: an algorithm for inducing modular rules.

• CobWeb: an incremental clustering algorithm, based on probabilistic categorization trees.

• K-Means: a simple clustering algorithm that randomly select clusters centers and iteratively generates a certain number of clusters.

• Apriori: an algorithm for mining frequent itemsets for Boolean association rules.

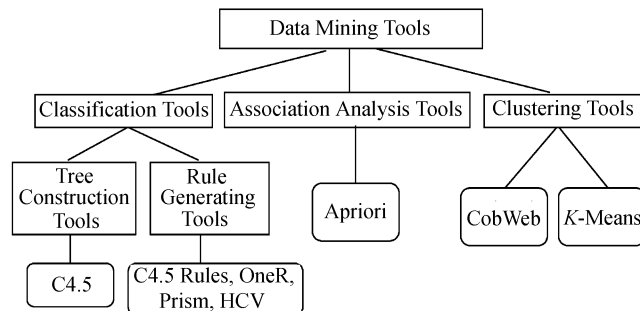Fig.3. System workflow of the OIDM agent.

Fig.4. Data mining models integrated in an OIDM agent.

These eight algorithms are organized into a hierarchy, as shown in Fig.4, to help users clarify their mining task. Algorithm selection follows this hierarchy. If the user has no knowledge about data mining or is not sure about which algorithm is the most suitable, the system can help him/her choose one through providing some typical sample mining tasks and asking the user to choose a similar one.

3.4 OIDM Data Processing Module

All data mining software packages require the input data follow a specific data format (such as csv — a comma delimited format) before the algorithm can actually run on a given dataset. Furthermore, most algorithms require the user to provide the domain knowledge for the raw data such as what are the possible values for a particular nominal attribute. OIDM provides the Data Processing Model (DPM) to help the user. DPM can extract domain knowledge automatically from the input data and ask the user to refine the domain knowledge if necessary. Through asking the user a serial of common questions (such as which attribute will be treated as the class label, and whether a specific attribute is nominal or continuous), DPM can convert the original data file (if the field delimiter is other than the comma) and construct input files that meet a specific algorithm's input format automatically. The only input from the user would be the data files and answers for some specific questions. Two common input files can be generated through DPM: arff (for the WEKA package) and names & data (for C4.5 and HCV).

3.5 OIDM Multi-Level Summarization Module

Summarizing the mining results can be very useful, as some mining mechanisms generate complicated results. For example, it is quite common to generate more than 100 classification rules in a typical classification problem. Consuming the whole results at one time is unnecessary sometimes. Consequently, instead of showing the whole details of the mining results, we use a Multi-Level Summarization Model, which provides the user with two levels of results. At the first level, general information of the mining results is provided, such as the classification accuracy, the coverage of the results, and the statistical information, from which the user can easily get the performance information of the selected algorithm. If the user is particularly interested in a specific result, he/she can continue to drill into the second-level results which include the details of the mining results, such as the classification rules, clusters, and the coverage and accuracy of the association rules. By using the multi-level summarization module, the system can be more informative and practical.

3.6 OIDM Parameter Tuning

To improve the mining results, OIDM provides two types of interactions: 1) selecting different types of mining algorithms, and 2) setting different parameters for a specific algorithm (see Fig.3). Experienced users usually know which mining algorithm to use before launching OIDM. Therefore, parameter tuning is more useful to them. OIDM provides a unique set of common parameters for each mining algorithm. Consequently, different parameter tuning options are presented based on the selected algorithm. Simple explanations for the parameters are also provided. The default parameter values are set initially. For detailed explanations of the parameters, the user may refer to the online algorithm manuals, which are also linked in OIDM.

4 A Running Example of OIDM

In this section, we present a demonstration where an OIDM agent is used to solve a specific data mining problem (classification based).

[1] Start Trip (see Fig.5).

[2] **Q**: Which data mining tool would you like to use?
   (a) Clustering
   (b) Classification
   (c) Association analysis
   (d) Not sure

   A: Not sure

[3] **Q**: Which of the following categories of problems is your problem similar to?
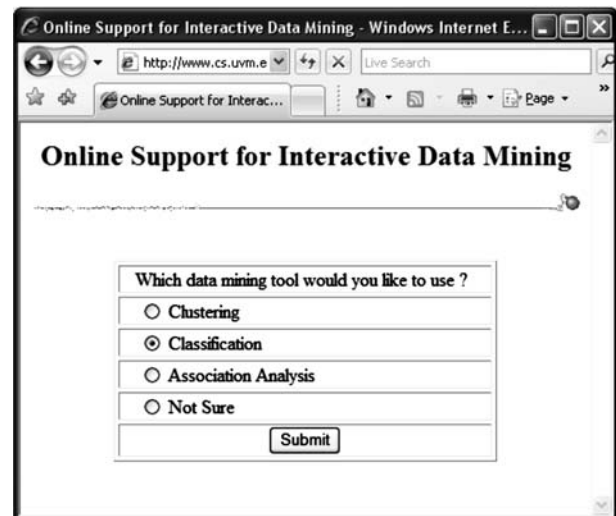   (a) Segment a customer database based on similar buying patterns

Fig.5. OIDM starting page.

(b) Find out common symptoms of a disease

(c) Find out whether customers buying beer will always buy diapers

(d) None of the above

A: Find out common symptoms of a disease

[4] **Q**: Would you prefer the output in the form of:

(a) A decision tree

(b) A set of IF-THEN rules

A: A decision tree

[5] **Q**: Which classification tool would you like to use?

(a) C4.5

A: C4.5

[6] Data uploading page (see Fig.6)

[7] Data processing page (see Fig.7)

[8] Result page (see Fig.8)

[9] **Q**: Are you satisfied with the results?

(a) Yes

(b) No

A: No

[10] **Q**: Choose a different method?

(a) Choose a different method

(b) Tune parameters

A: Tune parameters

[11] Parameter tuning page (see Fig.9)

[12] Result page (see Fig.8)

Q: Are you satisfied with the results?

(a) Yes

(b) No

A: Yes
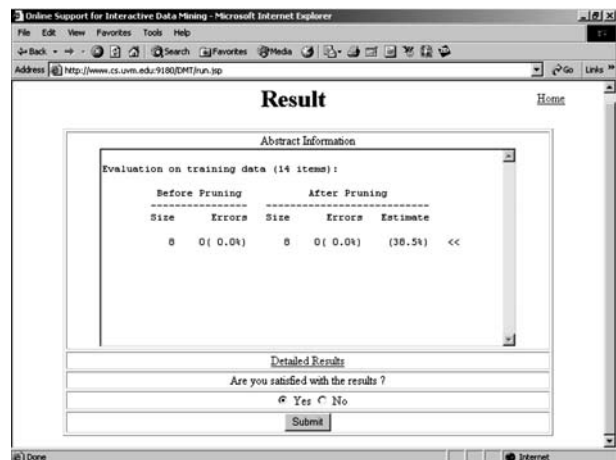
[13] Start page

Fig.6. Data uploading page.

Fig.7. Data processing page.

Fig.8. Multi-level result page.
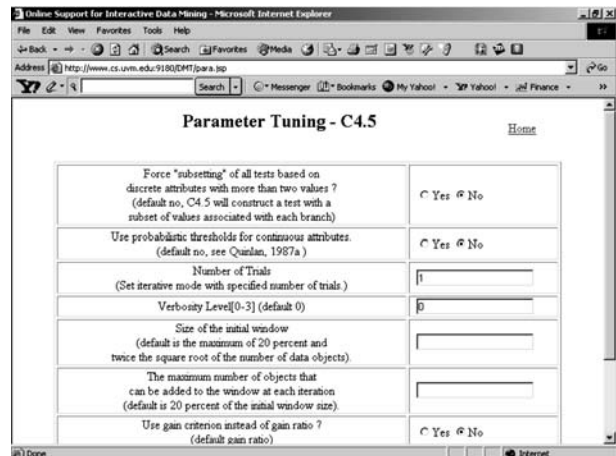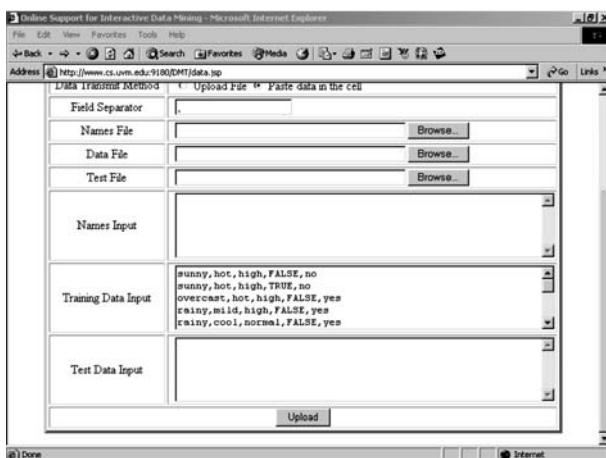

Fig.9. Parameter tuning page.

5 System Evaluation

To evaluate the system runtime performance, we

select five datasets from the UCI data repository[27], and use OIDM to carry out a predefined mining process (C4.5 decision tree construction). In Table 1, we report the system runtime (tree construction) and the corresponding prediction accuracies. The overall results suggest that OIDM is very effective in assisting build decision models in an iterative manner. The expected tree construction time is usually less than twenty seconds even for datasets with more than sixty thousands instances.

To further assess the system performance from open-domain users, we have released the website of OIDM (single agent version) online[①] and collected the feedback from the users. Basically, all suggestions conclude that OIDM is a useful agent. It is expected that two types of people would be particularly interested in OIDM: 1) Students who have taken a data mining or artificial intelligence course and are interested in conducting research in data mining. Their feedback indicates that OIDM is helpful in understanding the basic concepts of data mining and the different mining algorithms. 2) Senior researchers whose feedback suggests that OIDM is helpful in generating preliminary experimental results and conducting data mining research.

In summary, the above feedback suggests that OIDM is a useful and effective data processing agent, which could benefit both junior and senior researchers from different perspectives.

Table 1. OIDM Runtime and Prediction Accuracies

Datasets	No. Training	No. Test	Runtime (s)	Accuracy (%)
Adult	40 957	7 885	17.89	85.50
Connect	60 005	7 552	10.27	81.46
Letter	17 562	2 438	9.48	86.67
Mushroom	7 298	826	0.37	100.00
Splice	2 890	300	1.07	93.00

6 Conclusions

Recent development in computer networking and storage technologies has raised the needs of discovering patterns and knowledge from large data collections. Such data mining requirements have now rapidly spanned from traditional computational science to business intelligence, life science, environment science[28], and security informatics. In practice, the efforts on performing data mining are usually nontrivial, especially for applications which require a ubiquitous mining platform whereas users are generally from the outside of the engineering and computational science domains. Devising simple yet efficient data mining agents thus becomes increasingly important for data mining to meet the real-world requirements.

In this paper, we have reported an agent-based intelligent data analysis system with a primary focus on devising an online interactive data mining agent — OIDM, which supports online interactive data mining, through which the user can acquire optimal mining results without any programming work. We argued that existing research merely focused on developing autonomous mining agents. Due to the inherent complexity of the data mining techniques and a large number of algorithms available for selection, interactive data mining agents provide a solution to involve users in the mining process in a simple yet effective way. As a mining agent, OIDM can interact with the user and help him/her fulfill the learning task. Autonomous, interactive, complete, and compact are the four goals that guide the system design of OIDM. At the beginning, the user may be ignorant about either the data mining process or the knowledge in the data, and OIDM can take specific steps according to the information gathered from the user and perform the learning task automatically until satisfactory results are found. Compared with other similar tools, OIDM possesses the following unique and useful features: 1) OIDM is a mining agent capable of solving typical data mining problems such as classification, clustering and association mining, whereas most other data mining toolboxes only address one type of data mining problems, and they recommend the best algorithm. 2) The interaction model of OIDM is compact. In other words, it can guide the user to find the right algorithm in only a few steps. 3) OIDM uses a multi-level summarization mechanism to present the mining results, which is useful in helping the user understand the results. 4) Each OIDM agent is an independent entity to fulfill a variety of data mining functionalities, and multiple OIDM agents can stack together to form an agent-based distributed or parallel data mining system with an arbitrary number of agents. The evaluation results suggest that, though data mining is a complex task, an OIDM agent provides a simple and flexible solution through actively involving the user in the mining process.

References

[1] Wu X. Knowledge Acquisition from Databases. Ablex Publishing Corp., 1995.

[2] Fayyad U M, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds.) Advances in Knowledge Discovery and Data Mining. 1996, AAAI Press, pp.1–34.

[3] Zhu X, Davidson I. Knowledge Discovery and Data Mining: Challenges and Realities. IGI Global, 2007.

[4] Wang F, Carley K, Zeng D, Mao W. Social computing: From

[①]The single agent based OIDM system is currently available at http://www.cs.uvm.edu:9180/DMT/index.html.

social informatics to social intelligence. *IEEE Intelligent Systems*, 2007, 22(2): 1541–1672.

[5] Chen H, Wang F, Zeng D. Intelligence and security informatics for homeland security: Information, communication, and transportation. *IEEE Trans. Intelligent Transportation Systems*, 2004, 5(4): 329–341.

[6] Chen H. Intelligence and Security Informatics for International Security: Information Sharing and Data Mining. Springer, 2006.

[7] Chen H, Reid E, Sinai J, Sike A, Ganor B. Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security. Springer, 2008.

[8] Xu K, Munoz-Avila H. CaBMA: A case-based reasoning system for capturing, refining, and reusing project plans. *Knowledge and Information Systems*, 2008, 15(2): 215–232.

[9] Zhuang Y, Fong S, Shi M. Knowledge-empowered automated negotiation system for E-commerce. *Knowledge and Information Systems*, 2008, 17(2): 167–191.

[10] Quinlan J R. C4.5: Programs for machine learning. *Machine Learning*, 1994, 16(3): 235–240.

[11] Clark P, Niblett T. The CN2 induction algorithm. *Machine Learning*, 1989, 3(4): 261–283.

[12] Wu X, Yu P, Piatetsky-Shapiro G, Cercone N, Lin T, Kotagiri R, Wah B. Data mining: How research meets practical development. *Knowledge and Information Systems*, 2003, 5(2): 248–261.

[13] Petrie C. Agent-based engineering, the Web, and intelligence. *IEEE Expert: Intelligent Systems and Their Applications*, 1996, 11(6): 24–29.

[14] Zhong N, Ohsuga S, Liu C, Kakemoto Y, Zhang X. On meta levels of an organized society of KDD agents. In *Proc. the 1st European Symposium on Principles of Data Mining and Knowledge Discovery*, Trondheim, Norway, June 24–27, 1997, pp.367–375.

[15] Ong K, Zhang Z, Ng W, Lim E, Agents and stream data mining: A new perspective. *IEEE Intelligent Systems*, May/June 2005, 20(3): 60–67.

[16] Klusch M, Lodi S, Moro G. The role of agents in distributed data mining: Issues and benefits. In *Proc. IEEE/WIC International Conference on Intelligent Agent Technology* (*IAT 2003*), Beijing, China, Oct. 13–17, 2003, p.211.

[17] Hand D. Decomposing statistical question. *Journal of the Royal Statistical Society*, Series A, 1994, 157: 317–356.

[18] Craw S. CONSULTANT: Providing advice for the machine learning toolbox. In *Proc. the BCS Expert Systems Conference*, Cambridge, UK, 1992, pp.5–23.

[19] Verdenius F. Applications of inductive learning techniques: A survey in the Netherlands. *AI Communications*, 1997, 10(1): 3–20.

[20] White S, Sleeman D D. Providing advice on the acquisition and reuse of knowledge bases in problem solving. In *Knowledge Acquisition Workshop*, Singapore, Nov. 22–23, 1998.

[21] Motoda H. Active mining, a spiral model of knowledge discovery. Invited talk of the *2002 IEEE International Conference on Data Mining*, Maebashi City, Japan, Dec. 9–12, 2002.

[22] Ware M, Frank E, Holmes G, Hall M, Written I. Interactive machine learning: Letting users build classifiers. *International Journal of Human Computer Studies*, 2001, 55(3): 281–292.

[23] Hellerstein J, Avnur R, Chou A, Hidber C, Olston C, Raman V, Roth T, Haas P. Interactive data analysis: The control project. *IEEE Computer*, 1999, 32(8): 51–59.

[24] Micacchi C, Cohen R. A framework for simulating real-time multi-agent systems. *Knowledge and Information Systems*, 2008, 17(2): 135–166.

[25] Nguyen N, Katarzyniak R. Action and social interactions in multi-agent systems. *Knowledge and Information Systems*, 2009, 18(2): 133–136.

[26] Resconi G, Kovalerchuk B. Agents' model of uncertainty. *Knowledge and Information Systems*, 2009, 18(2): 213–229.

[27] Newman D, Hettich S, Blake C, Merz C. UCI repository of machine learning. Irrine, CA: University of California, Department of Information and Computer Science, 1998.

[28] Zhang K, Fan W. Forecasting skewed biased stochastic Ozone days: Analyses, solutions, and beyond. *Knowledge and Information Systems*, 2008, 14(3): 299–326.

[29] Heijst V, Terpstra G, Wielinga P, Shadbolt N. Using generalised directive models in knowledge acquisition. In *Proc. EKAW 1992*, Heidelberg and Kaiserslautern, Germany, May 18–22, 1992, pp.112–132.

Xin-Dong Wu is a professor and the chair of the Computer Science Department at the University of Vermont, USA. He holds a Ph.D. degree in artificial intelligence from the University of Edinburgh, Britain. His research interests include data mining, knowledge-based systems, and Web information exploration. He has published over 180 refereed papers in these areas in various journals and conferences, including IEEE TKDE, TPAMI, ACM TOIS, DMKD, KAIS, IJCAI, AAAI, ICML, KDD, ICDM, and WWW, as well as 23 books and conference proceedings. His research has been supported by the U.S. National Science Foundation (NSF), the U.S. Department of Defense (DOD), the National Natural Science Foundation of China (NSFC), and the Chinese Academy of Sciences, as well as industrial companies including U.S. West Advanced Technologies and Empact Solutions. Prof. Wu is the founder and current steering committee chair of the IEEE International Conference on Data Mining (ICDM), the founder and current editor-in-chief of Knowledge And Information Systems (KAIS, by Springer), the founding chair (2002∼2006) of the IEEE Computer Society Technical Committee on Intelligent Informatics (TCII), and a series editor of the Springer Book Series on Advanced Information and Knowledge Processing (AI&KP). He was the editor-in-chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE, by the IEEE Computer Society) between January 1, 2005 and December 31, 2008, and served program committee chair for ICDM'03 (the 2003 IEEE International Conference on Data Mining) and as program committee co-chair for KDD-07 (the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining). Prof. Wu is the 2004 ACM SIGKDD Service Award winner, the 2006 IEEE ICDM Outstanding Service Award winner, and 2005 chair professor in the Changjiang (or Yangtze River) Scholars Program at the Hefei University of Technology appointed by the Ministry of Education of China. He has been an invited/keynote speaker at numerous international conferences including IEEE GrC 2009, IDEAL 2009, JCKBSE 2008, HAIS 2008, NSF-NGDM'07, PAKDD-07, IEEE EDOC'06, IEEE ICTAI'04, IEEE/WIC/ACM WI'04/IAT'04, SEKE 2002, and PADD-97.

Xing-Quan Zhu received his Ph.D. degree in computer science from Fudan University, China, in 2001. He is currently an associate professor of the Faculty of Engineering and Information Technology, University of Technology, Sydney (UTS), Australia. Before joining UTS, he was a tenure track Assistant professor in the Department of Computer Science & Engineering, Florida Atlantic University, USA (2006∼2009), a research assistant professor in the Department of Computer Science, University of Vermont, USA (2002∼2006), and a postdoctoral associate in the Department of Computer Science, Purdue University, USA (2001∼2002). Dr. Zhu's research mainly focuses on data mining, machine learning, and multimedia systems. Since 2000, he has published more than 90 refereed journal and conference proceedings papers in these areas. He is an associate editor of the IEEE Transactions on Knowledge and Data Engineering.
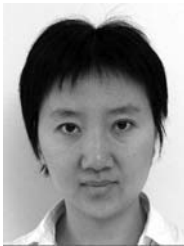
Qi-Jun Chen received her Master's degree in computer science from the University of Vermont, USA. She is currently a database administrator of the West Virginia University. Her research interests include inductive learning from large databases.

Fei-Yue Wang received his Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, New York in 1990. He jointed the University of Arizona in 1990 and became a professor and the director of the Program for Advanced Research in Complex Systems (PARCS) in 1999. In 1999, he found the Intelligent Control and Systems Engineering Center at the Chinese Academy of Sciences, China, under the support of the Outstanding Oversea Chinese Talents Program. Since 2002, he is the director of the Key Laboratory of Complex Systems and Intelligence Science at the Chinese Academy of Sciences. Currently, he is the vice president for research, education, and academic exchange at the Institute of Automation, Chinese Academy of Sciences. He is a member of ACM Council, editor-in-chief of IEEE Intelligent Systems and IEEE Transactions on ITS. Dr. Wang is a member of Sigma Xi and fellow of IEEE, INCOSE, IFAC, ASME, and AAAS. His current research interests include social computing, Web and services science, modeling, analysis, and control of complex systems, especially social and physical/cyber systems. He was the editor-in-chief of the International Journal of Intelligent Control and Systems from 1995 to 2000, editor in charge of the Series in Intelligent Control and Intelligent Automation from 1996 to 2004, EiC, associate EiC, or associate editors of 10 IEEE Transactions and Magazines. Since 1997, he has served as general or program chair of more than 20 IEEE, INFORMS, ACM, ASME international conferences. He was the president of IEEE ITS Society from 2005 to 2007, the president of Chinese Association for Science and Technology (CAST, USA) in 2005, and the president of the American Zhu Kezhen Education Foundation from 2007∼2008. In 2007, he received the National Prize in Natural Sciences of China and was elected as the Outstanding Scientist by ACM for his work in intelligent control and social computing.