FACULTY OF COMPUTER
& INFORMATION SCIENCES
AIN SHAMS UNIVERSITY
Abbassia, Cairo, Egypt

KNOWLEDGE DISCOVERY WITH ARTIFICIAL NEURAL NETWORKS

A Thesis submitted to the Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University

In partial fulfillment of the requirements for the degree of Master in Computer and Information Sciences

BY

Ayad Fekry Ayad

B.Sc. in Computer Science
Demonstrator, Computer Science Department
Faculty of Computer and Information Sciences
Ain Shams University

Under the Supervision of

Prof. Dr. Abdel-Badeeh Mohamed Salem

Prof. Of Computer Science, Computer Science Department,
Faculty of Computer and Information Sciences
Ain Shams University

Prof. Dr. Mostafa Mahmoud Syiam

Professor and Head of Computer Science Department,
Vice Dean for Student affairs
Faculty of Computer and Information Sciences
Ain Shams University

2004

Acknowledgments

My utmost thanks to ALLAH for all his gifts in all my life.

My gratitude should go to Prof. Mohammed Said Abdel-Wahab, Dean of the faculty of Computer and Information Sciences, for his valuable support.

My greatest thanks go to Prof. Abdel-Badeeh Salem and Prof. Mostafa Syiam for their teaching, supporting, guiding and encouragement they provide me throughout my work. Their comments and suggestions helped me not only in this thesis, but also in my career. They have suffered a lot with me since it is my first step in research studies. They taught me how to read and write a paper and how to conduct experiments and analyze the results.

I would like to thank my fathers Kerolous and Marqus for helping and supporting me to complete this thesis.

Special Thanks to Mr. Tamer Mustafa who helped me during writing this thesis.

Thanks are not enough to be given to my parents. Rather, there are no words that are valuable to equalize their love and support for me. They are behind any success in my life.

Table of Contents

List Of Figures

List Of Tables

List of Publications

1. Abdel-Badeeh M. Salem, Mostafa M. Syiam, and Ayad F. Ayad ***"Improving Self-Organizing Feature Map (SOFM) Training Algorithm Using K-means initialization***" In Proc. 5th International Conference on Enterprise Information Systems ICEIS, vol.1, 2003, pp.399-405, France.

   Published also in Proc. 7th IEEE International Conference On Intelligent Engineering Systems (INES), vol.40, 2003, pp.41-46. Egypt.

2. Abdel-Badeeh M. Salem, Mostafa M. Syiam, and Ayad F. Ayad ***"A Hybrid Dynamic Self-Organizing Map For Clustering Of Document Collections"*** In Proc. 2nd WSEAS Intern. Conf. on ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING, and DATA BASES (AIKED 2003), vol.2, 2003, pp.201-206, Greece.

3. Abdel-Badeeh M. Salem, Mostafa M. Syiam, and Ayad F. Ayad ***"Unsupervised Artificial Neural Networks For Clustering Of Document Collections"*** The 17th International FLAIRS Conference, 2004, USA.

   Accepted also for publication in 6th International Conference on Enterprise Information Systems ICEIS, 2004, France.

Abstract

Data mining is a part of a large area of recent research in artificial intelligence and information management. The purpose is to find new knowledge from databases where the dimensionality, complexity, or amount of data is prohibitively large for manual analysis. A large data set may be of very high dimensionality and consists of complex structure that even the most well planned data mining techniques might have difficulty extracting meaningful patterns from it. An unsupervised technique such as cluster detection becomes useful in such situations.

Clustering is a useful tool when it is used to deal with a large complex data set with many variables and unknown internal structure. In such situations, clustering would be the best tool to obtain an initial understanding of the structure inherent in the data. Once automatic cluster detection has discovered regions of the data space that contains similar records, other data mining tools and techniques could be used to discover rules and patterns within the clusters.

The Self-Organizing Map (SOM) has been used as a tool for mapping high-dimensional data into a one- (or two-) dimensional feature map. It is then possible to visually identify the clusters from the map. The main advantage of such a mapping is that it would be possible to gain some idea of the structure of the data by observing the map, due to topology preserving nature of the SOM.

Usually, SOM can be initialized using random values for the weight vectors. In this thesis we present a different approach for initializing SOM. This approach depends on the K-means algorithm as an initialization step for SOM.