

Hybrid Knowledge-based Support with Learning Abilities

by

Maung Aung Soe Paing (Lawrence)

M-94-6317-2

A Thesis submitted in partial fulfillment of the requirements of the Degree of

**Master of Science
in
Software Engineering**

**Faculty of Science and Technology
UNIVERSITY OF MACAU**

October, 1997

Table of Contents

1.	Introduction	1-1
1.1	Thesis Outline	1-1
2.	CBR General Concept	2-1
2.1	What is Case-based Reasoning	2-1
2.2	CBR Software Tools	2-1
2.2.1	Remind	2-2
2.2.2	ESTEEM	2-2
2.2.3	KATE	2-3
2.3	Comparison On 'Representation Of Cases'	2-3
2.3.1	Flat Vs Structured (KATE)	2-4
2.3.2	Nested Cases (ESTEEM)	2-5
2.3.3	Partially Nested Flat Cases (ReMind)	2-6
3.	Data Mining General Concept	3-1
3.1	Why is Data Mining?	3-1
3.2	The KDD Process	3-1
3.3	An Overview of Data Mining Methods	3-3
3.4	Related Fields and Tools	3-3
4.	KATE Tool Sets	4-1
4.1	Overview	4-1
4.2	How does KATE works?	4-2
4.2.1	Model Editor	4-2
4.2.2	Questionnaire Generator	4-4
4.2.3	Runtime Generator	4-4
4.3	Case Retrieval Functionality	4-4
4.3.1	KATE-INDUCTION	4-5
4.3.2	KATE-CBR	4-5
4.3.3	Dynamic Consultation	4-5
4.4.1	Importing Cases into Kate	4-5
4.4.1	Databases Formats	4-5
4.4.2	Text Database Organization	4-6
4.4.3	Defining Reference Cases and Test Cases	4-6
5.	CBR Applications and Examples	5-1
5.1	Testing the Decision Tree	5-1
5.2	Testing CBR	5-2
5.3	Case-based Reasoning Applications in KATE	5-2
5.4	CBR and Induction Test Result from Case Studies	5-9
5.5	Discussion	5-12
5.6	Conclusion	5-13
6.	Adaptation in KATE	6-1
6.1	Incorporating Adaptation Functions In KATE	6-1

6.2	Objectives of Developing the Rule-Based Adaptation	6-2
6.3	Application Of User Defined Rule Base	6-3
6.4	Future Extension to KATE-ADAPTATION	6-6
7.	Clementine	7-1
7.1	Overview & Architecture	7-1
7.2	Data Access	7-2
7.2.1	Data Types	7-2
7.3	Data Manipulation	7-3
7.4	Display and Visualization	7-4
7.5	Modeling	7-4
7.5.1	Neural networks	7-5
7.5.2	Rule Induction	7-6
7.5.3	Regression Modeling	7-6
7.6	Testing Models	7-6
7.7	Browsing Models	7-6
7.8	Exporting Data and Results	7-7
7.9	Model Deployment	7-7
8.	Data Mining Applications and Examples	8-1
8.1	Data Mining Models on Piles Foundations Database	Case 1-1
8.2	Data Mining Models on Abalone Database	Case 2-1
8.3	Data Mining Models on 1985 Auto Imports Database	Case 3-1
8.4	Data Mining Models on Car Evaluation Database	Case 4-1
8.5	Data Mining Models on Echocardiogram Database	Case 5-1
8.6	Data Mining Models on Protein Localization Sites Database	Case 6-1
8.7	Data Mining Models on Glass Identification Database	Case 7-1
8.8	Data Mining Models on Hepatitis Database	Case 8-1
8.9	Data Mining Models on Relative CPU Performance Database	Case 9-1
8.10	Data Mining Models on Pima Indians Diabetes Database	Case 10-1
9.	CBR and Data Mining	9-1
9.1	Introduction	9-1
9.2	Induction	9-1
9.2.1	Induction from KATE	9-2
9.2.2	Induction from Clementine	9-3
9.3	Comparison on Induction Results from KATE-INDUCTION and Clementine	9-4
9.4	Interfacing Neural Network Models and Regression Analysis with CBR Adaptation Process	9-5
9.5	Weak Integration	9-7
9.6	Strong Integration	9-8
10.	Conclusions	10-1
10.1	Summary of Contributions	10-1
10.1.1	Rule-based Adaptation	10-1
10.1.2	Prototypes on CBR and Data Mining Models	10-2
10.1.3	Integration of KATE Adaptation and Data Mining Methods (Neural Network & Regression Analysis)	10-3

10.1.4	Knowledge Discovery Support Environment - KDSE	10-3
10.2	Limitation of This Work	10-3
10.3	Future Research Directions	10-4
11.	References	Ref-1
12.	List of Figures and Tables	FT-1