GALATASARAY UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

ALTERNATING LEAST SQUARES MODEL RECOMMENDER SYSTEMS: AN APPLICATION ON CREDIT CARD MARKET

İlkay KÖRPE

June 2017

ALTERNATING LEAST SQUARES MODEL RECOMMENDER SYSTEMS: AN APPLICATION ON CREDIT CARD MARKET

(HAREKETLİ EN KÜÇÜK KARELER MODELİ ÖNERİCİ SİSTEMLER:

KREDİ KARTI PAZARI ÜZERİNDE BİR UYGULAMA)

by

İlkay KÖRPE

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

June 2017

This is to certify that the thesis entitled

ALTERNATING LEAST SQUARES MODEL RECOMMENDER SYSTEMS: AN APPLICATION ON CREDIT CARD MARKET

prepared by **İlkay KÖRPE** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering** at the **Galatasaray University** is approved by the

Examining Committee:

Assist. Prof. Dr. Tuncay GÜRBÜZ (Supervisor)
Department of Industrial Engineering
Galatasaray University -------------------------

Prof. Y. Esra ALBAYRAK
Department of Industrial Engineering
Galatasaray University -------------------------

Assist. Prof. Dr. Umut ASAN
Department of Industrial Engineering
İstanbul Technical University -------------------------

Date: -------------------------

ACKNOWLEDGEMENTS

The main aim of this study, in a broad sense, is proposing an effective model to analyze big datasets. To be more specific, I will be extracting hidden patterns from purchase logs, create customer and product clusters, form relations in between, forecast/recommend future transactions and propose real-life marketing applications accordingly. Studies focusing on such implicit datasets are quite uncommon in literature so I hope this one will shed light on future researches on this matter.

I want to express my gratitude to some acquaintance of mine, to whom I owe greatly: Assist. Prof. Dr. Tuncay Gürbüz, for his supervision and mentoring, Hasan Can Saral for his never ending support even in the most challenging of his time, and beloved Nazlı Kilislioğlu, my ultimate motivation, for being right by my side.

May 2017
İlkay KÖRPE

TABLE OF CONTENTS

LIST OF SYMBOLS

ALS	**:** Alternating Least Squares
BDA	**:** Big Data Analytics
B-SGD	**:** Bias-Stochastic Gradient Descent
CB	**:** Content-Based
CF	**:** Collaborative Filtering
CRM	**:** Customer Relationships Management
DBMS	**:** Database Management System
IDC	**:** International Data Corporation
IT	**:** Information Technology
LDA	**:** Latent Dirichlet Analysis
LSA	**:** Latent Semantic Analysis
MCC	**:** Merchant Category Code
MSE	**:** Mean Squared Error
PCA	**:** Principal Component Analysis
RoI	**:** Return-on-Investment
SGD	**:** Stochastic Gradient Descent
SVD	**:** Singular Value Decomposition
W-ALS	**:** Weighted Alternating Least Squares

LIST OF TABLES

ABSTRACT

Following the introduction of MapReduce and Apache Hadoop,it has been possible to process immense datasets that are beyond the capabilities of traditional database management system techniques. This created a new area of study: Big Data. Big Data is generally used to define the massive and unstructured datasets, unsuitable to process with subject traditional methods.

Growing interest on data intensified on the areas where it is available the most: e-commerce businesses, movie review sites, music player platforms to name a few, where user interaction is digital so that it can be logged and traced. Practices mainly aim analyzing user profiles, predicting preferences and making appropriate recommendations.

Though it is relatively easier to analyze feedbacks and predict preferences in these cases, where user ratings, scores, favorites or likes/dislikes are available, the bigger part of the value lies within the indirect data, as direct feedbacks are usually not in grasp. Businesses should harness any information available and build proper correlations to feed the recommendation system. In this study, credit card transaction logs will be studied to predict card holder's next transaction sector and propose marketing offers correspondingly. I hope it will shed light on future researches on recommendation systems with implicit data.

ÖZET

Yakin geçmişte, özellikle bireysel internet kullanımının yaygınlaşması ile birlikte veri üretimi, eşi benzeri görülmemiş bir hıza ulaşmıştır. Bunu takiben veri işleme algoritmalarında yaşanan gelişmeler ile birlikte veri kullanımı, asli iş kolu veri ile doğrudan ilintlili olmayan işletmelerin de odağına girmiş; üzerinde çalışılan verinin ölçeği, önceki sistemler ile mümkün olamayacak boyutlara ulaşmıştır. Bu ölçek, sadece boyut anlamında bir büyüklüğü değil; yapı, format, kaynak, doğruluk, anlamlılık açılarından çeşitliliği de kapsamaktadır. Bu yeni 'veri' kavramı, geleneksel kavramdan ayrışmış ve kendi terminolojisini yaratmıştır: 'Büyük Veri'.

Büyük Veri ile ilgili yaşanan gelişmeler, işletmelerin pazarlama aktivitelerine de yeni bir yön vermiştir. Kitlesel pazarlama, yerini giderek kişisel pazarlamaya bırakmaktadır. Her birey için;üründen beklenti, iletişim yatkınlığı, kanal tercihi ve bunun gibi pazarlama aktivitelerine yön veren farktörlerin ciddi anlamda değişkenlik gösterdiğini farkeden işletmeler pazarlama yaklaşımlarını da bu alt segmentlere göre şekillendirmeye başlamış, veri kullanımına hakim olan azınlık bir kesim ise bu işi bireye kadar özelleştirmeyi başarmıştır.

Bu çalışmada, günümüzde sanal sektörlerde kendine uygulama alanı bulan; fakat aslında daha geniş bir iş kolu yelpazesi için katma değer potansiyeli taşıyan önerici sistemler üzerinde odaklanılmıştır. Arama motorlarındaki reklam yerleştirmeleri, e-ticaret sitelerindeki 'önerilen ürünler', medya veritabanlarındaki 'bu ürünü alanlar şunları da beğendi' kısımları önerici sistemlerin akla gelen ilk örneklerindendir. Bu sistemler, geçmiş verilerin analizi ile geniş bir ürün katalogundan, kullanıcıların/tüketicilerin kullanmaya/tüketmeye meyilli olduğu veya kullanacağı/tüketeceği öngörülen ürünleri tahmin eder.

Bu tezde, önce farklı önerici sistem çeşitleri ve algoritmaları yüzeysel olarak tanıtılmakta, ardından da kredi kartı pazarında bir uygulama ile Hareketli En Küçük Kareler algoritması üzerine kurulu bir İşbirliğine Dayalı Filtreleme Modeli önerici sisteminin işleyiş detayları incelenmektedir. İlk adımda; girdi olarak kullanılan 1 milyon kredi kartı işlemi verisi kart hamili – sektör matrisine dönüştürülmektedir. Daha sonra sisteme 'özellik' olarak adlandırdığımız 3. bir boyut tanıtılmaktadır. Özellik, kart hamilleri ile sektörler arasındaki ilişkileri tanımlamada istasyon görevi görecektir. Bu amaç doğrultusunda önce kart hamili – sektör matrisi, kart hamili – özellik ve özellik – sektör matrislerine faktörize edilir; ardından da Hareketli En Küçük Kareler algoritması ile her iki matris parelel ve kısmen bağımsız olarak çözülür. Elde edilen değerler ile kart hamillerinin sektörlerle olan ilişkileri hesaplanır ve tahminleme yapılır. Tezin sonuç bölümünde bu tahminler gerçekleşmelerle kıyaslanarak modelin tutarlılığı değerlendirilmekte ve modelin gerçek hayatta değer katabileceği uygulama önerileri sunulmaktadır.

1. INTRODUCTION

While cascading every passing minute in size, diversity and importance; data has become one of the fundamental assets of the era we live in, the era of information. As well as creatingnew opportunities, the everlasting expansion of data has its own issues. Avenues, opened by this information flow come with a price. Amidst this huge amount of data, the task of making certain decisions becomes a challenge. (Jain et al., 2016)

Businesses have been benefiting from data for quite a long period of time, but it is recent that the challenge has diverted from actually gathering the data, to identifying the relevant and processing it in time. Today, we are surrounded with a vast ocean of information. The scale of data production is so enormous that 2.5 quintillion bytes of data are being created every day. The day by day increment of growth is so tremendous that 90% of the data in the world today has been created in the last two years alone(Syed et al., 2013). This perception of magnitude creates its own terminology: ‘Big Data’.

‘Big Data’ refers to datasets so voluminous they cannot be reasonably analyzed using traditional database management systems or software programs. Furthermore, Big Data consists of structured/unstructured and verified/unverified data (Syed et al., 2013). What is considered Big Data differs across various domains, and whether particular data are big or not is determined by whether these data push the capability limits of the information systems that work with these data (Vasarhelyi et al., 2015).As difficult as it is to truly define, there are four specific features of Big Data that challenge the capabilities of modern information systems (IBM, 2012; Laney, 2001; Zhang et al., 2015). These features include:

1.Volume – the massive size of a typical database
2.Velocity – data added on a continuous basis
3.Variety – types of data, both structured and unstructured
4.Veracity – reliability, authenticity, and validity of data.

The main challenge with the Big Data is to utilize it efficiently, that is identifying the relevant information among a massive pile, processing it rapidly, and coming to a conclusion in time. As well as developing more detailed strategies and making more solid decisions, the main goal with the Big Data is also to discover the uncharted, create patterns beyond know-how. Companies that effectively and efficiently utilize Big Data have the potential to gain significant competitive advantages including cost avoidance, increased profits, clear thinking, and new product/service development (Dennehy, 2016). The authors predicted that data-savvy managers and professionals with deep analytical skills will be much needed for businesses but hard to find (Chen, Chiang, &Storey, 2012; Dhar, 2013). McAfee & Brynjolfsson (2012) declared that businesses that do not base their decisions upon data analytics would have no place in future businesses due to the fierce market competition.

We have already stated, Big Data is big; but where does all that data come from? Following the digitalization trend in last decades, every tool we use in our daily life leaves a digital trace: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. And through the extension of individual internet access, this data spreads worldwide. By 2016, the number of people who use the internet has reached a count of 3.2 billion (Jain et al. 2016). All these actors come together to make us, the very individuals, data factories on our own.

The remarkable point here is that these actors, while feeding the growth, also provide information an exceptional diversity. The ‘individual data factories’ produce their own unique data, carrying the characteristics of the individual (McAfee & Brynjolfsson, 2012). Once managed to rule this information in its unique way, businesses can carry CRM one step further to one-to-one marketing. For example: in this study, with the help of credit card transaction data, we are going to analyze customer profileson card holder level and produce every card holder's own marketing proposal.

The literature has various studies on 'recommender systems', tools analyzing available data related with consumers and goods to predict interest on consumer-item level. Even though some recent researches studied recommender system applications in real-life cases, most papers are still mainly focused on the techniques instead and literature still lacks application-based articles.

Likewise, the researches on real-life applications commonly take explicit feedback datasets to work on; ratings, reviews, likes/dislikes on books, movies, series, e-commerce products etc., where users explicitly express their preference. However, on most real-life cases, such explicit feedback is not available and one needs to rely on implicit information, such as; clicks, purchases and time spent. Recommendation system applications on implicit feedback datasets would be regarded as pioneers in literature.

'Purchase' does not necessarily mean the consumer was content with his transaction in the end. For example; one may buy a book or a movie judging by its cover, but not enjoy it when actually using it. However, in our case, it can be assumed that a credit card holder is always witting on the sector he makes a purchase and thus purchase always reflects the user preference. As a result, I believe, credit card transactions will be a suitable case for preliminary studies on implicit recommender systems.

The rest of the study is organized as follows: Section 2: Preliminaries, where we survey the literature and set ground for our thesis, Section 3: Alternating Least Squares, where we define the steps of the model, Section 4: Application, where we apply the model in a real case and study it in details and Section 5: Conclusion where we evaluate the model and its value.

2. PRELIMINARIES

As groundwork for our study, we surveyed the literature for relevant papers related with recommender systems. Since researches on Big Data set ground for our topic, we begin our survey from mid 1990s, when the pioneer studies on Big Data was first published. We shallowly mentioned the papers there and deepened the investigation through the subheadings more specific to our work: types of recommender systems in general, then collaborative filtering and its models in particular and finally Alternating Least Squares model in detail. During the preliminary research, we did our best to cover the factors affecting our choice of model, comparatively with their alternatives.

2.1 Big Data

While cascading every passing minute in size, diversity and importance; data has become one of the fundamental assets of the era we live in, the era of information. Businesses have been benefiting from data for quite a long period of time, but it is recent that the challenge has diverted from actually gathering the data, to identifying the relevant and processing it in time.

Today, we are surrounded with a vast ocean of information. The scale of data production is so enormous that 2.5 quintillion bytes of data are being created every day. The day by day increment of growth is so tremendous that 90% of the data in the world today has been created in the last two years alone (Syed et al., 2013). A report from International Data Corporation (IDC), Gantz & Reinsel (2011) indicates that the overall created and copied data volume in the world was 1.8ZB, which increased by nearly nine times within a five year period. The world generated over 1ZB of data in 2010, and by 2014 7ZB per year (Richard et al., 2011). This perception of magnitude creates its own terminology: 'Big Data'.

Introduction of MapReduce and Apache Hadoop to the market has enabled processing extremely large datasets that has never been possible before due to restrictions on traditional database management system (DBMS) capacities (Agneeswaran, 2012) and combined with several other successful systems, lead the way on big data analysis (Kumar et al., 2013). Though no following research was published, Gantz & Reinsel's (2011) prediction that the return-on-investment (ROI) for the big data market would reach $16.1 billion (a growth about six-times faster than Information Technology businesses overall) in 2014 represents a value aspect of the notion. Therefore interest on big data has been on the rise lately.

2.1.1 Defining Big Data

There have been extensive discussions in both enterprise industrial organizations and academia about a consensus definition of "big data" (Team, 2011; Grobelnik, 2012). The term has superficially been applied to datasets that grow so large that they become awkward to work with using traditional database management systems (Elgendy & Elragal, 2014). According to Min et al. (2014), big data typically comprises masses of unstructured data that needs more real-time analysis. Manyika et al. (2011) defined big data as the next frontier for innovation, competition, and productivity (Intel IT Centre – Peer Research, 2012). Richard et al. (2011) stated that big data technology could be described as a new generation of technologies and architectures, designed so that enterprise organizations could economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, storage and analysis. This definition is largely agreed to by many researchers and enterprise industrial R&D managers (Seref & Duygu, 2013; Min et al., 2014; Manyika et al., 2011; Janusz, 2013)

While most researchers mainly concern about volumes, some argue that size is not the main characteristic or challenge of big data. Wu et al. (2014) study the technical challenges related to data samples, structures, heterogeneity of sources, mining models, algorithms and system infrastructures that would support data analytics. International Data Corporation (IDC), outlines some attributes of big data as the four Vs, that is, big data development sources (Variety – V1), big data acquisition (Velocity – V2), big data storage (Volume – V3), big data analysis (Veracity – V4), and finally modulating towards big data value adding or implementation benefits to industry (Value-adding – V5) (Gantz & Reinsel, 2011; Richard et al., 2011). This implies that big data is the data of which the volume, acquisition speed or representation limits the capacity of using classical database management methods to conduct effective analysis (Mayer-Schönberger & Cukier, 2013) and therefore efficient methods or technologies need to be developed and used to analyze and process big data. Multivariate analysis techniques such as regression, factor analysis, clustering, and discriminant analysis have been widely associated with such applications (Chen et al., 2012).

Russom (2011) claimed Big Data Analytics (BDA) as the use of advanced techniques, mostly data mining and statistical analyses, to find (hidden) patterns in (big) data. While various researchers such as; Herodotou, et al. (2011), Zaniolo et al. (2013), Chandramouli et al. (2013), Jin et al. (2014), studied data mining in detail, this paper will be focusing on the big data analysis rather than diving deep into data mining. Our main aim will be analyzing credit card transaction history to anticipate customer preferences in order to predict future purchases. That is where Recommender Systems come in handy.

2.2 Recommender Systems

Recommender Systems are applications aiming to analyze propensity of a set of users towards given items (Burke, 2002). These systems became an important research area since the publication of landmark papers in the 1990s, when the term "collaborative filtering" was coined (Resnick & Varian, 1997). Since then, the number of research papers published has increased significantly in many application fields e.g. books, documents, images, movies, music, shopping, TV programs (Park et al., 2012), as well as the amount of commercial applications of recommender systems by large companies such as Amazon.com (Linden et al., 2003), Google (Das et al., 2007), Last.fm (Eyke, 2009), Netflix (Bennett & Lanning, 2007), among others.

While other work has aimed at creating hybrid systems which use a mix of both, recommender systems are mainly classified in two categories according to the approach being used: content-based filtering (CB) and collaborative filtering (CF) (Melville et al., 2002).

2.2.1 Content-based Filtering

Content based filtering is a technique where individual user profiles are taken into account. It analyzes a set of documents rated by an individual user and uses the contents of the documents, as well as the provided ratings, to infer a user profile that can be used to recommend additional items of interest. It enhances the user's interest and predicts whether the user would be interested in eating at any particular restaurant or interested in seeing any particular movie (Basu et al., 1998). It represents the comparison between the content contained in the item with the content of items of user's interest. By using Bayesian hierarchical model, better user profiles for upcoming users is made by collecting feedbacks from the old users (Zhang & Koren, 2007). In CB can deal with sparsity by converting sparse user filled matrix into full user rating matrix (Melvillle et al., 2002). However, the syntactic nature of CB, which detects similarities between items that share the same attribute or characteristic, causes overspecialized recommendations that only include items very similar to those of which the user is already aware (Nores, 2008).

2.2.2 Collaborative Filtering

Collaborative filtering is a technique for predicting unknown preferences of people by using already known preferences from many users (Resnick & Varian, 1997). In general, CF uses an information filtering technique based on the user's previous evaluation of items or history of previous purchases (Su & Khosgoftaar, 2009; Zhang et al., 2015). It computes similarity on two basis: user and item. The historical data available helps building the user profile and the item profile. It uses cosine and Pearson correlation similarity approach (Ahn, 2008). Both the user profile and the item profile are used to make a recommendation system (Zhang et al., 2015; Bennet & Lanning, 2007). However, this technique has been known to reveal three major issues: sparsity problem, the scalability problem and cold start (Claypool et al., 1999; Sarwaret al., 2000a, 2000b).

In this study, we will be using implicit feedback data, meaning our input is not explicit user preferences, but users' implicit purchase history instead, and we will be focusing on users' transaction logs rather than the content of the items purchased. Thus, a collaborative filtering approach will be more appropriate in our study than a content based filtering one. In addition, we will be fencing out most of the disadvantages of collaborative filtering technique thanks to the nature of our area of work, credit card transaction sectors. Most of the implicit feedback data carry a risk of misdirection. For example; in movie or book recommendations, a purchase log does not necessarily mean that the user preferred the item he bought and that in the end he was happy he did. However in banking sector, it can be assumed that a credit card holder is always witting on the sector he makes a purchase and thus purchase always reflects the preference of user. Likewise, the set of sectors hardly ever evolve and we are using a set of customers with sufficient historical data. Hence it is unlikely to face cold-start problem either. In addition collaborative filtering is proved to be providing more accuracy than content-based techniques on most cases (Aberger, 2015). As a result, a collaborative filtering recommender system will be the choice in this research. CF introduces three main algorithms to deal with its challenges: memory-based CF, model-based CF and hybrid CF (Zhang et al., 2015).

2.2.2.1 Memory-Based Collaborative Filtering

In Memory-Based Collaborative Filtering, people with similar interests are combined to form a group and every user is a part of that group (Sarwar et al., 2001). It uses the user rating data to determine the similarity between users or items (neighborhood based methods) and make predictions or recommendations according to similarity values determined (Resnick et al., 1994). It is easy to implement and scales well with correlated items. There is no need of considering the content of items being recommended. There are many limitations of memory-based CF like cold start problem, sparsity and their dependencies on human ratings (Su & Khosgoftaar, 2009).

2.2.2.2 Model-Based Collaborative Filtering

On the other hand, Model-Based Collaborative Filtering develops models based on training data (such as data mining algorithms, machine learning, etc.) and then intelligent predictions are made for CF tasks for the real world data relying on learnt models (Breeze et al., 1998; Basu et al., 1999). It overcomes the challenges memory-based collaborative filtering face, however with a price. Model building is generally costlier and it inevitably loses some useful information for dimensionality reduction techniques (Su & Khosgoftaar, 2009). Providing this advantage, model-based CF will be our choice in this study.

2.2.2.3 Hybrid Collaborative Filtering

As is evident in its name, hybrid collaborated filtering systems, combines different techniques of collaborative approaches and other recommender techniques (usually content based approaches), to get better results. Various problems of each technique can be avoided by using hybrid approach (Adomavicius & Tuzhilin, 2005).

2.3 Alternating Least Squares Model Collaborative Filtering Recommender System

In the context of real-time recommendations operating on very large data-sets, the Memory-based CF approaches are not fast and not as scalable as how we would like them to be (Su & Khoshgoftaar, 2009). They present serious scalability problems given that the algorithm has to process all the data to compute a single prediction (Cacheda et al., 2011). These algorithms are not appropriate for real time recommendation systems with a large number of users (Anbazhagan &Arock, 2016).

Model-based CF methods were introduced in order to overcome the shortcomings of Memory-based CF methods (Karydi & Margaritis, 2014). Model-based CF methods' main advantage is its ability to deal with sparsity problems, which is very common in most real-life situations. Matrix factorization based CF algorithms have been proven to be effective to address the scalability challenges of CF tasks (Srebro et al, 2006; Rennie&Srebro, 2006; Tak´acs et al, 2008). As a matter of fact; Koren (2009), Thai-Nghe et al.(2011) and Lim (2013) claimed matrix factorization to be the most accurate approach to reduce dimensionality and overcome the sparsity problems.

There are various algorithms concerning matrix factorization practiced in literature. Some examples are: Singular Value Decomposition, (SVD), Principal Component Analysis (PCA), Latent Semantic Analysis (LSA), Latent Dirichlet Analysis (LDA), Stochastic Gradient Descent (SGD), Alternating Least Squares (ALS), Bayesian Networks, Clustering methods and Association Rule-based methods (Su & Khoshgoftaar, 2009).

Though no comparative research covering all the algorithms has been published yet, a relevant and important study analyzed the performance of two algorithms thoroughly: Aberger (2015) benchmarked Stochastic Gradient Descent (SGD) and Alternating Least Squares (ALS) algorithms. He took the experiments one step further and introduced Bias-Stochastic Gradient Descent (B-SGD) and Weighted Alternating Least Squares (W-ALS) methods as well to run a more solid benchmark. In conclusion, Aberger

(2015) stated while B-SGD performed better than ALS in majority of cases, on really sparse datasets ALS distinguishingly outperformed other methods both accuracy and performance-wise. Its parallel processing mechanics is also a remarkable advantage in terms of scalability. Since our input data is similar in sparsity and regarding other related data, the choice of algorithm in this study will be Alternating Least Squares.

3. ALTERNATING LEAST SQUARES MODEL RECOMMENDER SYSTEMS

3.1 Phase I: Customer X Item Matrix

To start with, we need to provide the input of our analyses: purchase logs with two information: customer ID and item ID To start with. Once the input is provided, the model starts with phase one: building the customerXitem matrix (A).

Table 3.1: Customer X Item Matrix

Customer ID	Item ID
C1	I1
C1	I2
C1	I3
C1	I6
C1	I1
C1	I7
C2	I2
C2	I3
C2	I2
C2	I4
C2	I6
C2	I2
C2	I3
C2	I3
C3	I1
C3	I3
C3	I1
C3	I6
C3	I7
…	…
…	…

$$
CI = \begin{array}{cc}
 & \begin{array}{ccccccccc} I1 & I2 & I3 & I4 & I5 & I6 & I7 & \cdots & Im \end{array} \\
\begin{array}{c} C1 \\ C2 \\ C3 \\ \cdots \\ Cn \end{array} & \left[\begin{array}{ccccccccc}
2 & 1 & 1 & 0 & 0 & 1 & 1 & \cdots & \cdots \\
0 & 2 & 1 & 1 & 0 & 1 & 0 & \cdots & \cdots \\
2 & 0 & 1 & 0 & 0 & 1 & 1 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots
\end{array} \right]
\end{array}
$$

3.2 Phase II: Factorization

Following the introduction of a third dimension, ‘feature’, the second phase begins: factorization of the customer ID x item matrix (CI) into Customer ID x feature matrix (CF) and feature x item matrix (FI).

$$
CI = CF \times FI \tag{3.1}
$$

where

Table 3.2: Customer X Feature Matrix

$$
CF= \begin{array}{cc}
 & \begin{array}{cccccc} F1 & F2 & F3 & F4 & \cdots & Fl \end{array} \\
\begin{array}{c} C1 \\ C2 \\ C3 \\ \cdots \\ Cn \end{array} & \left[\begin{array}{cccccc}
cf_{11} & cf_{12} & cf_{13} & cf_{14} & \cdots & \cdots \\
cf_{21} & cf_{22} & cf_{23} & cf_{24} & \cdots & \cdots \\
cf_{31} & cf_{32} & cf_{33} & cf_{34} & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \\
\cdots & \cdots & \cdots & \cdots & & \cdots
\end{array} \right]
\end{array}
$$

and

Table 3.3: Feature X Item Matrix

$$
FI = \begin{array}{cc}
 & \begin{array}{cccccc} I1 & I2 & I3 & I4 & \dots & Im \end{array} \\
\begin{array}{c} F1 \\ F2 \\ F3 \\ \dots \\ Fl \end{array} & \begin{bmatrix} fi_{11} & fi_{12} & fi_{13} & fi_{14} & \dots & \dots \\ fi_{21} & fi_{22} & fi_{23} & fi_{24} & \dots & \dots \\ fi_{31} & fi_{32} & fi_{33} & fi_{34} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \\ \dots & \dots & \dots & \dots & & \dots \end{bmatrix}
\end{array}
$$

Feature, the new dimension, serves as an instrument of clustering, helping creating customer-item profiles. For example: it is very plausible that there will be a high correlation between purchases in gas stations, car washes and auto spare parts. Likewise, customers owning a vehicle will obviously have a distinguishingly higher tendency to spend in these sectors than customers without a vehicle. So a feature, ‘having a vehicle’, can help defining a relational profile between customers and items in this scenario.

The ‘Big Data’ approach here is that we do not define features beforehand and thus we neither need to have know-how on our topic, nor need to be familiar with the customer profiles, sector, transaction types etc. As a matter of fact, we will not have a linguistic definition of these features even when our analysis is over. They are just purely data driven patterns. The only user-determined input here is the number of features to be used in the model.

The number of features is subject to change according to the structure of data, mainly the number of distinct customers and items. As Aberger (2015) stated, perhaps the most challenging part of machine learning in practice is picking the proper number of features and the proper algorithmic parameter values. There is an equilibrium point of feature numbers where the efficiency is at its maximum. Too few features, and we will not be able go deep into detail to catch all the patterns and end up grouping divergent profiles in same clusters, resulting in inaccurate recommendations. Too many features, and we will be separating similar profiles into different clusters, hence losing the chance to benefit from these data to make a predictive recommendation.

3.3 Phase III: Alternating Least Squares

As previously stated, instead of linguistic definitions of features, our focus is their expressions in terms of customers and items. Therefore we are going to approximate the product of factorization matrices to Customer X Item matrix and find the $cf_{ik}$and fi_{kj} variables for all i, j and$k$values The approximation cost function will be as follows:

$$f = \sum_{j=1}^{m} \sum_{i=1}^{n} (ci_{ij} - \sum_{k=1}^{l} cf_{ik} fi_{kj})^2 \qquad (3.2)$$

There are multiple methods to optimize this multiple-unknown-variable, non-convex equation. Alternating Least Squares (ALS) approach proposes fixing one of the factorization matrices with default values in every element and solving the other, then fixing the solved matrix and solving the fixed and so on until the equation converges and further iterations no longer minimize the cost function. Hence ALS approach simplifies our problem and basically turns it into an iterative linear regression, enabling parallel processing on each dimensions of the matrix individually. Due to its ability to deal with scalability and other advantages mentioned in 2.3, ALS will be our choice of algorithm in this study. Hastie et al. (2014) defines ALS as follows:

Inputs: Data matrix X, initial iterates A_0 and B_0, and k = 0.

Outputs: $(A^*, B^*) = \mathrm{argmin}_{A,B} F(A, B)$

Repeat until Convergence

for i=1 to m do

$$A_i \leftarrow \left(\sum_{j\in\Omega_i} B_j B_j^T\right)^{-1} \left(\sum_{j\in\Omega_i} X_{ij} B_j\right)$$

end for

for j=1 to n do

$$B_j \leftarrow \left(\sum_{i\in\Omega_j} A_i A_i^T\right)^{-1} \left(\sum_{i\in\Omega_j} X_{ij} A_i\right)$$

end for

3.4 Phase IV: Prediction

By the time the iterations are over, we will have both the Customer X Feature matrix and Feature X Item matrix optimized according to the cost function, which means we will know each customer's (cf_{ik}) and each item's(fi_{kj}) proximity with each feature and calculating their scalar productwill pave way to building relations between customers and items.

$$r_{ij} = \sum_{k=1}^{10} cf_{ik} fi_{kj} \tag{3.3}$$

Since we have the numerical values of relations for each customer-item combinations in the definition set, we can rank items for customers or use score cut-offs to choose items to predict.

4. APPLICATION

In this thesis, we are going to apply the recommendation model on credit card market to study the efficiency of the model with an implicit dataset. For that, a major player in banking industry in Turkey, BANK XYZ is chosen for its wide range of credit card transaction data. 78.081 random customers out of 1.6 million with transactions in at least 4 different MCCs (Merchant Category Code, which will be our item in the application) in September 2017 are selected. (Early attempts proved our analysis needs sufficient input information on card holder level to produce a consistent proposal. When card holders with lack of enough transaction logs to catch a pattern enter the model, it fails to identify the customer profile due to high deviation and thus resulting in inconsistent assignment to certain profiles.) These 78.081 customers add up to 983.899 credit card transactions in 229 different MCCs in September 2016, which makes 446.738 distinct customer-MCC combinations. Following some observations and computational limits we are going to use 10 features to define the matrixes and do 10 iterations, where both are subject to further studies in detail. Therefore, in our application; $i \in \{1,2,3,\ldots, 78081\}$,$j\in \{1, 2, 3, \ldots, 229\}$,$k\in\{1, 2, 3, \ldots, 10\}$.

4.1 Phase I: Customer X MCC Matrix

To show the details of how our model works, its parallel processing and recommendation algorithm, we will make up a short dataset of transactions, since the actual dataset is too large to examine in detail. As it is usually not suitable to work on such small datasets due to the models nature, we give a hand to the model by manipulating the data accordingly.

Consider a data of 12 customers in 7 different MCC's making up to a sum of 125 transactions as expressed by the following Customer x MCC matrix (CM):

Table 4.1: Customer X MCC Matrix

$$
\mathrm{CM} = \begin{array}{c|ccccccc|}
 & \mathrm{M1} & \mathrm{M2} & \mathrm{M3} & \mathrm{M4} & \mathrm{M5} & \mathrm{M6} & \mathrm{M7} \\
\mathrm{C01} & 4 & 4 & 3 & 1 & 3 & & \\
\mathrm{C02} & 2 & 1 & 3 & 1 & 4 & & \\
\mathrm{C03} & 4 & 8 & 4 & 2 & 5 & & \\
\mathrm{C04} & 7 & 4 & 1 & 5 & 3 & & \\
\mathrm{C05} & 5 & 4 & 2 & 1 & & & \\
\mathrm{C06} & & & & & & 7 & 3 \\
\mathrm{C07} & & 2 & & & & 4 & 3 \\
\mathrm{C08} & & 1 & & & & 1 & 6 \\
\mathrm{C09} & & 1 & & & & 2 & 5 \\
\mathrm{C10} & 1 & & & 1 & & & \\
\mathrm{C11} & 2 & & & 1 & & & \\
\mathrm{C12} & & 1 & 2 & & 1 & &
\end{array}
$$

As it is evident from the matrix, customers [C01;C05] all have common transactions on MCCs [M1;M5], except for one: C05,M5. Likewise, even if it is not as piled up as the first one, there is also another observable purchase pattern of customers [C06;C09] on MCCs {M2,M6,M7} where sole missing purchase is C06;M2. On the other hand, rest of thecustomers in the input does not bear an obvious pattern and looksrather divergent at first sight.

4.2 Phase II: Factorization

After building the Customer X MCC matrix (CM), we introduce the third dimension: feature and factorize the matrix into Customer X Feature (CF) and Feature X MCC (FM) matrices. Considering the patterns at first glance, 3 features seem rather promising to express relationships between customers and MCCs and form clusters.

We assign initial random integers to CF and FM between -10 and 10, totally arbitrary regarding only the scale of transaction counts, ranging between 0 and 8.

Table 4.2: Customer X Feature Matrix

Let

$$CF = \begin{array}{cc} & \begin{array}{ccc} F1 & F2 & F3 \end{array} \\ \begin{array}{c} C01 \\ C02 \\ C03 \\ C04 \\ C05 \\ C06 \\ C07 \\ C08 \\ C09 \\ C10 \\ C11 \\ C12 \end{array} & \begin{bmatrix} -7 & 10 & 3 \\ -8 & 4 & 7 \\ -5 & -3 & 6 \\ 0 & -9 & 6 \\ 8 & 9 & -5 \\ -7 & 9 & 2 \\ -2 & -2 & 5 \\ 6 & 3 & 4 \\ 5 & -8 & 7 \\ 1 & 0 & 8 \\ -7 & -3 & 9 \\ -2 & 6 & -2 \end{bmatrix} \end{array}$$

And

Table 4.3: Feature X MCC Matrix

Let

$$FM = \begin{array}{cc} & \begin{array}{ccccccc} M1 & M2 & M3 & M4 & M5 & M6 & M7 \end{array} \\ \begin{array}{c} F1 \\ F2 \\ F3 \end{array} & \begin{bmatrix} -2 & -4 & 3 & -2 & -3 & -1 & -10 \\ -2 & 7 & 8 & 7 & 7 & -9 & 5 \\ -7 & 9 & 5 & 6 & -9 & -7 & 3 \end{bmatrix} \end{array}$$

4.3 Phase III: Alternating Least Squares

Now, our goal is to minimize the 'error', predicted-observed distance, which is formulized by the following cost function:

$$f = \sum_{j=1}^{12} \sum_{i=1}^{7} (cm_{ij} - \sum_{k=1}^{3} cf_{ik} fm_{kj})^{2} \tag{4.1}$$

Proceeding ALS method, we will hold FM constant, iterate CF once, hold CF constant, iterate FM once, back to holding FM constant and iterating CF once and so on for 20 iterations. The method enables iterating each vector of CF and CM independently. Minimizing f_i for fm_{kj} where $cm_{ij}$is known and cf_{ik} is constant iterates FM matrix and minimizing f_j for $cf_{ik}$where cm_{ij} is known and fm_{kj} is constant iterates CF matrix.

$$f'_i = \sum_{j=1}^{12} (cm_{ij} - \sum_{k=1}^{3} cf_{ik} fm_{kj})^{2} \tag{4.2a}$$

$$f''_{\boldsymbol{j}} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{7}} (cm_{ij} - \sum_{k=1}^{3} cf_{ik} fm_{kj})^{2} \tag{4.2b}$$

4.3.1 Initial State

Regarding the random initial feature vectors, we have CF, FM, and prediction matrix and f, f’ and f’’ cost functions as follows:

Table 4.4: CF, FM, CM matrixes and f, f’, f’’ cost functions in the initial state

	M1	M2	M3	M4	M5	M6	M7
F1	-2	-4	3	-2	-3	-1	-10
F2	-2	7	8	7	7	-9	5
F3	-7	9	5	6	-9	-7	3

	F1	F2	F3								f''_i=
C01	-7	10	3	-27	125	74	102	64	-104	129	62.022
C02	-8	4	7	-41	123	43	86	-11	-77	121	46.353
C03	-5	-3	6	-26	53	-9	25	-60	-10	53	10.757
C04	0	-9	6	-24	-9	-42	-27	-117	39	-27	20.653
C05	8	9	-5	1	-14	71	17	84	-54	-50	17.829
C06	-7	9	2	-18	109	61	89	66	-88	121	51.152
C07	-2	-2	5	-27	39	3	20	-53	-15	25	6.161
C08	6	3	4	-46	33	62	33	-33	-61	-33	14.527
C09	5	-8	7	-43	-13	-14	-24	-134	18	-69	26.505
C10	1	0	8	-58	68	43	46	-75	-57	14	21.049
C11	-7	-3	9	-43	88	0	47	-81	-29	82	26.011
C12	-2	6	-2	6	32	32	34	66	-38	44	10.658
f'_i =				15.247	59.842	23.939	34.518	72.248	40.302	67.581	f= 313.677

4.3.2 1st Iteration - CF

We start the first iteration with CF, so we hold FM constant. We manipulate each C_jF vector to decrease f''_j:

Table 4.5: CF, FM, CM matrixes and f, f', f'' cost functions in 1st iteration - CF

FM:

	M1	M2	M3	M4	M5	M6	M7
F1	-2	-4	3	-2	-3	-1	-10
F2	-2	7	8	7	7	-9	5
F2	-7	9	5	6	-9	-7	3

CF and CM:

	F1	F2	F3	M1	M2	M3	M4	M5	M6	M7	f''_i =
C01	- 5,18	- 2,83	2,28	0,06	21,43	-26,78	4,23	-24,79	14,69	44,49	4.184
C02	- 6,12	1,92	- 2,47	25,69	15,69	-15,35	10,86	54,03	6,13	63,39	7.770
C03	- 3,89	- 3,46	1,59	3,57	5,65	-31,40	-6,90	-26,86	23,90	26,37	3.620
C04	- 0,02	- 4,89	6,62	-36,52	25,43	-6,08	5,53	-93,75	-2,31	-4,39	11.789
C05	5,07	3,82	- 5,30	19,32	-41,24	19,27	-15,20	59,23	-2,35	-47,50	8.582
C06	- 4,94	- 2,80	1,64	4,00	14,92	-29,02	0,12	-19,54	18,66	40,32	2.991
C07	- 1,91	- 2,41	1,16	0,52	1,21	-19,21	-6,09	-21,58	15,48	10,53	1.061
C08	1,85	1,32	- 1,68	5,42	-13,28	7,71	-4,54	18,81	-1,97	-16,94	1.202
C09	4,74	- 6,40	7,42	-48,62	3,02	0,12	-9,76	-125,80	0,92	-57,14	22.151
C10	1,02	- 0,08	0,14	-2,86	-3,38	3,12	-1,76	-4,88	-1,28	-10,18	173
C11	- 6,34	- 3,54	1,80	7,16	16,78	-38,34	-1,30	-21,96	25,60	51,10	5.532
C12	- 1,78	1,20	- 2,18	16,42	-4,10	-6,64	-1,12	33,36	6,24	17,26	1.756
			f'_i =	5.397	3.781	5.650	617	35.739	1.809	17.818	f= 70.812

4.3.3 1st Iteration - FM

We continue the first iteration with FM, so we hold CF constant. We manipulate each FM_i vector to decrease f''_i:

Table 4.6: CF, FM, CM matrixes and f, f', f'' cost functions in 1st iteration - FM

	F1	F2	F3	M1	M2	M3	M4	M5	M6	M7	$f''_i =$
F1				-1,80	-3,59	1,38	-1,96	-3,45	-0,60	-0,67	
F2				-2,31	8,29	3,84	6,97	1,50	-4,69	5,31	
F2				-1,40	6,34	6,04	5,87	-0,01	-5,45	3,07	
C01	- 5,18	- 2,83	2,28	12,67	9,59	-4,24	3,81	13,60	3,95	-4,56	316
C02	- 6,12	1,92	- 2,47	10,04	22,23	-15,99	10,88	24,02	8,13	6,71	1.485
C03	- 3,89	- 3,46	1,59	12,77	-4,64	-9,05	-7,16	8,21	9,90	-10,89	718
C04	- 0,02	- 4,89	6,62	2,06	1,50	21,18	4,82	-7,33	-13,13	-5,63	749
C05	5,07	3,82	- 5,30	-10,53	-20,14	-10,35	-14,42	-11,71	7,93	0,62	1.414
C06	- 4,94	- 2,80	1,64	13,06	4,92	-7,66	-0,21	12,83	7,16	-6,52	509
C07	- 1,91	- 2,41	1,16	7,38	-5,77	-4,88	-6,24	2,96	6,13	-7,96	311
C08	1,85	1,32	- 1,68	-4,03	-6,35	-2,53	-4,29	-4,39	1,86	0,61	144
C09	4,74	- 6,40	7,42	-4,14	-23,03	26,78	-10,34	-26,03	-13,27	-14,38	2.705
C10	1,02	- 0,08	0,14	-1,85	-3,44	1,95	-1,74	-3,64	-1,00	-0,68	46
C11	- 6,34	- 3,54	1,80	17,07	4,83	-11,47	-1,68	16,55	10,60	-9,02	857
C12	- 1,78	1,20	- 2,18	3,48	2,52	-11,02	-0,94	7,96	7,32	0,87	288
$f'_i =$				988	1.984	2.254	607	1.973	820	914	$f =$ 9.541

4.3.4 2nd Iteration - CF

Following the second iteration, we go back to holding FM constant and manipulating CF:

Table 4.7: CF, FM, CM matrixes and f, f', f'' cost functions in 2nd iteration - CF

	M1	M2	M3	M4	M5	M6	M7
F1	-1,80	-3,59	1,38	-1,96	-3,45	-0,60	-0,67
F2	-2,31	8,29	3,84	6,97	1,50	-4,69	5,31
F2	-1,40	6,34	6,04	5,87	-0,01	-5,45	3,07

	F1	F2	F3	M1	M2	M3	M4	M5	M6	M7	f''_i =
C01	- 2,29	- 2,87	2,46	7,31	0,03	0,68	-1,08	3,57	1,43	-6,15	77
C02	- 1,22	1,52	- 2,61	2,34	0,43	-11,61	-2,34	6,52	7,83	0,88	293
C03	- 3,82	- 1,78	1,92	8,30	11,13	-0,51	6,35	10,49	0,18	-1,00	99
C04	- 0,02	- 4,86	6,64	1,97	1,88	21,42	5,14	-7,29	-13,38	-5,41	761
C05	4,47	4,55	- 3,75	-13,31	-2,10	0,99	0,94	-8,56	-3,58	9,65	553
C06	- 3,23	- 1,90	1,99	7,42	8,46	0,27	4,77	8,27	0,00	-1,82	290
C07	- 1,89	- 1,46	1,37	4,86	3,37	0,06	1,57	4,32	0,51	-2,28	87
C08	1,66	1,61	- 1,26	-4,94	-0,60	0,86	0,57	-3,30	-1,68	3,57	52
C09	3,77	- 4,92	7,84	-6,40	-4,62	33,66	4,34	-20,46	-21,92	-4,58	2.307
C10	- 0,17	- 0,08	0,15	0,28	0,90	0,36	0,66	0,47	-0,34	0,15	2
C11	- 3,16	- 1,93	2,28	6,95	9,80	2,00	6,13	7,98	-1,48	-1,13	218
C12	- 1,11	1,29	- 1,20	0,70	7,07	-3,83	4,12	5,78	1,16	3,91	128
			f'_i =	560	309	1.829	122	819	899	327	f = 4.866

4.3.5 2^{nd} Iteration - FM

Like in the first iteration, we continue by manipulating FM while holding CF constant:

Table 4.8: CF, FM, CM matrixes and f, f', f'' cost functions in 2^{nd} iteration - FM

	M1	M2	M3	M4	M5	M6	M7
F1	-1,19	-1,14	1,41	-1,03	-0,36	-0,84	1,18
F2	-1,91	1,10	4,53	1,13	-0,38	0,77	2,52
F2	-1,75	1,32	3,49	1,31	-0,18	0,17	2,32

	F1	F2	F3	M1	M2	M3	M4	M5	M6	M7	f''_i =
C01	- 2,29	- 2,87	2,46	3,90	2,70	-7,64	2,34	1,47	0,13	-4,23	137
C02	- 1,22	1,52	- 2,61	3,12	-0,38	-3,94	-0,44	0,33	1,75	-3,66	83
C03	- 3,82	- 1,78	1,92	4,59	4,93	-6,75	4,44	1,71	2,16	-4,54	167
C04	- 0,02	- 4,86	6,64	-2,31	3,44	1,13	3,23	0,66	-2,60	3,13	112
C05	4,47	4,55	- 3,75	-7,45	-5,04	13,83	-4,38	-2,66	-0,89	8,04	478
C06	- 3,23	- 1,90	1,99	3,99	4,22	-6,22	3,79	1,53	1,59	-3,98	167
C07	- 1,89	- 1,46	1,37	2,64	2,36	-4,50	2,09	0,99	0,70	-2,73	76
C08	1,66	1,61	- 1,26	-2,85	-1,78	5,24	-1,54	-0,98	-0,37	3,09	57
C09	3,77	- 4,92	7,84	-8,81	0,64	10,39	0,83	-0,90	-5,62	10,24	273
C10	- 0,17	- 0,08	0,15	0,09	0,30	-0,08	0,28	0,06	0,11	-0,05	1
C11	- 3,16	- 1,93	2,28	3,46	4,49	-5,24	4,06	1,46	1,56	-3,30	75
C12	- 1,11	1,29	- 1,20	0,96	1,10	0,09	1,03	0,13	1,72	-0,84	10
			f'_i =	356	141	642	75	47	121	255	f = 1.637

4.3.6 End of Iterations

At the end of 20th iteration, what we have at hand is as follows:

Table 4.9: CF, FM, CM matrixes and f, f', f'' cost functions at the end of 20th iteration

	F1	F2	F3	f''_i =
C01	- 4,57	- 0,85	1,73	3
C02	- 2,95	0,29	0,53	14
C03	- 6,45	- 0,77	2,22	14
C04	- 0,04	- 0,23	- 0,27	100
C05	- 4,91	2,64	- 1,07	13
C06	0,14	- 0,37	2,00	45
C07	- 1,03	- 0,53	2,72	12
C08	- 0,37	0,30	0,77	21
C09	0,60	0,62	- 0,31	25
C10	0,13	- 0,09	0,22	2
C11	- 0,52	0,56	- 0,29	5
C12	- 0,95	1,28	- 1,42	3

	M1	M2	M3	M4	M5	M6	M7
F1	-0,66	-0,76	-0,80	-0,22	-0,40	0,02	0,37
F2	-0,44	0,86	-0,67	0,13	-0,23	-0,04	1,49
F2	-0,14	0,62	-0,60	0,04	-0,20	0,22	1,50

	M1	M2	M3	M4	M5	M6	M7	
C01	3,15	3,81	3,19	0,96	1,68	0,32	-0,36	
C02	1,75	2,82	1,85	0,71	1,01	0,05	0,14	
C03	4,29	5,62	4,34	1,41	2,31	0,39	-0,20	
C04	0,17	-0,33	0,35	-0,03	0,12	-0,05	-0,76	
C05	2,23	5,34	2,80	1,38	1,57	-0,44	0,51	
C06	-0,21	0,82	-1,06	0,00	-0,37	0,46	2,50	
C07	0,53	2,01	-0,45	0,27	-0,01	0,60	2,91	
C08	0,00	1,02	-0,37	0,15	-0,08	0,15	1,47	
C09	-0,63	-0,12	-0,71	-0,06	-0,32	-0,08	0,68	
C10	-0,08	-0,04	-0,18	-0,03	-0,08	0,05	0,24	
C11	0,14	0,70	0,21	0,18	0,14	-0,10	0,21	
C12	0,26	0,94	0,75	0,32	0,37	-0,38	-0,57	
f'_i =	61	32	6	28	29	60	41	f= 257

4.3.7 Phase IV: Prediction and Evaluation

Studying the results, we notice that the model fills the missing spots in the patterns as foreseen. Except for the observed transactions, customer C05 is expected to have a transaction in MCC5 with a distinguishingly higher chance than any other MCC, cm'_{55} is equal to 1,57 while cm'_{65} is -0,44 and cm'_{75} is 0,51. Likewise customer C06's score on MCC2 is 0,82 while no other non-observed MCC has a positive value.

On the other hand, we recognize that the model cannot produce solid predictions on non-observed MCCs for the rest of the customers. Customers C01, C02, C03, C04, C07, C08 and C09 shape the patterns they fit in and as a result; they do not have missing spots in their patterns. This is due to the fact that the dataset we worked on is too small comparing with an actual set and manipulated to intensify the predictions on C05-MCC5 and C06-MCC2. On a real-life case, these 'saturated patterns' or 'perfect matches' are unlikely to be this frequent.

For the customers C10, C11, C12, another reason plays a role in the inability of prediction: their transaction logs are too divergent from the rest of the dataset and 3 features are not enough to form a cluster of their own. Unable to associate these 3 customers with the built relations, our model could not manage to produce predictions.

4.4 RESULTS

To evaluate the success of our model on the actual data,we will compare the recommendations with the actual purchase realizations in the following month, October 2016. We will consider a prediction accurate, if the subject customer has at least one transaction in the recommendation MCC in October 2016. If the subject customer has no transactions in the recommendation MCC, it will be an inaccurate prediction.

First of all, out of 78.081 input customers, 2.740 turned out to not have made any purchase transactions in the following month. Therefore, we will exclude these customers from the evaluation set, and check the recommendations of the other 75.341 customers. Out of this 75.341, 58.331's top scored recommendation were accurate, at least one transaction realization in the recommended MCC in October 2016, which is equivalent to %77 accuracy.

While respectable in accuracy, the downside is that, due to model's clustering nature, %89 of the recommendations are actual realization transactions in input data. In consequence, it is hard to make an innovative use of the model as is.

However, if we take account of the secondary recommendations (recommendations that are not the top scored) as well, we can increase the ratio of 'new' recommendations, at the cost of accuracy obviously. For instance: taking the top scored MCC among top 5 predictions, we can raise the 'new' predictions' ratio against 'same' predictions to %69; but the accuracy drops to %34. This approach brings a new dimension to our study: the optimization of accuracy and new recommendation ratio.

Assuming %50 accuracy an acceptable level, our goal will be to cover maximum number of customers with new recommendations. In a set of top scored MCCs among top 3 predictions, using a score cut-off of 1.04, it is possible to cover 22.991 customers with new predictions while ensuring the %50 accuracy ratio. This accounts for %31 of the whole customer set with at least one transaction in October 2016.

5. CONCLUSION

5.1 Thesis Contribution

As seen during the observations, our prediction model permits managing the accuracy vs. extent balance in user's initiative and creates marketing proposals acceptable in both accuracy and extent levels.

In banking industry, there are various ways to benefit from this model. Most basically, it is reasonable to track these prediction realizations and at the end of user-defined *n* days, to propose a reward in return of a sum of transactions worth user-defined *x* TLs in recommended MCCs to the customers who has not purchased in that that MCC yet.

One could come up with more advanced proposals as well. For example; rather than encouraging the customers to purchase in their recommended MCCs, cross-sales activities using the recommendations as hooks can be more efficient and more extensive. One such activity can be for example: credit card limit increase. Identifying customers with %95 outstanding balance/limit ratios and offering them a reward in return of a sum of transactions in recommended MCCs before leading them to increase their credit card limit would augment the success of the proposal rather than a raw limit increase communication.

As well as customer driven proposals, it is also possible to benefit from the model from MCC perspective too. For sector based mass campaigns, the model can help identify the customers who show relatively less tendency to participate and enable cost reduction.

For example; we can exclude such customers from a routine food/café based campaign SMS mailing and dispose of the SMS cost.

In conclusion, Alternating Least Squares Model Collaborative Filtering Recommendation System is an easy-to-implement, adaptive, flexible and efficient tool to manage CRM activities on customer level. It is easy to implement because it is not very heavy in computation load and possible to work it on ordinary computers even with million data logs. Actually the real potential lies within continuous data flow and real-time analyses and these require an appropriate technical infrastructure, but still periodic runs and updates can make a remarkable contribution as well.

It is adaptive, because it is machine-learning through and through, no room for know-how. Providing similar data input, one can use the model in various activities in various sectors, not necessarily purchase transactions.

It is flexible, since users can easily scale the accuracy and extent of the model as they will. It is possible to cover a large portion of customer base, also to select a niche segment with maximum prediction accuracy.

Finally, to conclude, our prediction model has significant potential in terms of CRM and customer-level analyses. Businesses that are able to feed the model with the required data, should benefit from this or suchlike models to increase their efficiency in their line of work.

5.2 Future Work

To improve the model even further, the area that carries the biggest potential is the inputs. We are using only two information, customer and MCC, to make predictions; however there are various constant knowledge like; age, gender, occupation etc. that have undeniable correlation with card holders' preferences. And it is not limited only by the customer; some characteristics of MCCs also affect card holders' transactions, such as: location, seasonality, ability to do installments etc. There are numerous known factors at hand that may steer the model to a better accuracy but that we are totally neglecting as is. Further studies on this topic, should focus on working with multiple dimensions.

Another immature point of this work is the choice of number of features. Number of features is a delicate parameter, influencing the results of the model immensely. Too few features, and it will not be possible to go deep into detail to catch all the patterns and end up grouping divergent profiles in same clusters, resulting in inaccurate recommendations. Too many features, and we will be separating similar profiles into different clusters, hence losing the chance to benefit from these data to make a predictive recommendation. In this study, we decided the number of features to use on several observations; however the choice deserves a dedicated study of its own. Artificial neural networks applications face a very similar challenge regarding the determination of number of branches in neural tree, so researchers may also benefit from the works on that area.

REFERENCES

Aberger, C. R. (2015). Recommender: An Analysis of Collaborative Filtering Techniques, **URL:***http://cs229.stanford.edu/proj2014/Christopher%20Aberger,%20Recommender.pdf*

Addo-Tenkorang, R., and Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering 101*.pp. 528–543.

Adomavicius, G. and Tuzhilin,A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowledge and Data Eng., vol. 17,* pp. 734–749.

Agneeswaran, V. (2012). Big-Data – Theoretical, Engineering and Analytics Perspective. (S. Srinivasa, & V. Bhatnagar, Eds.)LNCS , 7678, pp. 8-15.

Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, *Information Sciences: an International Journal, Vol. 178, Issue 1,* pp 37 –51.

Anbazhagan, M. and Arock, M. (2016).A Study and Analysis of Collaborative Filtering Algorithms for Recommender Systems.*IJCT A, ISSN: 0974-5572,* pp. 127-136.

Anders, U. and Korn, O. (1999).Model Selection in Neural Networks.*Neural Networks, 12,* pp. 309–323.

Basu, C., Hirsh, H. and Cohen, W. (1998). Recommendation as classification: using social and content based information in recommendation. 1*5th National Conference on Artificial Intelligence (AAAI ’98),* pp. 714–720, Madison, Wis, USA.

Berry, M. J. A. and Linoff, J. S. (2004).Data Mining Techniques for Marketing. Sales and Customer Relationship Management (2nd ed.,).

Breese, J., Heckerman, D. and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering.*14th Conference on Uncertainty in Artificial Intelligence (UAI '98)*

Cacheda, F., Carneiro, V., Fernández, D.L. and Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems, *ACM Transactions on the Web.*

Cai, D., He, X., Wen, J. R. and Ma, W. Y. (2004). Block-level link analysis, *27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 440–447.

Chandramouli, B., Goldstein, J. and Duan, S. (2012). Temporal Analytics on Big Data for Web Advertizing.*The 28th International Conference on Data Engineering (ICDE)*, pp. 90-101.

Chandramouli, B., Goldstein, J. and Quamar, A. (2013). Scalable Progressive Analytics on Big Data in the Cloud. *The 39th International Conference on VLDB*, pp. 1726-1737. Trento.

Chen, H., Chiang, R. and Storey, V. (2012). Business Intelligence and Analytics: from Big Data to Big Impact. *MIS Quarterly , 36 (4)*, pp. 1165-1188.

Cheng, Y., Qin, C. and Rusu, F. (2012). GLADE: Big Data Analytics Made Easy. *SIGMOD*, pp. 697-700.

Cho, Y. H., Kim, J. K. and Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction.*Expert Systems with Applications*, 23, pp. 329–342.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. and Sartin, M. (1999).Combining content-based and collaborative filters in an online newspaper, *ACM SIGIR Workshop on Recommender Systems '99.*

Cuzzocrea, a., Song, I. and Davis, K. (2011). Analytics over Large-Scale Multi-dimensional Data: The Big Data Revolution! *DOLAP*. pp. 101-103. Glasgow.

Dennehy, B. (2016). How to use Big Data to create business advantage in the real world.*Washington society of certified public accountants annual meeting*, Seattle, WA.

Dhar, V. (2013).Data Science and Prediction.*Communications of the ACM , 56 (12)*, pp. 64-73.

Elgendy, N. and Elragal, A. (2014). Big Data Analytics: A Literature Review Paper. *The 14th Industrial Conference on Data Mining (ICDM).* Petersburg

Elragal, A. and Haddara, M. (2014). Big Data Analaytics: A Text Mining Based Literature analysis. *NOKOBIT*

Fisher, D., DeLine, R., Czerwinski, M.and S., D. (2012). Interactions With Big Data Analytics. *Interactions,* pp. 50-59.

Frias-Martinez, E., Chen, S. Y. and Liu, X. (2009). Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability. *International Journal of Information Management,* 29, pp. 48–56.

Ghazal, A., Rabl, T., Hu, M., Raab, F., Poess, M., Crolotte, A., et al. (2013). Big Bench: Towards an Industry Standard Benchmark for Big Data Analytics. *SIGMOD,* pp. 1197-1208. NY: ACM.

Gupta, R., Gupta, H. and Mohania, M. (2012). Cloud Computing and Big Data Analytics: What Is New from a Database Perspective.(S. Srinivasa, & V. Bhatnagar, Eds.)*LNCS , 7678,* pp. 42-61.

Han, X., Li, J., Yang, D. and Wang, J. (2013).Efficient Skyline Computation on Big Data.*TKDE , 25 (11),* pp. 2521-2535.

Hastie, T., Mazumder, R., Lee, J. D., and Zadeh, R. (2014).Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares.*Statistics Department and ICME* Stanford University.

He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., et al. (2011). RC File: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems. *The 27th International Conference on Data Engineering (ICDE)* , pp. 1199-1208.

Herodotou, H., Lim, H., Luo, G., Boisov, N., Dong, L., Cetin, F., et al. (2011). Starfish: A Self-tuning System for Big Data Analytics. *5th Biennial Conference on Innovative Data Systems Research (CIDR 11),* pp. 261-272. CA.

Hu, Y., Koren, Y. and Volinsky, C. (2008).*Eighth IEEE International Conference on Data Mining*

Huai, Y., Lee, R., Zhang, S., Xia, C. and Zhang, X. (2011). DOT: A Matrix Model for Analyzing, Optimizing and Deploying Software for Big Data Analytics in Distributed Systems. *SOCC,* pp. 1-14.

Ibnkahla, M. (2000).Applications of neural networks to digital communications-a survey.*Expert Systems with Applications, 80,* pp. 1185–1215.

J. Bennett, J. and Lanning, S. (2007). The Netflix Prize, *ACM SIGKDD Explorations Newsletter – Specia issue on visual analytics, Vol. 9 Issue 2*, pp. 51 – 52.

Jain, A., Jain, V., and Kapoor, N., (2016) A Literature Survey on Recommendation System Based on Sentimental Analysis. *Advanced Computational Intelligence: An International Journal (ACII), Vol.3, No.1*, New Delhi.

Karydi, E. and Margaritis, K. G. (2014). Parallel and Distributed Collaborative Filtering: A Survey, *Information Retrieval Distributed, Parallel, and Cluster Computing.*

Kim, H. K., Kim, J. K. and Ryu, Y. U. (2009).Personalized recommendation over a customer network for ubiquitous shopping. *IEEE Transactions on Services Computing, 2*, pp. 140–151.

Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R. and Suh, J. H. (2002).A personalized recommendation procedure for internet shopping support.*Electronic Commerce Research and Applications, 1*, pp. 301–313.

Koren, Y. (2009). Matrix Factorization Techniques for Recommender Systems, *IEEE Computer Society*, *IEEE 0018-9162/09*, pp. 42-99.

Kumar, A., Niu, F. and Ré, C. (2013). Hazy: Making It Easier to Build and Maintain Big Data Analytics. *Communications of The ACM , 56 (3),* pp.40-49.

Laptev, N., Zeng, K. and Zaniolo, C. (2013). Very Fast Estimation for Result and Accuracy of Big Data Analytics: the Earl System. *The 29th International Conference on Data Engineering (ICDE),* pp. 1296-1299.

Lihua, W., Lu, L., Jing, L. and Zongyong, L. (2005). Modeling user multiple interests by an improved GCS approach. *Expert Systems with Applications, 29,* pp. 757–767.

Lim, D. (2013). Matrix Factorization Presentation, University of California, San Diego, **URL***: http://www.academia.edu/3695261/Matrix_Factorization_Presentation*

Lopez-Nores, M., Garca-Duque, J., Frenandez-Vilas, R.P. and Bermejo-Munoz, J. (2008).A flexible semantic inference methodology to reason about user preference in knowledge-based recommender systems.*Knowledge-Based systems, 21,* pp. 305–320.

Lu, J. and Li, D. (2013). Bias Correction in a Small Sample From Big Data. *TKDE* , 25 (11), pp. 2658-2663.

Mahmood, T. and Ricci, F. (2009).Improving recommender systems with adaptive conversational strategies, *20th ACM conference on Hypertext and Hypermedia*, pp. 73–82.

Malhotra, N. K. (2007). Marketing research: An applied orientation (5th ed.). Pearson Education Inc.

McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution. *HBR*, pp. 3-9.

Melville, P., Mooney, R. J. and Nagarajan, R. (2002).Content-boosted collaborative filtering for improved recommendations.*Eighteenth National Conference on Artificial Intelligence (AAAI-02),* pp. 187. Edmonton, Alberta.

Mozafari, B., Zeng, K., D'Antoni, L. and Zaniolo, C. (2013).High-Performance Complex Event Processing over Hierarchical Data.*ACM Transactions on Database Systems , 38 (4),* pp. 21-39.

Narang, A., Srivastava, A. and Katta, N. (2012).High Performance Offline & Online Distributed Collaborative Filtering.*The 12th International Conference on Data Mining (ICDM),* pp. 549-558.

Oard, D. W. and Kim, J. (1998). Implicit Feedback for Recommender Systems, *5thDELOS Workshop on Filtering and Collaborative Filtering,* pp. 31–36.

Park, D. H., Kim, H. K., Choi, I. Y., Kim, J. K. (2012). A literature review and classification of recommender systems research.*Expert Systems with Applications 39.*pp. 10059–10072

Rennie, J. D. M. and Srebro, N. (2005). Fast maximum margin matrix factorization for collaborative prediction, *22nd International Conference on Machine Learning, ICML ‘05,* Bonn, Germany

Resnick, P. and Varian, H. R. (1997). Recommender systems, *Communications of the ACM, vol. 40, no. 3,* pp. 56–58.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994). Grouplens: An open architecture or collaborative filtering of netnews, *ACM Conference on Computer Supported Cooperative Work,* pp. 175–186.

Ricci, F., Rokach, L., Shapira, B. and Kantor, P. B (2010).Recommender Systems Handbook; First Edition; Springer-Verlag New York, Inc. New York, NY, USA.

Roddick, J., Spiliopoulou, M., Lister, D. and Ceglar, A. (2008).Higher Order Mining.*SIGKDD Explorations* , 10 (1), pp. 5-17.

Russom, P. (2011). Big Data Analytics.TDWI , 4th Quarter, pp.1-38.

Sarwar, B., Karypis, G., Konstan, J. A. and Riedl, J. (2000a). Application of dimensionality reduction in recommender system- a case study.*ACM Web KDD-2000 Workshop.*

Sarwar, B., Karypis, G., Konstan, J. A. and Riedl, J. (2000b). Analysis of recommendation algorithms for e-commerce.*ACM E-Commerce*, pp. 158–167.

Sarwar, B.M., Karypis, G., Konstan, J. A. and Riedl, J. (2001). Item based collaborative filtering recommendation algorithms. *10th International Conference on World Wide Web (WWW '01),* pp. 285–295.

Shardanand, U. and Maes, P. (1995).Social information filtering: Algorithms for automating 'Word of Mouth'.*Human Factors in Computing Systems Conf.*

Singh, S., & Singh, N. (2012).Big Data Analytics.*International Conference on Communication, Information & Computing Technology (ICCICT)*, pp. 1-4.Mumbai.

Srebro, N., Rennie, J. D. M. and Jaakkola, T. (2005).Maximum margin matrix factorization.*Advances in Neural Information Processing Systems*, vol. 17, pp. 1329–1336.

Su, X. and Khoshgoftaar, T. M. (2009).A survey of collaborative filtering techniques.*Advances in Artificial Intelligence.*pp. 19.

Syed , A. et al. (2013). The future revolution on Big Data, *International Journal of Advanced research in Computer and Communication Engineering, 2* pp. 2446–2451.

Tak´acs, G., Pil´aszy, I., N´emeth, B. and Tikk, D. (2008). Investigation of various matrix factorization methods for large recommender systems, *IEEE International Conference on Data Mining Workshops, ICDM '08,* Pisa, Italy, pp. 553–562.

Thai-Nghe, N., Drumond, L., Horvàth, T., Nanopoulos, A. and Schmidt-Thieme, L. (2011). Matrix and Tensor Factorization for Predicting Student Performance, *3rd International Conference on Computer Supported Educaction (CSEDU).*

Thusoo, A. S. (2010). Hive-a petabyte scale data warehouse using Hadoop. *26th International Conference on Data Engineering (ICDE)*, pp. 996-1005.

Tomasi, G. and Bro, R. A. (2006). Comparison of algorithms for fitting the PARAFAC model, *Comput.Statist.Data Anal.50,* pp. 1700–1734.

Vasarhelyi, M. A. et al. (2015). Big Data in accounting: An overview.*Accounting Horizons, 29,* pp. 381–396.

Wang, J., Zhao, P., Hoi, S., & Jin, R. (2014). Online Feature Selection and Its Applications. *TKDE , 26 (3),* pp. 698-710.

Wold, J. P., Bro, R. A., Veberg, A., Lundby, F., Nilsen, A. N. and Moan, J. (2006). Active photosensitizers in butter detected by fluorescence spectroscopy and multivariate curve resolution, *J. Agric. Food Chem., 54*, pp. 10197–10204.

Wu, X., Zhu, X., Wu, G. and Ding, W. (2014).Data Mining with Big Data.*TKDE , 26 (1),* pp. 97-107.

Zhang, Y. and Koren, J. (2007).Efficient Bayesian Hierarchical User Modeling for Recommendation Systems.*30th annual International ACM SIGIR Conference on Research and Development*, pp. 47 – 54.

Zhang, Y., Zhang, M. and Liu, Y. (2015).Incorporating Phrase-level Sentiment Analysis on Textual Reviews for Personalized Recommendation, *Eighth ACM International Conference on Web Search and Data Mining*, pp. 435 – 440.

BIOGRAPHICAL SKETCH

İlkay Körpe was born on January 1st,1990 in İstanbul. He received majority of his education in Galatasaray Institutions. He went toGalatasaray High School in 2004 and graduated in 2009. The same year, he started Galatasaray University and studied in Department of Industrial Engineering for four years. As his partial fulfillmentof the requirementsfor hisbachelor of science, he wrote his thesis on matrix factorization algorithms. In 2013, after his graduation from universityhe began his further studies on Industrial Engineering in Graduate School of Science and Engineering in Galatasaray again and received his Master of Science in 2017.