

# **Context-aware Recommendation for Data Visualization**

L. A. D. P. Athukorala

(178053R)

Degree of Master of Science (Research)

Department of Computer Science And Engineering

University of Moratuwa

Sri Lanka

June 2019

# **Context-aware Recommendation for Data Visualization**

L. A. D. P. Athukorala

(178053R)

Thesis submitted in partial fulfillment of the requirements for the  
degree of Master of Science (Research) in Computer Science and Engineering

Department of Computer Science And Engineering

University of Moratuwa

Sri Lanka

June 2019

## Declaration

I declare that this is my own research Thesis and this Thesis does not incorporate without acknowledgement any material previously published submitted for a Degree or Diploma 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.

Signature: .....

Date:

L.A.D.P.Athukorala

I have read the Thesis and it is in accordance with the approved university Thesis outline.

Signature of the supervisor:.....

Date:

Dr. D. Meedeniya

Signature of the supervisor:.....

Date:

Dr. G. I. U. S. Perera

## **Abstract**

Today projects with data analysis play a significant role to give us suggestions to our daily problems. While understanding those analysis data user needs to get the meaning and the nature of the data. Data visualization is the best option to observe the data. The human eye can easily analyze those data with the help of visualization. Moreover when visualizing a dataset better to have an understanding of data types and user intention or preferences. Recommendation systems are the best approach to address the above problem. In this ,study we discuss recommendation application which gets the help of machine leaning and mapping algorithm. Context awareness is a help while giving recommendations to chart types. Even though from users perspective suggestions can be changed. Therefore the proposed solution improves with the help of user’s feedbacks. For each test-run system is collecting user feedbacks and use them to improve the training dataset. At the initial stage, there are only a few training data. Users can interact with the system and based on their feedbacks the outcome of the system will get more accurate. Based on user feedbacks recommendation will get more reliable in the long-run. In this study, we are looking at how much accuracy it has in the initial stage and how it varies with the number of test runs in the system. Therefore user interaction plays a significant role to help recommendations. Feedbacks from users help when improving the recommendations. The System recommendations are provided using a combined method of machine learning and rule based components and the evaluation has shown an accuracy over 80%. As this is a trending research area, contribution made through this study can be useful for the industry and the research community.

## **Keywords:**

Human-centered computing, Recommender systems, Content awareness, Data Visualization, Information systems applications.

## *Acknowledgements*

I would never have been able to finish my dissertation without the guidance, support and encouragement of numerous people including my mentors, my friends, colleagues and support from my family. At the end of my thesis I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Dulani Meedeniya and Dr. Indika Perera, for the continuous support given for the success of this research both in unseen and unconcealed ways. This would not have been a success without your tremendous mentorship and advice from the beginning. Your wide knowledge and logical way of thinking have been of great source of inspiration for me. You have always extended his helping hands in solving research problems. The in-depth discussions, scholarly supervision and constructive suggestions received from you have broadened my knowledge. I strongly believe that without your guidance, the present work could have not reached this stage.

**Thank you!**

# TABLE OF CONTENTS

Chapter 1	Introduction.....	1
1.1	Overview.....	1
1.2	Motivation.....	2
1.3	Problem Statement.....	2
1.4	Objectives.....	3
1.5	Research Question.....	3
1.6	Expected outcome.....	4
1.7	Scope.....	4
Chapter 2	Literature Review.....	5
2.1	Introduction.....	5
2.2	Data sets.....	5
2.2.1	Point Data Sets.....	5
2.2.2	Aerial Data Set.....	6
2.2.3	Discrete Data.....	7
2.2.4	Continuous Data.....	7
2.2.5	Example Data sets.....	7
2.3	Data Pre-Processing.....	7
2.3.1	Data Cleaning.....	8
2.3.2	Data Transformation.....	9
2.3.3	Data Reduction.....	10
2.4	Recommender systems.....	10
2.4.1	Importance of Recommender Systems.....	11
2.4.2	Content awareness.....	11
2.4.3	Recommendation Algorithms.....	11
2.4.3.1	Content-based filtering.....	12
2.4.3.2	Collaborative filtering.....	12

2.4.3.3	Demographic.....	13
2.4.3.4	Knowledge-based.....	14
2.4.3.5	Community-based.....	14
2.4.3.6	Cluster Models.....	14
2.4.3.7	Rating method.....	15
2.4.3.8	Hybrid systems.....	15
2.4.4	Existing recommender systems.....	16
2.4.5	Related work in Recommender systems.....	19
2.5	Machine Learning for Interactive Systems.....	20
2.5.1	Supervised learning.....	20
2.5.2	Unsupervised learning.....	21
2.5.3	Semi-supervised learning.....	22
2.5.4	Related Work on interactive machine learning.....	23
2.6	Information visualization.....	25
2.6.1	Information structures.....	25
2.6.1.1	Tabular structures.....	25
2.6.2	Data Visualization strategies.....	28
2.6.3	Related work on data visualization.....	36
2.7	Existing tools on data visualization and recommender systems.....	38
2.7.1	Data recommender tools.....	38
2.7.2	Data visualization tools.....	39
2.7.3	Machine learning support tools.....	40
2.8	Evaluation of visualizing data in recommender systems.....	41
2.8.1	Comparison of related work on visualizing and recommender data.....	41
2.8.2	Comparison of recommender algorithms.....	42
2.8.3	Comparison of data visualization techniques.....	44
2.9	Evaluation of related work.....	45
2.10	Quality attributes.....	46

2.10.1	Accuracy .....	46
2.10.2	Performance .....	46
2.10.3	Reliability.....	47
2.10.4	Usability.....	47
2.10.5	Scalability .....	47
Chapter 3	Research Methodology .....	48
3.1	Introduction.....	48
3.2	System Design .....	48
3.2.1	The process of the system .....	48
3.2.2	Architecture design .....	49
3.3	Pre-processing.....	50
3.3.1	Duplicate tuple removal .....	50
3.3.2	Normalization.....	51
3.3.3	Filling missing values .....	51
3.3.4	Outlier removal .....	52
3.4	Context identifier .....	52
3.5	Rule-based Component.....	55
3.5.1	Why rule-based .....	56
3.5.2	User Intention.....	57
3.5.3	Select a chart based on user's intention .....	58
3.5.4	Steps of process.....	58
3.6	Machine Learning Component.....	63
3.6.1	Why machine learning .....	63
3.6.2	Decision tree based approach.....	64
3.6.3	Flow of module .....	67
3.7	Feedback component .....	69
3.8	Integration of recommendations .....	73
3.9	Visualization Component.....	75

3.9.1	Visualization process .....	76
3.9.2	Designing Visualizations .....	77
3.10	Feedback Validation .....	81
Chapter 4	Implementation .....	83
4.1	Introduction.....	83
4.2	Machine Learning module .....	83
4.2.1	Setup the module.....	83
4.2.2	Module Pseudocode code.....	84
4.3	Rule-based approach.....	84
4.3.1	The theory behind the Rule-base representation.....	84
4.4	Integrating recommendation .....	90
4.5	Column grouping .....	91
4.6	Visualization component.....	91
4.6.1	Chart option provider .....	91
4.6.2	Chart visualizer .....	92
4.6.3	Pseudocode for visualization .....	92
4.7	Backend admin controllers.....	92
4.7.1	Chart adder.....	92
4.7.2	Rule adder .....	93
4.7.3	Dataset viewer.....	93
Chapter 5	Experiments and Results.....	95
5.1	Usability study .....	95
5.2	The accuracy of the system.....	96
5.3	Time consumption .....	101
5.4	Memory and CPU usage .....	102
Chapter 6	Conclusions.....	104
References.....		105