MSTCEC
002
ZHANG

Knowledge Discovery in Medical Diagnosis Data

by

Zhang Siqi

Master of Computer and Information Science

2005



Faculty of Science and Technology
University of Macau

TABLE OF CONTENTS

List of Figures iii

List of Tables iv

Glossary..... v

CHAPTER 1 Introduction 1

1.1 Background 1

1.2 Development of Medical Intelligence Diagnosis 3

1.3 Introduction of Data Mining and KDD 6

1.3.1 Basic Definition 6

1.3.2 The Six-Step DMKD Process Model 9

1.3.3 The Data-Mining Step of the KDD Process 10

1.4 Works of this Thesis 17

CHAPTER 2 Data Preprocessing 19

2.1 The Characters of Medical Records 19

2.2 The Definition of Data Preprocessing 21

2.3 Data Preprocessing in Medical Data 23

2.3.1 Data Intergration 23

2.3.2 Makeup the Missing Data 24

2.3.3 Discretization 25

2.3.4 Data Conversion 31

2.3.5 Feature Selection 32

2.4 Summary 36

CHAPTER 3 Decision Tree 38

3.1 Data Classification 38

3.2 Decision Tree Inducation Techniques 39

3.3 Multiclass Feedback Decision Tree 44

3.3.1 Multiclass as Multiple Two-Class Problems 44

3.3.2 Feedback Decision Tree 45

3.4 Extracting Classification Rules 49
3.5 Reprocessing the Result of Mining 50
3.6 Summary 51
CHAPTER 4 Simulation and Experimental Study 53
4.1 Interoduction 53
4.2 User-System Interaction 54
4.3 Data Preprocessing 55
4.3.1 Data Integration 55
4.3.2 Extracting and Clearing the Database 56
4.4 Knowledge Discovery 58
4.4.1 Buliding Decision Tree and Extraction Rules 58
4.4.2 Decision Tree with Feature Selection 59
4.4.3 Diagnosis Rules with Interestingness Measure 61
4.4.4 Evaluations of Feedback DT 61
4.5 Summary 62
CHAPTER 5 Conclusion 63
Bibliography 65
Appendix : Publications 73