UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A Local Network Neighbourhood Artificial Immune System

by

Alexander J. Graaff

Submitted in partial fulfillment of the requirements for the degree

Philosophiae Doctor

in the Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

Pretoria, South Africa

June 3, 2011

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A Local Network Neighbourhood Artificial Immune System

by

Alexander J. Graaff

Abstract

As information is becoming more available online and will forevermore be part of any business, the true value of the large amounts of stored data is in the discovery of hidden and unknown relations and connections or traits in the data. The acquisition of these hidden relations can influence strategic decisions which have an impact on the success of a business. Data clustering is one of many methods to partition data into different groups in such a way that data patterns within the same group share some common trait compared to patterns across different groups. This thesis proposes a new artificial immune model for the problem of data clustering. The new model is inspired by the network theory of immunology and differs from its network based predecessor models in its formation of artificial lymphocyte networks. The proposed model is first applied to data clustering problems in stationary environments. Two different techniques are then proposed which enhances the proposed artificial immune model to dynamically determine the number of clusters in a data set with minimal to no user interference. A technique to generate synthetic data sets for data clustering of non-stationary environments is then proposed. Lastly, the original proposed artificial immune model and the enhanced version to dynamically determine the number of clusters are then applied to generated synthetic non-stationary data clustering problems. The influence of the parameters on the clustering performance is investigated for all versions of the proposed artificial immune model and supported by empirical results and statistical hypothesis tests.

Keywords: Data Clustering, Artificial Lymphocytes, Affinity Maturation, Clonal Selection, Somatic Hyper Mutation, Artificial Immune Networks, Immune Network Topologies, Clustering Performance Measures, Dynamic Clustering, Non-stationary Data Clustering.

Thesis supervisor: Prof. A.P. Engelbrecht
Department of Computer Science
Degree: Philosophiae Doctor

A Local Network Neighbourhood Artificial Immune System

deur

Alexander J. Graaff

Opsomming

Soos wat inligting meer aanlyn toeganglik raak en vir altyd meer deel vorm van enige besigheid, is die eintlike waarde van groot hoeveelhede data in die ontdekking van verskuilde en onbekende verwantskappe en konneksies of eienskappe in die data. Die verkryging van sulke verskuilde verwantskappe kan die strategiese besluitneming van 'n besigheid beinvloed, wat weer 'n impak het op die sukses van 'n besigheid. Data groepering is een van baie metodes om data op so 'n manier te groepeer dat data patrone wat deel vorm van dieselfde groep 'n gemeenskaplike eienskap deel in vergelyking met patrone wat verspreid is in ander groepe. Hierdie tesis stel 'n nuwe kunsmatige immuun model voor vir die probleem van data groepering. Die nuwe model is geinspireer deur die netwerk teorie in immunologie en verskil van vorige netwerk gebaseerde modelle deur die model se formasie van kunsmatige limfosiet netwerke. Die voorgestelde model word eers toegepas op data groeperingsprobleme in statiese omgewings. Twee verskillende tegnieke word dan voorgestel wat die voorgestelde kunsmatige immuun model op so 'n manier verbeter dat die model die aantal groepe in 'n data stel dinamies kan bepaal met minimum tot geen gebruiker invloed. 'n Tegniek om kunsmatige data stelle te genereer vir data groepering in dinamiese omgewings word dan voorgestel. Laastens word die oorspronklik voorgestelde model sowel as die verbeterde model wat dinamies die aantal groepe in 'n data stel kan bepaal toegepas op kunsmatig genereerde dinamiese data groeperingsprobleme. Die invloed van die parameters op die groepering prestasie is ondersoek vir alle weergawes van die voorgestelde kunsmatige immuun model en word toegelig deur empiriese resultate en statistiese hipotese toetse.

Sleutelwoorde: Data Groepering, Kunsmatige Limfosiete, Affiniteit Volwassewording, Klonale Seleksie, Somatiese Hiper Mutasie, Kunsmatige Immuun Netwerke, Immuun Netwerk Topologiee, Groepering Prestasie Maatreels, Dinamiese Groepering, Groepering van Dinamiese Data.

Tesis studieleier: Prof. A.P. Engelbrecht
Departement Rekenaarwetenskap
Graad: Philosophiae Doctor

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

– Soli Deo Gloria –

"To God Alone the Glory"

Acknowledgments

My sincere gratitude to God for all the privileges in my life and the opportunity to undertake a research study such as this. The research study, especially the research on the natural process of immunology, gave me an intense appreciation and admiration for the detail that God has put into His creation. All glory to the Creator that I can be part of His creation and that He enables me to understand a diminutive part of it.

A warm thanks to Professor Andries Engelbrecht for his continuous support, guidance and patience. Prof. Engelbrecht was always willing to listen and assist in refining my proposed ideas and methods (even the most bizarre ideas).

Also, thanks to Mr. Nelis Franken for all the brainstorming sessions on statistical hypothesis testing and probability distributions.

Last but not least, I want to thank my family and my wife, Daleen, for their support during this study. You are always there to give balance to my life.

Contents

List of Figures

List of Tables

List of Algorithms