Mona maghraby

بسم اللّه الرحمن الرحيم

مركز الشبكات وتكنولوجيا المعلومات

قسم التوثيق الإلكتروني

Mona maghraby

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها

علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

Mona maghraby

بعض الوثائق الأصلية تالفة
وبالرسالة صفحات لم ترد بالأصل

A MULTI-AGENT SYSTEM FOR KNOWLEDGE DISCOVERY IN DATABASES

By

Abdalla Ahmed Abdalla

Supervised By

Dr. Abdelaziz Khamis

Computer Science Dept.,
Institute of Statistical
Studies and Research

Dr. Hesham Hassan

Computer Science Dept.,
Faculty of Computers and
Information

A Thesis Submitted to the institute of Statistical Studies and Research, Cairo University, in partial fulfillment for the degree of M.Sc., in the department of Computer Science.

December 2000

CERTIFICATE

I certify that this work hasn't been accepted in substance for any academic degree and is not being concurrently submitted in candidature for any other degree.

Any portions of thesis for which I'm indebted to other sources are mentioned and explicit references are given.

Student: Abdalla Ahmed Abdalla

APPROVAL SHEET

A MULTI-AGENT SYSTEM FOR KNOWLEDGE DISCOVERY IN DATABASES

By

Abdalla Ahmed Abdalla

This thesis for the M.Sc. Degree in Computer Science and Information Systems, Department of Computer Science, Institute of statistical Studies and Research, Cairo University, has been approved by:

Name	Signature
Prof.Dr. Ebada Ahmed	
Prof.Dr. Abdelaziz Khamis	A. Khamis
Prof.Dr. Reem Bahgat	Reem Bahgat

Date: / /

Acknowledge

I have to thank my god for what he gave me allover my life.

I wish to express my deepest gratitude, sincere appreciation to ***Prof.Dr. Abdelaziz Khamis***, and ***Dr. Hesham Hassan*** for their valuable advice and guidance.

I also thank my family for being patient with me while working in this thesis for too long periods of time.

Abstract

The general objective of this work is to develop a multi-agent system for knowledge discovery, which is characterized by the capabilities that enable it to work efficiently within open and distributed systems.

In order to achieve the above objective, we use one of the most recent agent analysis and design methodology called ***GIAI*** and its extension ***Organizational Abstractions*** in the analysis and design of our system.

The most beneficial outcome of using this methodology is to obtain an architectural design of the system described in terms of a set of models and rules (organizational rules, role model, acquaintance model, agent model, and service model). The architectural design of the system addresses the issues of open systems and supports organization metaphor. So our proposed system (**MACKDD**: **M**ulti-agent **A**rchitecture for **C**ooperative **K**nowledge **D**iscovery in **D**atabases) has several advantages over the previous systems that integrate multi-agent technology with either knowledge discovery or information retrieval (e.g., an organized society of autonomous knowledge discovery agents (GLS), MACRON system, agent based knowledge discovery and intelligent agent-assisted decision support systems).

Our system achieves the following advantages:

- **Flexible organization (a set of interactive sub-organizations).** Each sub-organization or agent has a specific role in the system.
- **Open multi-agent system capabilities.** Our system incorporates functional manager agents. Those agents model other agents' capabilities and their changes. Also our system use organizational rules to control self-interested agents, and rational based monitors to monitor environmental changes that invalidate the plan.
- **Complex agent architecture.** Instead of using a society of simple agents, we use a complex architecture in building our agents, to achieve autonomy,

proactivity, and cooperation with other agents to maximize system's output. Those goals are achieved through providing the agent the following components:

1. **Planner.** It takes as input a set of goals and produces a plan that satisfies those goals. It uses TAEMS [Task Analysis, Environment Modeling and Simulation] model to represent agents' tasks. This model can represent tasks' quantitative and quality knowledge.
2. **Coordinator.** It is responsible for coordinating agents' plans to maximize system's output. Our contribution to the coordinator is extending (GPGP [Generalized Partial Global Planning]) with team capabilities.
3. **Directed goal criteria scheduler.** It supports dynamic changing goal criteria.
4. **Execution monitor**. It takes as input the agent's next intended action and prepares, monitors and completes its execution.

- **Achieving second-generation's goals of knowledge discovery systems.** Those goals are achieved through the following:
  1. Representation of optimization techniques by using TAEMS and GPGP.
  2. Knowledge management and refinement.

In this thesis we describe a variety of data mining techniques and their evolution. Also we discuss the different techniques and methodologies of cooperation and agents' analysis and design.

We implement our system by using two packages (ZEUS toolkit, WEKA). We use ZEUS to implement agents' characteristics, and we use WEKA to implement different steps of knowledge discovery process. Also we use the Java language to implement some modifications to the previous packages to reflect our system's features.

We provide some experiments that reflect the most important features of our system.

Table of Contents

Chapter 3: Agent Technology

Chapter 4: MACKDD Architecture

Chapter 5: Experiments

Chapter 6: Conclusion and Future Work

Chapter 1

Introduction

Agent based knowledge discovery provides a new technique for performing data mining. By combining techniques from distributed AI and machine learning, software agents equipped with learning algorithms mine datasources. A group of these agents will be able to cooperate to discover knowledge from databases.

A multi–agent approach to knowledge discovery is promising for variety of reasons [Decker et al. 95]:

- Multiple agents offer concurrency which is a big win in time in constrained situations or when the search space is very large. Query plans can often be decomposed into relatively independent sub-plans with few interdependencies. An agent executing a sub-plan functions relatively autonomously but needs to coordinate with others where interdependencies exist.
- When a system is dealing with enormous quantities of data, distributed computation at the sites where the data resides may often be viable approach compared to migrating data to a centralized processing location. Agents can reside at the data sources and perform distributed coordinated knowledge discovery.
- Agent-based architectures offer modularity, robustness, and other advantages of a distributed System.

Traditional knowledge discovery systems can't work efficiently within dynamic environments because the mining tasks are generally mapped onto static, pre-specified mining plans. Such systems don't have the capabilities that enable it to exploit dependencies among mining methods, monitor environment changes, to adopt itself with dynamically changing goal criteria, and provide intelligently trade–of solution quality for resource limitations. Therefore a sophisticated view of the knowledge discovery process (KDD in short) is required, where the process must be dynamic, incremental, and constrained with resource limitations. In order to achieve an integration of data mining techniques and software agents a number of issues need to be addressed before such integration can take place. Such issues include:

KDD process: The discovery process is a multi-step process, and the KDD agents should be divided into three groups based on the learning phases (pre-processing, knowledge elicitation, refinement and management), therefore the planning should be done in a multi-step mode. Furthermore since the discovery process is a repetitive process, this implies that revising a plan and re-planning are necessary [Zhong et al. 97].

Dynamism*:* A central problem faced by many agents who plan is that the world is constantly changing. This is especially true in the case of multi-agent environments, since agents are inherently dynamic entities. They appear and disappear, and their abilities may change over time. For example, changing abilities of KDD agents (i-e availability of more efficient data mining algorithms). Also, the construction of plan and scheduling and execution of the plan's actions take time. During this time the environment may change, invalidating the plan, this implies that implementation of rational based monitors to detect changes in the environment that are relevant to the plan is required [Veloso et al. 98].

Knowledge sharing*:* There is an interoperation between the different agents in the entire KDD System. In order to facilitate such interoperation, it is required that the knowledge communication is interpreted in a clear manner. KQML is a language developed for this purpose [Finin et al. 94].

System Coordination: It is important for all various agents under the KDD system architecture to be coordinated to contribute towards the overall KDD process. In many cases, agents' plans may be either conflicting or ambiguous because of