

MASTER'S THESIS

Data mining of classic Chinese medicinal formulae using artificial neural network

Dong, Mouzhi

Date of Award:
2000

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

**Data Mining of Classic Chinese Medicinal
Formulae Using Artificial Neural Network**

DONG Mouzhi

**A thesis submitted in partial fulfillment of the requirements
for the degree of
Master of Philosophy**

August 2000

Hong Kong Baptist University

Abstract

The objective of the research is to analyze whether Machine Learning technology is suitable to apply in the field of tonics for deficiency syndromes formulae study. In the thesis, emphasis is placed on analyzing whether the effects learned by neural networks can match those derived from the theory of Chinese Medicine.

While neural network, a useful Machine Learning model, is getting more and more popular in various fields, its application in Chinese Medicine is still a new topic. A major goal of the research is to find the compatibility relationship in formulae after learning a series of classic Chinese medicinal formulae by neural network.

Firstly, 87 classic Chinese medicinal formulae database are built up as whole data set. Secondly, the formulae are divided into two disjoint sets. 79 formulae in the training set, and 8 formulae in the testing set. Finally, not only single remedy relationship is found in the training formulae, but herb pairs as well. After training by both Adaptive Linear Neuron networks (ADALINE) and backpropagation neural network, prediction accuracy is compared. Multilayer backpropagation where the first layer is sigmoid and the second layer is linear will suit it well because of nonlinear property of compatibility.

Although the current research is focused on the effect of tonics for deficiency syndromes, it could be extended to other kinds of Chinese herb analysis. Based on the findings of the present study, the machine learning technology may be a valuable approach for the data mining in Chinese Medicine.

Table of Contents

Content	Page
DECLARATION	i
ABSTRACT	ii
ACKNOWLEDGEMENT	iii
TABLE OF CONTENTS	iv
LIST OF TABLES.....	vi
LIST OF FIGURES	viii
CHAPTER ONE INTRODUCTION	1
1.1 DEVELOPMENT OF CHINESE MEDICINE	1
1.2 THEORY SYSTEM OF CHINESE MEDICINE	1
1.2.1 <i>Bian Zheng Shi Zhi</i>	1
1.2.2 <i>Chinese Herbs – The basic component to treat diseases</i>	3
1.2.3 <i>Formulae – The Measure to Cure Diseases</i>	4
1.3 DATA MINING AND MACHINE LEARNING	7
1.3.1 <i>Data Mining: A Stage to Discover Knowledge from Database</i>	7
1.3.2 <i>Data Preparation for Data Mining</i>	8
1.3.2.1 <i>Formulae: Enough Data for Data Mining</i>	8
1.3.2.2 <i>Classic Chinese Formulae: Reliable Data for Data Mining</i>	9
1.3.2.3 <i>Tonic Formulae: Right Data for Data Mining</i>	10
1.3.3 <i>Machine Learning</i>	13
1.3.3.1 <i>The concept of learning</i>	13
1.3.3.2 <i>Classification of learning</i>	15
1.4 NEURAL NETWORKS: A DATA MINING ENGINE	16
1.4.1 <i>Attribute Representation in Formulae Study Suits for Neural Networks</i>	17
1.4.2 <i>Nonlinear Features in Formulae Study Suits for Neural Networks</i>	17
1.5 OBJECTIVES OF THE RESEARCH	19
CHAPTER TWO METHODOLOGY.....	20
2.1 THE KNOWLEDGE DISCOVERY PROCESS.....	20
2.2 DATA SELECTION	22
2.3 DATA REPRESENTATION	24
2.3.1 <i>Herb Effect Representation</i>	24
2.3.2 <i>Representation for the Dosages of Formulae in Daily Use</i>	29

2.4 ENRICHMENT	30
2.5 CODING.....	57
2.5.1 Coding.....	57
2.5.2 Input Coding	57
2.6 DATA MINING	60
2.6.1 Formulae Collection	61
2.6.2 Relative doses of Herbs as Input and Herb Effects as Output	61
2.6.3 Divide the formulae into two disjoint sets: training set and testing set	67
2.6.3.1 Any formula with only one herb appearance in 87 formulae have to be put into the training set...	67
2.6.3.2 Any formula with only one herb effect appearance in 87 formulae have to put into the training set	68
2.7 USE THE LEARNING ALGORITHM WITH THE TRAINING SET AS EXAMPLES TO GENERATE A HYPOTHESIS H	70
2.7.1 A Single Layer Neural Network	70
2.7.1.1 Architecture of Adaptive Linear Neuron networks.....	70
2.7.2.1 Least Mean Square Error Algorithm	72
2.7.2 Multilayer Neural Networks.....	75
2.7.2.1 Backpropagation Architecture	75
2.7.2.2 Backpropagation Algorithm	78
CHAPTER THREE RESULT AND DISCUSSION	81
3.1 STABLE AND RELIABLE TRAINING SET	81
3.2 FIND COMPATIBILITY FROM WEIGHT MATRIX.....	83
3.2.1 The Single Remedy Effect.....	85
3.2.2 Synergism and Assisting Compatibility.....	87
3.3 PREDICT THE FORMULAE EFFECT IN TESTING FORMULAE SET	90
3.4 CONCLUSION	96
3.4.1 Neural Network Result Match Theory of Chinese Medicine.....	96
3.4.1.1 ADALINE Network Suits for Compatibility Study.....	96
3.4.1.2 Backpropagation Neural Network Predict well	97
3.4.2 Artificial Neural Network Application Need Design.....	97
3.4.2.1 Training Epoch Increase with More Layer Used	97
3.4.2.2 The number of nodes in hidden layer of backpropagation neural network.....	99
3.4.2.3 Selecting Suitable Transfer Function to Adapt 79 Formulae Training Model	102
3.4.2.4 The Results Reflect the Reliability of Prediction.....	104
3.5 FUTURE STUDY.....	104
APPENDICES	105
APPENDIX 1	105
Translation for input neurons.....	105
APPENDIX 2	107
Translation for output	107
REFERENCES	110
CURRICULUM VITAE	113