Degree of Master of Philosophy

Investigation and Development of Fuzzy Logic Based Analytics for Data Warehousing

PARANA PALLIYA GURUGE DINESH ASANKA

158062B

Supervisor(s):

Dr. Amal Shehan Perera

2021-02-19

Department of Computer Science and Engineering

Faculty of Engineering

UNIVERSITY OF MORATUWA - SRI LANKA

Declaration

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Further, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: …… Date:

PPG Dinesh Asanka

The above candidate has carried out research for the MPhil Dissertation under my supervision.

Signature of the Supervisor:…….. Date:

Dr. Amal Shehan Perera

Acknowledgments

First, I express my sincere gratitude to my supervisor Dr. Amal Shehan Perera for guiding me throughout the project. Despite his busy schedule, he has given his fullest corporation whenever I wanted his assistance. I would also like to extend my appreciation to the panel members Dr. Dilum Bandara and Dr. Lochandaka Ranathunga for their valuable advice given to me throughout this project lasting more than four years. In addition, I am thankful to Prof. Sanath Jayasena for the valuable ideas given at the initial research stage. These ideas helped me to a great extent.

My loving parents and my wife always encouraged me towards higher studies. Especially, my wife and my two children made many sacrifices. I admit words cannot describe the sacrifices made. Anyhow, I would be failing in my duty if I leave them out of the honors.

This research is on data warehouse and I thank the three companies that contributed with their experience in data warehousing. The first data warehouse where I got the necessary experience was Ocean Software. There, I had an opportunity to work in the UK as well and got the most experience. At Pearson, I was able to implement some of the concepts that I learned importantly, administration and the configuration aspects of the data warehouse. At Virtusa, I was able to work on end-to-end aspects of data warehouse starting from the requirement gathering to deployment including the non-functional requirements of data warehouse. I deeply appreciate the kindness shown and thank all the co-workers and the management of these organizations.

Jagath Robotics (Pvt) Ltd provided me the data set which was used in this research as a proof of concept. Data gathering is a challenging task in any research. Collecting data from a reputed source helped me to proceed with the research.

I was awarded the Microsoft Most Valuable Professional by Microsoft. This allows me to gain experience with the Microsoft BI platform. The experience helped me a lot for this research. Hence, I thank Microsoft for awarding me the MVP award for 11 consecutive years.

I was fortunate enough for opportunities to lecture on data warehouse at different universities and institutions in Sri Lanka like the Sri Lanka Institute of Information Technology (SLIIT), Institute of Information Technology (IIT), Kothalawala Defence University (KDU), University of Ruhuna, University of Kelaniya, and University of UVA Wellassa. These institutions

provided me with the opportunity to lecture and these lectures helped enable me with the necessary knowledge in the area of the data warehouse. I also had the opportunity to supervise many data warehousing projects at MSc and BSc levels. This was a great opportunity. I was able to present data warehouse related sessions at places such as SQL Saturday, Colombo Big Data User Group, Sri Lanka Association of Artificial Intelligence (SLAAI), SQL Server Sri Lanka User Group (SSSLUG), Sri Lanka Azure Road Show, Colombo Dev Day, Malaysia TechInsights which allowed me to gain and present knowledge with data warehouse. I thank the organizers of these user groups for inviting me to these events so that I was able to gain knowledge in the data warehousing.

I pay my gratitude to all the teachers who taught me starting from my first school to B.Sc, MBA and MSc degrees. They gave me a sound base for many technical concepts and research methodologies.

Abstract

Data warehouse is a widely used technology that provides the employees who take strategic decisions within an enterprise with access to any level of required data. Historically, data warehouses were built on crisp values with a key assumption that one attribute value falls into one nominal value. Fuzzy Logic can be built into the data warehouse by treating the dimension value as weightages of different labels. However, in most of the attempts to implement a fuzzy data warehouse, they were limited and non-comprehensive in the implementation when considering end to end aspects of the data warehouse.

Using fuzzy techniques, it is possible to represent fuzzy conceptual information in the original domain, that would lead to better analysis. In this research, different types of fuzzy membership functions are defined using different techniques and data warehouse facts and dimensions are designed accordingly. There can be multiple fuzzy functions for one dimension as well as for one fact table depending on the business domain. Apart from defining fuzzy membership function using data-driven methods, there are other approaches of defining fuzzy membership functions such as a derived method where multiple fuzzy memberships are combined to define several fuzzy membership functions. In the literature reviewed, concepts like ETL and OLAP cube were found to be discussed in a limited manner. Non-function techniques are also identified and addressed in the means of validation, configuration, performance, security, scalability in order to make better usability of the fuzzy data warehouse.

The scope of this research revolves around end-to-end features of fuzzy data warehousing starting from data extraction and transformation to data warehouse modeling. Implementing a fuzzy data warehouse, helps to enable users with better analyses. To verify whether the proposed fuzzy data warehouse can be applied, a feasibility study is carried out for the domains in which fuzzy data warehousing can be implemented. Concepts related to the outcome from this research are verified with the use of a Sri Lankan plantation data set for four years. The results show that concepts introduced by this research can be implemented in realistic scenarios.

Key Words: Data Warehouse, Fuzzy Logic, Fuzzy Membership Function, ETL, OLAP

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abbreviations

OLAP	Online Analytical Processing	NFR	Non-Functional Requirements
OLTP	Online Transactional Processing	ANN	Artificial Neural Network
ETL	Extract-Transform-Load	HMF	Horizontal Membership Function
BI	Business Intelligence	MF	Membership Function
MIS	Management of Information Systems	DBMS	Database Management System
CRM	Customer Relationship Management	SCD	Slowly Changing Dimension
HR	Human Resource	CDC	Change Data Capture
ERP	Enterprise Resource Planning	MDX	Multidimensional Expressions
KPI	Key Performance Indicators	MOLAP	Multidimensional Online Analytical Processing
FK	Foreign Key	ROLAP	Relational Online Analytical Processing
PK	Primary Key	HOLAP	Hybrid Online Analytical Processing

SVM	Support Vector Machine	MDDB	Multi-Dimensional Database
FD	Fuzzy Dimension	FC	Fuzzy Cube
AI	Artificial Intelligence	BI	Business Intelligence
ML	Machine Learning	RFI	Recency – Frequency - Intensity
RFM	Recency – Frequency - Monetary	SSIS	SQL Server Integration Services
SSAS	SQL Server Analysis Service	FIS	Fuzzy Inference System
DOM	Degree of Membership	IQR	Inter-Quartile Range
FIS	Fuzzy Inference System		