



Developing a Personalized Multi-Dimensional Bank Analysis System using Business Intelligence

Mohamed Galal El Saeed Rabiea

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 of Science in Computer and Information Sciences

Supervised by:

Prof. Dr. Mostafa Aref

Professor of Computer Science Department of Computer Science

Faculty of Computer and Information Sciences

Ain Shams University

Dr. Ghada Nasr Aly Hassan

Faculty of Computer and Information Sciences

Ain Shams University

2017

*To my parent's soul, my wife, Radwa, my brother, Ahmed, my sisters
Shaimaa and Lamiaa, and my friends.*

Acknowledgements

Thanks to Allah for giving me this opportunity, the strength and the patience to complete my master, after all the challenges and difficulties.

I would like to thank Prof. Mostafa Aref and Dr. Ghada Hassan for their help and guidance.

My sincere appreciation and gratitude to them for all effort, and the time that they have devoted to support and assist me throughout this thesis creation.

Abstract

Intelligent techniques have been used in the various business domains to improve analysis process, increase revenues and save time. In customer-centric institutions, one of the areas in which intelligent techniques and data mining algorithms have been used is the personalization for enhanced Customer Relationship Management (CRM) performance. However, the expanding of customer base, the diversity of products, the complex behavior of customer groups lead to facing big challenges in producing tailored actions to customer needs.

After applying different techniques to enhance the concept of personalization in the financial institutions, it was found that there are still shortages in customer satisfaction level for reaching the tailored action that satisfy customer needs. Also there are difficulties in applying the personalization approaches in the institutions with a large number of customers using the traditional approaches. Applying the personalization using similarity-based approach from a prospective of customer profile or customer behavior has a limited scope of actions. The limitation of developing a single framework to serve specific customer dimension is considered a very limited approach in framework re-usability.

In this work, we developed a multi dimension personalization framework (MDPF) architecture to improve personalized targeting. The framework presented improves on the automation of existing systems by using supervised and unsupervised data mining techniques, and enhances the level of targeting by considering more effective dimensions in multiple stages of the framework. This framework overcomes the shortage of old approaches in targeting the customers' needs. The developed framework is a hy-

brid approach combines the data-driven (Knowledge Discovery) approaches with the business-driven (Top-Down) approaches to generate a new business competitive opportunities through a complete understandable and profiled customer segments. The framework has been built based on the concept of framework dimensionality which leads to an enhancement in the project agility. Enhancing the agility process leads to adapting to any new changes in the framework quickly without rebuilding the framework from scratch.

Contents

1	INTRODUCTION	13
1.1	Research Objectives	14
1.2	Thesis Contributions	15
1.3	Thesis Outline	17
2	BACKGROUND AND RELATED WORK	19
2.1	The Technical Personalization process	21
2.2	Measures of the personalization Process	23
2.2.1	Controlling the Decision-Making Cycle	23
2.2.2	Automating the Personalization Cycle	24
2.2.3	Targeting Efficiency	24
2.2.4	The Efficiency of the Process Cost	24
2.2.5	Actions Scope	25
2.2.6	Customer Base Coverage	25
2.3	The Personalization Approaches	25
2.3.1	Profile-Based Approach (Computational)	26
2.3.2	Neighborhood Approach (Similarity-Based patterns)	26
2.3.3	Business-Driven Approach (Top-down)	28
2.3.4	Customer Is a Boss Approach (Customization Approach)	29
2.3.5	Knowledge Discovery Approach (Bottom-Up)	30
2.4	The Analysis Dimensions of Personalization	31
2.4.1	Socio-Demographics Segmentation Dimension	31
2.4.2	Behavioral Segmentation Dimension	31
2.4.3	Value-Based Segmentation Dimension	32

2.4.4	Time Series Analysis Dimension	33
2.4.5	Life-Stages Segmentation	33
2.5	Business Objectives of Personalization Process	33
2.5.1	Personalization for Forecasting	34
2.5.2	Personalization for Recommendation	35
2.5.3	Personalization for Decision Support and Planning	38
2.5.4	Personalization for Improving Loyalty	39
2.5.5	Personalization for Exploration and Discovery	41
2.6	Business Intelligence Life cycle	41
2.6.1	Top-Down Approach (Business Driven)	42
2.6.2	Bottom-Up Approach (Data Driven)	44
2.7	Data Modeling	46
2.7.1	Normalized Entity Relationship Modeling	46
2.7.2	Dimensional Data Modeling	47
2.8	Data Visualization	48
3	MULTI-DIMENSIONAL PERSONALIZATION FRAMEWORK DESIGN	51
3.1	Multi-Dimensional Personalization Framework in Banking	52
3.1.1	Dimensions in the MDPF	53
3.1.2	The MDPF Technical Architecture	55
3.2	MDPF Data Flow Design	60
3.3	Project Analysis Hybrid Approach	61
3.4	Case Study: Applying the MDPF for Credit Cards Holders	61
3.5	Credit Cards Dataset Description	62
3.6	Case Study Design	64
3.6.1	Classification for Auto Profiling	68
3.6.2	Business Plans	69
4	MULTI-DIMENSIONAL PERSONALIZATION FRAMEWORK IMPLEMENTATION	71
4.1	Tools Used in Applying the Framework	72

4.2	Data Preparation for Framework	72
4.2.1	Data Identification	72
4.2.2	Date Modeling	73
4.2.3	Data Transformation(Segmentation Fields)	76
4.3	The MDPF Implementation	78
4.3.1	Applying the Data Reduction Technique	78
4.3.2	The Clustering Step	81
4.3.3	Groups' Profiling	85
4.3.4	Associate the Segments to the Time Dimension	87
4.3.5	Golden Groups Profiling	88
4.4	MDPF Evaluation	91
5	CONCLUSION AND FUTURE WORK	99
5.1	Summary of Research	99
5.2	Discussion and Interpretation of Findings	100
5.3	Significance of the Study	101
5.3.1	The Personalization Enhancement	102
5.3.2	The Business Contributions	103
5.4	Recommendations for Further Research	103

List of Figures

2.1	Personalization Perspectives	21
2.2	Personalization Implementation Process	23
2.3	Life-Stage segmentation example in Banking	34
2.4	Kimball BI Lifecycle	42
2.5	Top-Down Analysis Approach	44
2.6	Bottom-Up Analysis Approach	46
2.7	Snowflake schema or normalized modeling	47
2.8	Star Dimensional Modeling Schema	48
2.9	Data Visualizaion	49
3.1	Multi-Dimensions used for segmentation	53
3.2	value-based Segmentation [28]	54
3.3	Multi-Dimension Personalization Framework	56
3.4	Two Steps Clustering algorithm [25]	58
3.5	Mixing data mining techniques and market research	59
3.6	MDPF Data Flow	60
3.7	Hybrid analysis approach for analysis	62
4.1	Merchants Classification	75
4.2	MDPF Framework	78
4.3	Clustering Stream in the Modeler	79
4.4	The silhouette measure of the clustering solution	82
4.5	Clusters Distribution	83
4.6	Cluster centers	84
4.7	Segments Profiling Table	86

4.8	Groups distribution over seasons	89
4.9	Family Men over seasons	90
4.10	Travellers over Summer season	91
4.11	oldies over Winter season	92
4.12	Fashion Ladies over Summer season	93
4.13	Fashion Ladies over Spring season	94
4.14	Fun Lovers over Summer season	95
4.15	Summer Travelholics	96
4.16	Personalization versus Waste for normal and golden groups	96
4.17	Dimensional Framework	97
4.18	Pipeline Framework	97

List of Tables

3.1	Customer Demographics	63
3.2	Account Information	63
3.3	Transactions Information	63
3.4	Accounts Association to each input KPI	65
3.5	Pairwise correlation coefficients among the original input fields	65
3.6	Varimax rotation (Total variance explained)	65
3.7	the rotated component matrix	66
3.8	Interpretation of the extracted principal components	66
3.9	Demographical profile for clusters	67
3.10	Segments to time series association	68
3.11	Demographical profile for each cluster	69
4.1	The tools used in the Framework	72
4.2	Operational System Fields	74
4.3	Customer Demographics	74
4.4	Accounts Information	74
4.5	Transactions Information	75
4.6	Merchants Classification	75
4.7	The used KPIs Fields for credit card segmentation	76
4.8	Clustering Fields sample	77
4.9	Varimax rotation	79
4.10	rotated component matrix	80
4.11	Mean of purchases percentage for each cluster by merchant type	83
4.12	Average purchase amount for each cluster by merchant type	84

4.13 Usage KPIs by cluster	85
4.14 The demographical profile of each cluster	85
4.15 Segments to time series association	87
4.16 Segments to time series classified by season	88
4.17 Segments to Season association	88
4.18 personalization for normal versus golden groups	92
4.19 MDPF Evaluation factors	94

List of Publications

- [1] Mohamed Galal, Ghada Nasr and Mostafa Aref. "A Survey on Personalization for Improved Intelligent CRM." *In Proceedings of First International Conference "Business in a dynamic world", January 2016, Limassol, Cyprus.*
- [2] Mohamed Galal, Ghada Hassan, and Mostafa Aref. Developing a personalized multi-dimensional framework using business intelligence techniques in banking. In Proceedings of the 10th International Conference on Informatics and Systems, pages 21-27. ACM, 2016
- [3] Mohamed Galal, Ghada Hassan, and Mostafa Aref. Intelligent Personalization Approaches For Complex Customer Behaviour: An Overview. International Journal of Intelligent Computing and Information Sciences, Ain Shams University, August 2017.