

On the Classification of Cancer Cell Gene via Expressive Value Distance (EVD)
Algorithm and Its Comparison to the Optimally Trained ANN Method

by

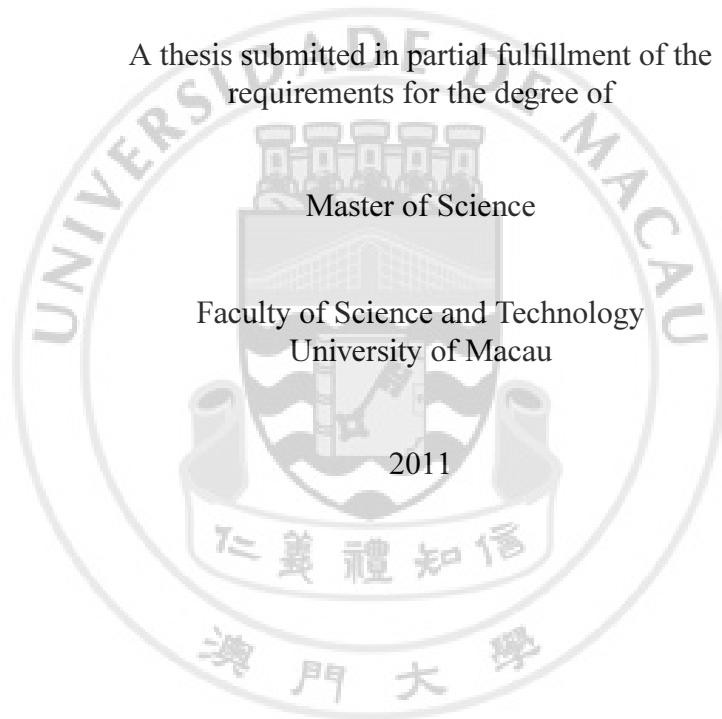
Zhang Tong

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science

Faculty of Science and Technology
University of Macau

2011



Approved by _____

Supervisor

Date _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address: Faculty of Science and Technology,
University of Macau,
Av. Padre Tomas Pereira, S.J. Taipa,
Macao, China.

Telephone: +853 6653 8961

Fax: +853 2883 8314

E-mail: ztony86@gmail.com, ma96530@umac.mo

Signature _____

Date _____



University of Macau

Abstract

On the Classification of Cancer Cell Gene via Expressive Value Distance (EVD)
Algorithm and Its Comparison to the Optimally Trained ANN Method

by Zhang Tong

Thesis Supervisor: Tam Sik Chung
Faculty of Science and Technology

In recent years, cancer can be detected and recognized by analyzing the sample's expression profile. The cancer gene expression data are high dimensional, high variable dependent, and very noisy. The dimension reduction method is often used for processing the high dimensional data. In this study, a new statistical dimension reduction method called Expressive Value Distance (EVD) is developed and proposed for the practical high-dimensional gene expression cancer data. The feature genes data extracted by EVD are arranged for training the optimally trained Artificial Neural Network (ANN). The trained ANN is then used to classify whether the unseen gene data is cancer or not. In comparison of ANN classification with and without EVD, it is found that both of the ANN can classify the cancer data in good accuracy. With the EVD method, the great amount of data (2000 genes) can be effectively reduced to 16 genes. Therefore, EVD is an effective dimension reduction method. Even the EVD method is not used, the optimally trained ANN is also an advanced method for classifying the high dimensional and complicated cancer data. In other words, it proves that optimally trained ANN is a very robust classification technique.

Keywords-Gene expression profile; Dimension reduction; Artificial neural network; Expressive value distance; Classification of cancer cell gene;

TABLE OF CONTENTS

LIST OF FIGURES	ii
LIST OF TABLES	iii
PREFACE	iv
CHAPTER 1: Introduction	1
1.1 Dimension Reduction	1
1.2 Review of Dimension Reduction methods	2
1.2.1 Principal Component Analysis	3
1.2.2 Kernel PCA	5
1.2.3 Factor Analysis	6
1.2.4 Maximum Variance Unfolding	7
CHAPTER 2: Classification Methods	8
2.1 Classification Methods	8
2.2 Artificial Neural Networks	8
2.3 Optimally Trained Artificial Neural Networks	9
CHAPTER 3: Expressive Value Distance (EVD)	13
3.1 New Dimension Reduction method	13
3.1.1 Vertical process	13
3.1.2 Horizontal process	14
3.2 Algorithm of EVD	15
CHAPTER 4: Experimental Result	17
4.1 Data	17
4.2 Experimental Results	21
CHAPTER 5: Conclusion	23
BIBLIOGRAPHY	24