Exploring big data opportunities for online customer segmentation

Georgia Fotaki
Marco Spruit
Sjaak Brinkkemper
Dion Meijer

contact: m.r.spruit@uu.nl

Department of Information and Computing Sciences,
Utrecht University, The Netherlands
Princetonplein 5, 3584 CC Utrecht, NL

ABSTRACT

In today's competitive business environment, more and more organizations move or extent their business online. Thus, there is an increasing need for organizations to build concrete online marketing strategies in order to engage with their customers. One basic step towards achieving the objectives related to online marketing is the segmentation of online customers, based on the customer data gathered online. Since there is an onslaught of customer information collected from online sources, new techniques are required for managing and analyzing the huge amount of data, and this is where the concept of Big Data can play an essential role.

This research sheds light on three fields: Online Marketing, Customer Segmentation, and Big Data Analytics. The three domains are integrated into the Online Customer Segmentation (OCS) framework, which attempts to show how online marketing objectives can be supported by techniques and tools applicable to extremely large datasets. For the creation of the OCS framework a set of main online marketing objectives is defined. Moreover, the differences among customer attributes gathered from offline and online channels are discussed and OCS categories are identified. Finally, the concept of Big Data is introduced and relevant techniques and tools suitable for analyzing customer segmentation categories and segmenting customers effectively are described. We demonstrate the OCS framework by applying it on a hypothetical business scenario using an online customer data set.

Keywords: Online Marketing, Online Customer Segmentation, Big Data, Data Mining.

Introduction

In today's fiercely competitive business environment, organizations struggle to improve customer experience, achieve customer retention, and grow their customer base. The increasing need to keep customers satisfied and treat them as individuals has triggered interest in customer engagement. Customer engagement is defined by Van Doorn et al. (2010) as "the behavioral manifestation from a customer towards a brand or a firm which goes beyond purchase behavior". Furthermore, more attention is paid on building concrete online marketing strategies, due to the tremendous growth of internet technologies. In their effort to leapfrog the competition, many organizations have already started employing tools that facilitate online marketing strategies and customer engagement such as Web analytics tools, Social Media monitoring software, Web Content Management systems, Audience Targeting tools and Online Customer Engagement Management tools (Fotaki et al., 2012). However, objectives for online marketing are still, for many organizations, not well defined and not yet aligned with the business strategy (Chaffey et al., 2009).

A vital process for Online Marketing is Customer Segmentation, which constitutes the process of dividing customers into distinct and homogeneous groups. Customer segmentation is considered an effective method for managing different customers with different preferences, while developing diverse marketing strategies (Chen et al., 2007; Tsitptis & Chorianopoulos, 2009). Online customers can be segmented according to their characteristics that are tracked online with the use of specific techniques and algorithms. There are various types of segmentation based on certain customer attributes gathered from several sources. Customer segmentation types intent to support different business tasks or activities regarding marketing goals (Tsiptis & Chorianopoulos, 2009). In the meantime, the business world is facing the challenge of dealing effectively with the onslaught of customer data and information stemming from online sources. Hence, the use of "Big Data" tools and techniques that are able to handle and analyse a huge amount of data in real time has become inevitable. Big Data is a new term primarily used to describe the data sets that are so large and complex that they require advanced and unique storage, management, analysis and visualization technologies (Chen et al., 2012). Most of the existing software tools for Online Marketing and Customer Engagement, such as OCEM tools, can track and handle huge amounts of online customer data. Therefore, they could integrate and utilize big data approaches in order to facilitate and enhance OCS, as well as other functions they might perform providing a holistic customer view.

There is a plethora of research conducted for customer segmentation in traditional Customer Relationship Management systems that basically work with data gathered from offline channels. However, OCS in real time has not received much attention in research. Moreover, since the notion of big data is still in its infancy, there is not much research on the field yet. It is apparent that there is a need to explore the utilization of big data for segmentation of online customers. With an aim to explore the applicability of big data for OCS, this research investigates the three fields of Online Marketing, Customer Segmentation and Big Data Analytics, and explores their interrelationships. The three terms are combined into one framework that attempts to show how online marketing objectives can be supported by an effective OCS, which can be implemented by techniques and tools applicable to large datasets.

The paper is structured as follows: Section 2 details the research questions that we investigate in this research as well as the research methods that we have applied. Section 3 presents the literature study to provide the basis for this research. In Section 4 we present a case study to detail our findings; Section 5 elaborates the findings; Section 6 provides the evaluation of our research findings and finally, Section 7 details the conclusions and provides directions for future research.

Research Design

The two research questions that this research aims to explore are the following: (i) "Which are the customer segmentation types that can assist each of the business goals regarding online marketing?" and (ii) "Which big data approaches and techniques can be used for each Online Customer Segmentation type?".

This research follows the Design Science Research Methodology (DSRM) (Peffers et al., 2007), which focuses on providing a certain artifact that contains the characteristics of the research outcomes. The DSRM consists of six activities, which we followed to conduct this research: (i) Problem identification to define the specific research problem and justify the value of a solution to this problem; (ii) Definition of the objectives for a solution to the problem; (iii) Design and Development of the artifact; (iv) Demonstration of how the artifact is capable of solving the problem;(v) Evaluation of the artifact to measure and observe how well the it supports a solution to the problem; and (vi) Communication of the problem, and the artifact and its effectiveness to professionals and researchers, and suggestions for future research.

In this research the artifact is a framework that constitutes an indirect mapping among big data techniques for OCS and online marketing goals. In order to design the framework three factors are taken into account; (i) online marketing business objectives, (ii) customer segmentation types, (iii) big data techniques and tools. Initially, a set of business objectives regarding online marketing and OCE is determined, based on both scientific theory and the data gathered during the interviews with business consultants and marketers of DEVCORP. Secondly offline customer segmentation types are described, based on literature review. Complementary to the literature review on customer segmentation, data from the OCEM tool were observed and analyzed in order to identify OCS types. Finally, a comprehensive literature review was made on big data and their differences and similarities to data mining are highlighted. A focused literature study regarding techniques used for customer segmentation is made and approaches that are appropriate to assist customer segmentation are presented and analyzed. During the whole process semi-structured and unstructured interviews with experts of DEVCORP were made, for the related topics.

The aforementioned data were analysed in order to be used as an input for the design of the framework that this research proposes. The proposed framework was validated by experts of DEVCORP, as well as by a customer intelligence expert and author of one of the books used for this research.

Literature Review

Our research intersects the fields of customer segmentation and online marketing on the one hand, and on the other hand available big data tools and techniques suitable for customer segmentation.

Customer Segmentation and Online Marketing

The initial point of satisfactory customer segmentation is the determination of business objectives and relevant KPIs related to marketing and the relationship with the customers (Tsiptsis & Chorianopoulos, 2009). Correspondingly, before proceeding to segmentation of online customers, it is essential that online and interactive marketers define a clear online marketing strategy and all the business tactics and processes involved in it. With the increasing impact of internet, online marketing plays a vital role for their success. However, as stated by Chaffey & Smith (2008), one of the main problems is the lack of clear online marketing objectives. The unclear responsibilities for many tasks related to online marketing and the treatment of internet as "just another channel to the market" constitute some of the most common challenges in managing an Online marketing Strategy (Chaffey et al., 2009).

As evident from literature, the main objectives of online marketers are related with the customer lifecycle, which is the stage through which a customer passes during a long term relationship with an organization. Those stages are: *Customer Acquisition* regards the processes to attract new customers; *Customer Retention* includes the activities to maintain relationships with existing customers; *Customer Development* regards activities aiming to extent customer's involvement with the organization and (Chaffey et al., 2009; Peterson, 2004). For a successful online marketing strategy it is essential: to acquire and attract new online customers; to improve churn rates, customer satisfaction and loyalty in order to achieve customer retention; to increase sales and improve conversion rates in order to reach the stage of customer development (Chaffey et al., 2009; Chaffey & Smith, 2008).

Customer segmentation is performed based on customer data, according to which customers can be grouped. There is not any one-size-fits-all segmentation scheme that is able to assist effectively all the business needs and goals. There are several segmentation approaches available, each of them suitable for supporting different business needs (Pachidi, Spruit & Weerd, 2014; Tsiptsis & Chorianopoulos, 2010). The main customer segmentation types based on attributes gathered from offline sources are described: *Value-based* is a segmentation process, through which customers are categorized according to their value and include attributes such as: Customer LifetimeValue, Value of Orders, Value of purchases etc. (Chan 2008; Kim et al. 2006; Kumar et al., 2010); *Behavioral Segmentation* is a process, through which customers are grouped according to usage, attitude and behavior regarding a product or promotion (Stroud, 2006; Tsiptsis & Chorianopoulos, 2009). Behavioral segmentation might include: product ownership, type of transactions &frequency of transactions, revenue history, payments, and product utilization etc.; *Loyalty or Engagement Segmentation* is used to determine different groupings of customers according to different degrees or loyalty to supplier or brand (Stroud, 2006). The segments for loyalty or engagement segmentation are often created by applying simple business rules and/or cluster models on survey or database information (Tsiptsis & Chorianopoulos, 2009). Attributes related to customer loyalty that can be: Engagement/Loyalty Score, Frequency of purchases, Number of Complains etc.; *Demographic* segmentation customers are grouped according to their demographic or social characteristics (Stroud, 2006). Customer attributes that belong to this category are: Age, gender, income, ethnicity, marital status, education other personal details of the consumer etc. *Attitudinal segmentation* is performed in order to explore customers' needs that can be fulfilled by the purchase of a product or service, views, attitudes, and preferences (Tsiptis & Chorianopoulos, 2009). Attributed included might be: Preferences, Interests, Customer needs, Motivations, Usage occasion, lifestyles, personality, preferable promotions.

Big Data Characteristics

The idea of “big data” is relatively new so there is not much scientific literature available. Although there is not a standard scientific definition of the term yet, some researchers and analysts have attempted to define big data. Most analysts and researchers focus on the most significant attributes –known as V’s- of big data, which characterize the nature of the data (Russom, 2011; Forrester, 2012; O’Reilly, 2012; Sathi, 2012).

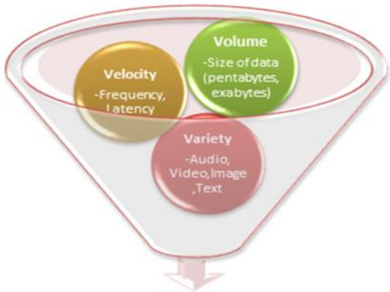

Figure 1: The three Vs of Big Data.

As shown in Figure 1 these are: *Volume* , which refers to the amount of data (terabytes, petabytes, etc.) that are created and can be processed by big data techniques (Sathi, 2012); *Velocity* that regards the frequency at which data are generated as well as the latency of the data (O’Reilly, 2012; Zaslavsky, Perera & Georgakopoulos, 2012; Russom, 2012); and *Variety* that regards the different types and sources of data such as images, text from social networks or mobile devices, web logs, streamed video or audio. An example of such a definition is the one given by Gartner[1], which describes big data *as “high volume, high velocity, and/or high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”.*

Based on the above description of Big Data, in this research we adopted the following definition of Big Data as: “*Data characterized by high-volume, high-velocity, and high-variety, which require advanced and exclusive tools and techniques for information management, processing and storage, analysis and visualization”.*

As of using various data mining techniques for big data and traditional data, there are various differences.

The key factor of their differentiation is the main characteristics of the data that each technique is able to handle and analyze. A deeper look into the definitions and description of the two terms reveals their main differences. Traditional data mining techniques and tools are basically intended for structured relational databases of a certain size while big data are characterized by large volume, quick rate of data generation and diversification in terms of types and sources of data. Moreover, big data can be found in any form; structured, semi-structured or unstructured and hence cannot be easily handled through Relational Database Management Systems (RDBMS). As stated by Chen et al. (2012) unstructured or semi-structured and diverse data require ad-hoc extraction processes, indexing and analytics within scalable and distributed environments, which is quite different than the ones that traditional data mining entails. Hence, the key difference between “big data” and “traditional data” lies in the characteristics of the data and the different tools and technologies that these characteristics call for. Techniques that are used for big data analysis are the same as those for traditional data mining techniques. For instance Classification, Clustering, Regression, Time Series Analysis, Association Rules, AB Testing, Visualization are some of the techniques applicable to big data (Manyika et.al,2011; Chen et al.2012). However, unlike the tools intended for traditional data, tools and technologies used for big data are able to analyze larger amounts of unstructured data, which stem from several sources and are found in various types.

Tools and Algorithms suitable for Customer Segmentation

There is a large body of scientific papers, reports and books, reports on several data mining algorithms and approaches that can be used for customer segmentation. Primarily, the data mining techniques for analyzing big data and traditional data are same; only the tools that implement those techniques in big data sets are different. Therefore, techniques and algorithms intended for traditional customer segmentation can also be suitable for data characterized by big volume, velocity and variety. In order to observe the trend on the techniques used for customer segmentation a focused literature review was conducted. The purpose was to find relevant studies that present and analyze techniques appropriate for certain segmentation types. The literature review resulted in 19 studies on customer segmentation using data mining algorithms. The algorithms can be implemented in very large databases with diversified data.

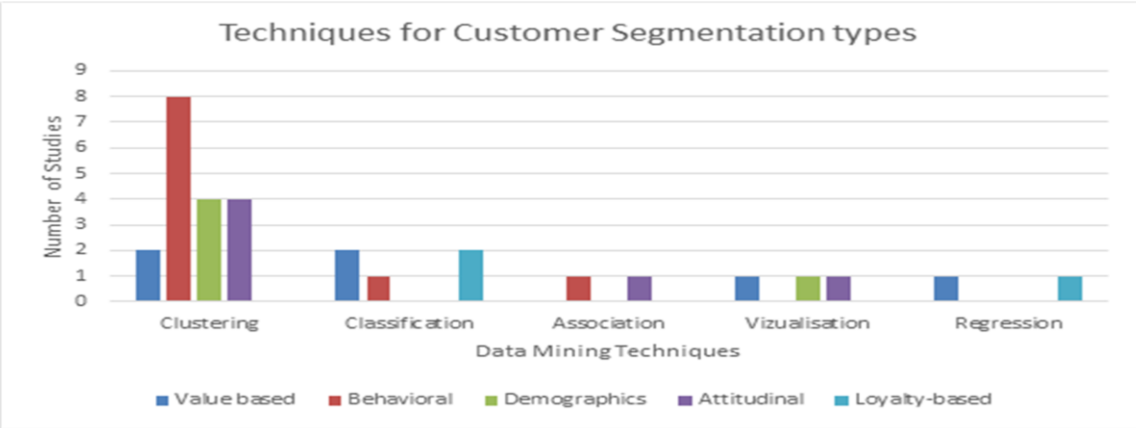

Figure 2: Frequency of techniques used for customer segmentation.

Figure 2 illustrates the techniques that were found in the studies and the customer segmentation type they analyze. The segmentation techniques found are the following: Clustering, Classification, Association, Regression and Visualization. Out of the studies using clustering, 73% use clustering for behavioral segmentation (Wu & Chou, 2011; Jansen, 2007; Ye, Yijun &Zhu, 2013; Cheng & Chen, 2009) 36 % for demographics (Chen et al., 2007, Tsiptsis & Chorianopoulos, 2009), 36% for attitudinal (Hong & Kim, 2012; Miguéis et al., 2012) and 18% for value-based (Chan, 2005). On the other hand, out of the studies using classification, 75% use classification for loyalty-based (Chan, 2008), 50% for value-based segmentation (Han, Lu &Leung, 2012), and 25% for behavioral (Kim et al., 2006). As far as the rest of the techniques are concerned; Association rules is used in one study for attitudinal segmentation (Pillai & Vyas, 2012); Visualization is used in another study for visualizing customer segments based on their value, demographics and needs and attitudes (Woo et al., 2005); Regression algorithms are used for value-based and loyalty-based segmentation (Hosseni & Tarokh, 2009). These statistics indicate that clustering and classification are the prevailing segmentation techniques, while the rest are used less.

Case Study

We conducted a case study at DEVCORP, a global provider of Web Content Management and Online Marketing software. Due to space limitations, the details of the case study are not presented in this paper, and hence we refer to Fotaki et al. (2013). After conducting a case study in DEVCORP and observing the different online customer data gathered by the OCEM tool from various online channels in order to create customer profiles, we designed the following structure as shown in Table which identifies the OCS types while showing the differences between online and offline customer segmentation types, in terms of the attributes they include. As it is illustrated Attitudinal, Behavioral, Demographic, Value-based, and Loyalty-based are found both offline and online, while referral and technical online segmentation types are newly defined. *Referral Segmentation* includes characteristics that show how the visitor ended up on a certain website, revealing their incentives and previous channels they visited. *Technical Segmentation* includes characteristics that indicate the technical information of the device, operating system, or browser that an online customer uses when entering a websites.

	Customer Segmentation types	**Attributes from offline sources**	**Attributes from online sources**
Online & Offline	**Attitudinal**	Preferences& Interests, Customer needs, Motivations, Usage occasion, lifestyle, personality, preferred promotions (based on market surveys)	Interactions Clicked, Viewed & Converted, Visited Channels & Sites, Page Views, Social media type, Clicked banners, Offers viewed , Product type purchased or viewed
	Behavioral	Frequency of transactions &purchases, Payments, Product ownership, Product utilization, Last Purchase, Revenues	Number of Visits, Last visit, Number of clicks, Number of Interactions, Subscriptions, Number of purchases
	Demographic	Age, gender, Income, education, job, Ethnicity, Region, Marital status, nr of children	Email Address, I.P (indicates location), Age, Genre, Facebook Id, Twitter Id
	Loyalty-based	Frequency of purchases, Nr of complaints, Loyalty score,	Engagement score based on business rules regarding customers' behavior
	Value-based	Customer Value, Monetary Value Propensity scores for profitability	Average value of purchases, Average value of orders
Online	**Referral**	------	Average Purchase value, Average Order Value
	Technical	------	Flash Version, Java Version, Operating System name, Operating System Version

Table 1: Customer segmentation attributes gathered from offline and online sources.

During the case study interviews were conducted with the experts of DEVCORP. Table shows the anonymized list of the interviewees and their fields of expertise.

Interviewee ID	Interviewee Job Title
M1	Chief Marketing Officer
M2	Product Marketer
B1	Business Consultant
B2	Business Consultant
B3	Online Marketer /Business Consultant
A1	Author / Customer Intelligence Expert
A2	Software Architect

Table 2: Details of the interviewees.

OCS Framework

Framework Overview

The Online Customer Segmentation (OCS) framework is presented in Figure 3, showing which OCS types can be preferably analyzed by certain techniques in order to assist main online marketing objectives. The table shows the level of usefulness of each OCS type based on online customer attributes gathered from online channels is indicated by different colors. In the colored cells the big data techniques suitable for analyzing each OCS type, are shown. The online marketing objectives illustrated, are the main online marketing objectives as they were found in related literature and during the interviews with the experts of DEVCORP. The framework suggests which segmentation categories are more preferable and useful for assisting each of the related online marketing objectives. In order to indicate the usefulness of each segmentation type for each of the online marketing objectives a High-Medium-Low scale is used, which is indicated by the different colors of the cells of the table. **H (green)** indicates that the segmentation type is highly useful for the corresponding objective, and therefore should be preferred; **M (yellow)** indicates that the segmentation type is useful in some cases and complementary to the highly useful segmentation types; **L (red) indicates** that the segmentation type is not that useful for the corresponding business goal, especially if it is used alone. The H-M-L values were identified based on the definition and measurements of each objective, usual actions related to the achievement of the objective and opinions of the interviewees regarding the objectives in relation with the customer attributes, on which more attention should be paid for the achievement of the objectives.

OCS FRAMEWORK	Online Customer segmentation types						
	Attitudinal	Behavioral	Demographic	Loyalty-based	Referral	Technical	Value-based
Objectives	Big Data Techniques (Implemented by big data tools)						
Increase New Customer Acquisition	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Improve Churn Rates	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Customer Satisfaction	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Customer loyalty	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Cross –up sales	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Improve conversion rates	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression

Level of usefulness of online customer segmentation types: *High* *Medium* *Low*

Figure 3: The Online Customer Segmentation Framework.

Online Marketing & OCS types

Firstly, online marketers aim to form relations with new customers and achieve customer acquisition. ***New Customer Acquisition*** is achieved when new visitors enter a website, and move on with interacting with it. Especially, for business sites and content sites it is essential to have an increasing number of new visitors in their websites (Peterson, 2006).As stated by B3 "*it is very* *important that online customers are directed to the right online channels and be exposed to content similar to their preferences*." Hence, online marketers should focus on Referral and Technical segmentation. Referral and technical attributes are the first to be gathered when a new customer profile is created. With referral not only can it be identified if the customer is new or not, but also which are the incentives that lead them to the website. Moreover by technical, marketers can adapt the content showed to the customer based on the technical perspectives of the device or browser that each customer uses, in order to create a friendlier interface for the new customer. Therefore, based on these segmentation types, marketers can show targeted content (banners, offers etc.) to each visitor so as to facilitate customer acquisition. After the first visit, in order to achieve acquisition preferences should also be taken into account. Therefore, attitudinal category would be also considered quite useful in combination with referral and technical info categories.

Secondly, the ***improvement of churn rates*** is related to customer retention, since customer retention is achieved when the likelihood of a customer to churn is decreased. Customer retention entails targeting the most valuable and profitable customers and keeping them buying or visiting the website (Chaffey et al., 2009). Therefore, focus should be shed on value-based segmentation, which allows for the identification and targeting of profitable and valuable customers. Moreover, behavioral segmentation type reveals whether a client is likely to churn or not, based on transactional behavior and number of visits. Online marketers can decide on providing most valuable customers with tailored propositions or promotions or award them with offers, by segmenting customers according to behavioral and value-based characteristics. As it was mentioned by B3 during the interview" *the improvement of churn rates is also related with having your clients exposed to content according to their preferences*." Therefore, segmenting them according to attitudinal segmentation type constitutes a complement to value-based and behavioral segmentation types and therefore rated as of medium usefulness

As M1 mentioned during the interviews "*it is important to give to the online customer the feeling that they are recognized and treated as individuals, in order to feel satisfied*". Therefore, ***satisfaction*** of online customer increases when they are exposed to relevant content, products and services. The most effective segmentation type in this case is attitudinal with which online marketers can decide on suitable content and offers based on their customer needs, avoiding having them exposed to irrelevant content that in most cases becomes annoying. Customers' satisfaction is difficult to be captured and measured online, since most of the times it can be obvious by a phone call or by filling in a satisfaction form. However, segmenting the customers according to behavioral characteristics or their engagement/loyalty score reveals the level of satisfaction, since the frequency of visits and interactions is an indication the level of visitors' satisfaction.

*"**Loyalty*** can be measured as the number of visits any visitor is likely to make over lifetime as a visitor" (Peterson, 2004), and can be typically measured by the number of visits per visitor. As stated by M1 "a loyalty or engagement score is a way to check customers' loyalty and engagement with the brand". Loyalty-based segmentation includes attributes that indicate the loyalty and the interest of the customer for the brand. Hence, it is very important to be used as a segmentation type for identifying loyal or disloyal customers and support the increase of customer loyalty. Moreover, the frequency of visits, or the time that a visitor spends on a website also indicates customers' loyalty (Peterson, 2006). Therefore, behavioral segmentation can also be used for distinguishing loyal and disloyal customers, and proceed to corresponding marketing tactics to increase customer loyalty.

The ***increase of cross-up sales*** is a very important objective for online retailers and entails the selling of additional and more expensive products to customers (Chaffey et al, 2009; Chaffey &

Smith, 2008; Peterson, 2006). In order to achieve the increase of cross-up sales, the online marketer has to understand the needs and preferences of the customers their buying behavior and identify the most valuable and loyal customers that would purchase products of higher price. Therefore, it is suggested that online marketers focus on the following segmentation types: Attitudinal, Value-based and Behavioral segmentation in order to: (i) provide customers with relevant products and promotions based on their needs and lifestyles and (ii) identify most active clients, who will be more willing to move to a purchase.(iii) identify most profitable clients. Moreover, Demographics, like the age, genre or the marital status, are also valuable for delivering relevant content to the customer and increase cross-up sales. However, as it as stated by B3 "demographics category contains attributes that are not easy to be captured online, and thus are only useful in some specific business cases". Therefore, since it can be used complementary to the attitudinal segmentation type is considered of medium usefulness.

The ***improvement of conversion rates*** is the most important objective of online marketers. B3 stated that "The performance of conversion rates is actually what online marketers are judged upon". A conversion entails the successful completion of certain activities by online customers that have a positive effect for the online organization (Peterson, 2004). The high visiting frequency of the website is the basic measurement of conversions, since conversion rates are calculated by the proportion of visitors that move on to a certain activity on a website, such as click on interactions, proceed on a purchase or on a subscription and obtain information (Winer, 2001). As mentioned by B3 and M2 the improvement of conversion rates all segmentation types can be important, depending always on the marketers' point of view. Attitudinal and referral segmentation types can play an important role for improving conversion rates, since the first indicates preferences and the latter the incentives that lead the visitor to the website, both allowing for distributing relevant content to the visitor. Behavioral segmentation is also important for online marketers, in order to identify customers with lower activity level and decide upon further marketing actions to trigger those customers to become more active on the website.

Techniques for OCS Types Analysis

The cells within the OCS framework in Figure 3 contain the techniques, which are preferable to be used for analyzing each of the customer segmentation types, in order to create effective customer segments. The techniques showed on the framework are suitable for analyzing structured or unstructured data of big volume, and velocity.The results are based on the literature review, while the opinions of the experts were also taken into account. First of all, the prevailing technique that suits in most cases for customer segmentation is **clustering**. Cluster analysis belongs to the unsupervised modeling techniques. In that case all customer attributes from a selected set can be simultaneously analyzed, in order to create customer segments with similar characteristics. A cluster model is able to manage a large number of attributes and reveal data-driven segments which are not known in advance. As it was also mentioned by A2 clustering is used when it is difficult to know the segments in advance, and thus a general algorithm is needed. Therefore, clustering is preferred for behavioral segmentation. A well-known big data tool that can be used for the implementation of clustering techniques is Apache Mahout (Anil, Dunning & Friedman, 2012). On the other hand, **classification** is also a very popular technique, which is used for supervised modeling. Classification is used when it is required for the segments to end in a specific result. This means that there is always a dependent variable according to which segments are created (Turban et al., 2010, Manyika et al., 2011).Classification is preferably used for loyalty-based segmentation or value-based segmentation. The amount of money spent a customer spend or the degree of loyalty could constitute the dependent variable. However, it can also be used in some cases for behavior segmentation or for attitudinal segmentation, but still is preferred for loyalty-based or value-based segmentation. Classification techniques can be implemented as well by the big data tool Apache Mahout (Anil et al., 2012).

Regression is a traditional statistical technique basically used for prediction and forecasting. Regression cannot be considered a straightforward segmentation technique. However, it is able to create customer segments by predicting the segment in which a customer belongs. It is very useful for loyalty segmentation, where the customers are segmented according to their degree of loyalty (loyal or disloyal). As it was also mentioned by A2 regression analysis is also useful for value-based segmentation where customers are grouped according to their profitability, predicting if a client is profitable or not. Regression techniques can be implemented as well by using language R for big data (Manyika et al., 2011). ***Association*** is a technique, which also belongs to unsupervised modeling and is used for recommendation engine for web page personalization (Cakir & Aras, 2012). Association is not a straightforward segmentation technique, since it is basically used for uncovering relations among the data. However, some characterize it as segmentation technique since it is capable of "predicting" customers' preferences according to previous behavior. Therefore, it is highly useful for segmentation based on customers' preferences, in order to find out what type of content the visitor might prefer to see or what type of product they might prefer to buy. Thus, it is very effective when used for products recommendations and web page personalization based on the user preferences. Apache Mahout is a big data tool that provides recommendation engines able of predicting patterns of user preferences out of very large amount of data (Owen et al. 2012). ***Visualization*** is not a segmentation technique itself, but is used complementary to segmentation in order to illustrate segments and make them more understandable. As A1 stated, it is often used to visualize the results of clustering techniques. For OCS, it is very important to visualize huge amounts of online customer data into segments with the use of visualization tools for big data, indicating preferences, behaviors or consumer characteristics. There are plenty of tools for visualization of Big Data such as Visual Insight & Jaspersoft.

As explained before technical and referral segmentation types were defined during the research. These segmentation types are not found in the literature, and thus it is not possible to make an assumption for the techniques suitable for analyzing these data, without implementing and comparing the techniques on real data. However, as it was also discussed during the interviews with A1 and A2, it is obvious that technical and referral segmentation can be assisted by both clustering and classification techniques, based on whether a dependent variable is needed to be used or not.

Evaluation

Example of OCS Framework

In order to provide an overview of how the OCS Framework could be utilized, a hypothetical business scenario was used, including analysis of online customer data. The hypothetical business scenario revolves around an organization which wants to increase conversion rates by selling more memberships through its website. For analyzing customer data a sample of 1200 customer profiles were obtained, and the IBM SPSS statistic data editor was used. In order to create clusters of customers, the TwoStep algorithm, which belongs to the clustering techniques, was applied. Figure 4 illustrates the result of a clustering technique (TwoStep algorithm), which was applied by using SPSS on online customer data that belong to the behavioral segmentation type. By inserting into the clustering algorithm the online behavioral data: member, number of visits, average time, visit clicks and click counts; five customer segments were revealed. The segments show which customers are members or not with high or moderate activity. Online marketers can target customers who are not members or have high activity on the website in order to have them expose to content regarding the purchase of memberships. This would increase the conversion rates of customers purchasing membership. More details can be found in Fotaki et al. (2013).

Clusters

Input (Predictor) Importance: 1,0 0,8 0,6 0,4 0,2 0,0

Cluster	1	4	5	3	2
Label	Moderate active visitor (uknown ...	Not a Member	Moderate active member	Higly active member	Higly active visitor (uknown ...
Description	Uknown Membership, lower visiting time and clicks	Visitor is not a member	Visitor is a member, lower visiting time and clicks	Visitor is a member, high visiting time and clicks	Uknown Membership, high number of clicks and visits
Size	66,5% (799)	16,3% (196)	11,9% (143)	3,6% (43)	1,7% (20)
Inputs	ismember 0 (100,0%)	ismember 2 (100,0%)	ismember 1 (100,0%)	ismember 1 (81,4%)	ismember 0 (55,0%)
	visits 291,99	visits 249,40	visits 383,76	visits 1.414,07	visits 728,50
	averagetime 0,71	averagetime 1,03	averagetime 0,50	averagetime 1,19	averagetime 41,80
	clickcount 762,60	clickcount 773,44	clickcount 1.003,36	clickcount 4.087,53	clickcount 7.556,20
	visitclicks 2,76	visitclicks 2,96	visitclicks 2,59	visitclicks 4,33	visitclicks 26,65

Figure 4: Online customer clusters created based on the behavioral customer segmentation type.

Marketing Experts Evaluation

The framework that this research proposes was evaluated by experts. Since the framework involves both the marketing and the technical perspective, interviews with experts of different expertise were conducted. The two aspects of the framework were separately evaluated. First of all, the part of the framework, which shows the usefulness of OCS types for the support of online marketing goals, was evaluated by two marketers and two business consultants (M1, M2, B1, and B3). The second part of the framework, which matches the OCS types with the big data techniques, was evaluated by a software architect and an author of one of the books used for the purpose of this research, specialized in customer intelligence (A2, A1).

M1, M2, B1 and B3 confirmed the inputs of the framework (online marketing objectives, and OCS types). Moreover, they were asked on which segmentation types they would focus more when they think of starting with a specific objective and how they would utilize the framework. More specific:

Regarding ***Increase of new customer acquisition*** objective, M1, M2 and B3 commented that technical and referral segmentation types are very important for the first visit of the customer, when actually the optimization of customers' journey in the website starts. As mentioned by one of the interviewees "for *the first contact is very important to categorize new visitors according to their referral and technical characteristics*". As mentioned by M2 and B1 apart from segmenting visitors according to the aforementioned attributes, it would also be useful to take into account their preferences and show them more relevant content in order to complete customer acquisition. Regarding ***Improvement of churn rates*** objective**,** M2 and B3 mentioned that value-based and behavioral characteristics are very important, since they are useful for categorizing customers according to their profitability and past behavior. As mentioned, value-based is more useful for the online retailers, since the calculated values are results of customers purchasing behavior, while for business sites it's easier to identify customers that are more likely to churn by segmenting them according to their behavioral characteristics. M1 mentioned that characteristics that indicate needs and preferences have to be considered as well, after identifying customer's profitability and the risk of churn according to behavioral and value-based characteristics.

As of the ***Increase of Satisfaction*** objective, all four interviewees responded that attitudinal are the first characteristics that they have to be taken into account in order to create segments showing the preferences of the customer. As was commented by one of the interviewees "*Customers should be guided correctly through the channels and get what they need, in order to be satisfied. Therefore characteristics that indicate preferences are very important*." M2 and B1 also consider behavioral characteristics important for the specific objective, as the frequency of the visits or the subscriptions constitute an indication for their satisfaction. B3 mentioned that also loyalty-based segmentation might give an indication of customers' satisfaction, but still when a customer is loyal does not mean that they are necessarily satisfied. Regarding ***Increase of Loyalty*** objective, all four interviewees agreed that loyalty-based segmentation type is very useful for the specific objective. As M1 stated "*a score that indicates the engagement, loyalty or the interest of the client in a certain brand should be always taken into account, so segments according to degree of loyalty can be created*". An engagement or loyalty score can be calculated by setting specific business rules that regard behavioral characteristics. Thus, the interviewees agreed that behavioral segmentation is also important to start with, in order to enhance the increase of loyalty.

Regarding ***Increase of Cross – up Sales*** objective, M2 and B3 mentioned that the first step is to see which customers are most likely to buy products. Therefore, a value-based segmentation and behavioral segmentation is very essential in order to see how much a customer is likely to spend or how willing they are to move on with purchases. M1 mentioned that behavioral segmentation is necessary in order to see what types of products a customer has already purchased, in order to avoid suggesting the same products or be able to decide upon additional offers. All four interviewees mentioned that they would use attitudinal based on segmentation in order to segment customers according to their preferences and suggest new products for purchase. 3 of the interviewees considered demographic segmentation important, but they believe that online customers' characteristics such as age or genre are sometimes difficult or it takes time until they are captured online and added in a visitor's profile. Regarding **Increase of conversion rates**, all the interviewees mentioned that it always depends on the online marketers' point of view and what they consider a conversion in each case. However, all agreed the attitudinal segmentation type and the behavior segmentation are important in most cases. B3 provided examples : " *Preferences and past behavioral is always the first thing you would see, when you consider as conversion visitors clicking on a specific banner, or moving on with purchasing a product or downloading content from your website. However, if somebody new enters referral or technical attributes would also be important to look into. Segmentation types like loyalty or a value-based segmentation, wouldn't add any value*". The low usefulness of value-based and loyalty-based segmentation is basically attributed to the fact that conversion rates are basically related with the activities that an online visitor completes while being on a website.

Technical Experts Evaluation

The technical aspect of the framework was evaluated by A1 and A2. A1 was asked to confirm the results of the literature review on customer segmentation on which the framework is based. Moreover, being an expert as well on the field of Customer Intelligence he was able to judge and confirm the results of the literature review and to give his point of view on which techniques would bear better results in each case. A2, who is a software architect in DEVCORP, responsible for the OCEM software among others, reviewed the second framework and commented on whether the big data techniques would be applicable for each of the OCS types and whether they could be effective or not. Both A1 and A2 agreed with the OCS types and the techniques used for segmentation. More specifically, A2 commented that such a categorization would fit in an OCEM product, which creates online visitors profiles by collecting online attributes, in order to facilitate the segmentation process. Based on his experience A2 mentioned that all classification, clustering, association, and regression are techniques relevant for segmenting customers according to the characteristics that are gathered online. More specific:

For ***classification*** *and* ***clustering*** *techniques,* A1 mentioned that cluster analysis is the most commonly used technique for segmenting customers. It is used when the natural groups that constitute the segments are not known in advance and a clustering algorithm is needed in order to define those groups and categorize the customers. He confirmed that clustering is mostly used for behavioral segmentation type and demographics. Furthermore, he commented that technical and referral segmentation types could be both analyzed by clustering as well as classification techniques. A2 also commented that clustering technique would fit in most cases. Specifically for technical and referral segmentation, clustering in his opinion would be the most useful technique. While bearing some examples he stated "*Clusters can be made out of search engines and referring sites or the keywords that are mostly used*". However, he mentioned that also classification would fit; depending also on the segments that somebody wants to define. Moreover, A1 mentioned "*We use classification when we want the segments to end in a specific result*". A1 confirmed that classification technique is very effective for classifying customers according to their loyalty score. For value-based segmentation A1 does not consider the classification very effective, because normally the marketers want to identify the most profitable clients, without knowing the customer value. However, it can be used in cases for behavioral segmentation as well. Similarly, A2 mentioned that he considers classification mostly useful when there is a dependent variable according to which segments can be created.

When asked for ***association****,* ***regression*** *and* ***visualization,*** A1 mentioned that these are not straightforward segmentation techniques, but there are several analysts and scientists who might treat them as such. Association is used for recommendation engines, and especially for basket analysis as it happens in the case of Amazon.com. Association is always useful to find out what visitors prefer to buy. A1 commented that comparing to offline, in an online environment it is much easier to track characteristics that indicate preferences and attitudes of the clients. In such cases association analysis is useful for characteristics that indicate customer preferences (attitudinal segmentation type). Respectively, A2 considers association analysis very helpful for creating recommendation and for micro-segmentation according to the user preferences. For regression, A1 mentioned that it cannot be used alone for segmentation. However, it estimates the probability that a customer belongs in a certain segment. It is often used to estimate churn probability and, thus, segment customers to churners or non-churners, based on value-based characteristics. For visualization A1 confirmed that visualization is very important for illustrating segments providing clear view, and is similar to cluster analysis. A2 also mentioned that visualization is very important for visualizing customer segments, especially when the amounts of customer data are huge.

Conclusion & Future Work

The main objective of the research was to answer two research questions: (i)Which are the customer segmentation types that can assist each of the business goals regarding online marketing? and (ii) Which big data approaches and techniques can be used for each Online Customer Segmentation type? The answer of these research questions resulted in the creation of a framework which provides a holistic overview of how OCS can support online marketing with the use of analytic approaches applicable in extremely large amount of online customer data. The framework is high level and can constitute the dawn for further and more detailed research in the field of Big Data and its application in online marketing.

Regarding the first research question, it was found that most of the customer attributes gathered online match with those gathered offline and fall under the main customer segmentation categories, as defined from traditional customer segmentation theory. These are: attitudinal, behavioral, demographics, loyalty-based, value-based. Moreover, two new segmentation types that include attributes gathered only online are the referral and the technical. Based on the literature and the discussions with the experts, it is understood that effective OCS plays an essential role for achieving online marketing objectives. The proposed framework shows the usefulness of each OCS type for each of the aforementioned online marketing objectives. Behavioral and attitudinal segmentation types are of high usefulness for most of the objectives.

In order to answer the second research question a literature study on big data and data mining techniques for customer segmentation was conducted. As it was found the techniques that can be applied on big data, stem from data mining and statistics. The difference lies on the characteristics of the data, while the analytic techniques that are used for big data analysis do not differ from those used for traditional data mining. The difference exists in the tools that are able to handle and analyze the large volume and the diversification of the data. Furthermore, the literature review showed that the main techniques that are also applicable in big data, for customer segmentation are: Clustering, Classification, Visualization, Regression, and Associations. Clustering and Classification are the techniques that are used more often and can be applied to the most of the customer segmentation types**,** while the rest of the techniques are not straight segmentation techniques, but their results can be used for customer segmentations. All techniques can be implemented on big data with the use of big data analytic tools, such as Apache Mahout or the R Language.

This research raised the following issues: (i) OCS is a core process for assisting an online marketing strategy. However, there is limited scientific research related to the field; (ii) huge amount of online customer data are continuously generated. However, there was no scientific research found for the use of big data tools in OCS; (iii) in the world of business a gap between online marketers and data analysts emerges. Normally, online marketers should be able to select combinations of OCS types, that would better serve their needs and goals, and then data analysts are called to do the analysis and provide them with the best solution. However, online businesses do not have yet well-defined online marketing goals. Therefore, marketers do not provide data analysts with the appropriate information, thus, proceeding with valuable and effective OCS becomes difficult.

It is apparent that a general guideline for effective OCS is needed, while the opportunities that big data offer for OCS should be further explored. The main deliverable of this research is a framework consisting of two parts; the first part showing OCS types able to assist online marketing objectives, while the second part shows which techniques that can be used for big data analysis are suitable for segmenting online customers according to each of, the OCS types. The framework could constitute a first step towards an effective OCS approach capable of assisting an online marketing strategy.

In future research, the frameworks could be tested as a whole on more than one real situation. Firstly, the actual usefulness of OCS types for each of the objectives, according to the framework can be tested. Starting with an online marketing objective certain OCS types can be selected to be analyzed for creating actionable customer segments. After the OCS types that should be analyzed have been chosen, the implementation of big data techniques and tools as should be tested. Big Data tools can be used in order to implement techniques to analyze the OCS types and create effective customer segments. Each of the techniques should be tested in order to find out what bears the best result in each case, and consequently model these situations into a coherent overview to help analysts select the most appropriate technique for each case (*e.g.* Vleugel, Spruit and Daal, 2010). This would require a longitudinal research since the actual effectiveness of the OCS in online marketing objectives can only be unveiled in the longer term. Furthermore, integration of the wealth of semi-structured (social) webdata should be explored to further inprove online customer segmentation, which will certainly raise the bar with respect to *big* data opportunities (*e.g.* Otten and Spruit, 2011). Finally, more OCEM tools can be observed in search of additional OCS types.

References

Anil, R., Dunning, T., & Friedman, E. (2012). *Mahout in action*. Shelter Island: Manning.

Chaffey, D. Chadwick-Ellis, F., Mayer, R., Johnston, K. (2009), Internet Marketing: Strategy, Implementation and Practice, Essex: Prentice Hall.

Chaffey, D., & Smith P.R. (2008). eMarketing eXcellence: Planning and optimizing your digital marketing. Oxford: Elsevier

Chan, C. C. H. (2005). Online auction customer segmentation using a neural network model. *International Journal of Applied Science and Engineering*, *3*(2), 101-109.

Chan, C., C., H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications* 34(4): 2754–2762.

Chen, H., Chiang, R., H., L., & Storey, V., C. (2012). Business Intelligence and Analytics: From Big Data to big impact, *MIS Quaterly* 36 (4), pp.1165-1188

Cheng, C.H., & Chen, Y.S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*(3), 4176–4184.

Chen, Y., Guozheng, Z., Dengfeng, H. & Chua, F. (2007). Customer segmentation based on survival character. *Journal of Intelligent Manufacturing* 18(4): 513–517.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases." *Al Magazine*, Vol. 17(3), pp. 37-54

Farooqi, R., & Raza, K. (2011). A Comprehensive Study of CRM through Data Mining Techniques. *Business*, *2*(i).

Fotaki,G., Spruit, M., Brinkkemper, S., & Meijer, D. (2013). *Exploring Big data opportunities for Online Customer Segmentation*. Technical report UU-CS-2013-021, Department of Information and Computing Sciences, Utrecht University.

Fotaki, G., Gkerpini, N., Triantou, A, I., & Brinkkemper, S. (2012). Online Customer Engagement Management. Utrecht University.

Gartner (2011). Big Data definition. Retrieved on Dec 6, 2013, from http://www.gartner.com/it-glossary/big-data/.

Hong, T. & Kim, E. (2010). Segmenting customers in online stores from factors that affect the customer's intention to purchase. *International Conference on Information Society*, *39*(2), 383–388.

Hosseni, M, .B, Tarokh M. J. (2011). Customer Segmentation Using CLV Elements. *Journal of Service Science and Management* 04(03): 284–290.

Jansen, S. M. H. (2007). Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior. *A Vodafone Case Study.*

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*(1), 101-107.

Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, *13*(3), 297–310.

Lee, J., & Park, S. (2005). Intelligent profitable customers' segmentation system based on business intelligence tools. *Expert Systems with Applications*, *29*(1), 145–152.

Leung, H., C. (2009). An Inductive Learning Approach to Market Segmentation based on Customer Profile Attributes. *Asian Journal of Marketing.*

Manyika, J., Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. "Big data : The next frontier for innovation, competition, and productivity." *McKinsey Global Institute* 364(May): 156.

Miguéis, V. L., Camanho, A. S., & Falcão e Cunha, J. (2012). Customer data mining for lifestyle segmentation. *Expert Systems with Applications*, *39*(10), 9359–9366.

Otten,S., & Spruit,M. (2011). *Linguistic engineering and its applicability to business intelligence: towards an integrated framework.* International Conference on Knowledge Discovery and Information Retrieval (pp. 460–464). Paris, France: SciTePress.

Pachidi,S., Spruit,M., & Weerd,I. van der (2014). Understanding Users' Behavior with Software Operation Data Mining. *Computers in Human Behavior*, *30*, 583–594.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77.

Peterson, E. T. (2004). Web analytics demystified: A marketer's guide to understanding how your web site affects your business. Celilo Group Media & CafePress.

Peterson, E.T. (2006). The Big Book of Key Performance Indicators. *Web Analytics Demystified.*

Pillai, J., & Vyas, O. P. (2012). CSHURI – Modified HURI algorithm for Customer Segmentation and Transaction Profitability, *2*(2), 79–89.

Russom, P. (2011). Big Data Analytics. *TDWI Best Practices Report,* Fourth Quarter.

Sathi, A. (2012). Big Data Analytics: Disruptive Technologies for Changing the Game. USA: MC Press Online.

Stroud, Dick. (2006). Customer Intelligence. *Journal of Direct Data and Digital Marketing Practice* 7(3): 286–288.

Tsiptsis, K., & Chorianopoulos, A. (2010). *Data Mining Techniques in CRM: Inside Customer Segmentation*. Wiley.

Turban E., Sharda, R., Delen D., & King, D. (2010). Business Intelligence: A Managerial Approach.(2nd ed.) New Jersey: Prentice Hall

Van Doorn, J.,Lemon, K.,N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P., C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research* 13(3): 253–266.

Vleugel,A., Spruit,M., & Daal,A. van (2010). Historical data analysis through data mining from an outsourcing perspective: the three-phases method. *International Journal of Business Intelligence Research, 1*(3), 42–65.

Winer, R. S. (2001). A framework for customer relationship management. *California management review*, *43*(4), 89-105.

Woo, J., Bae, S., & Park, S. (2005). Visualization method for customer targeting using customer map. *Expert Systems with Applications*, *28*(4), 763–772.

Wu, R.S & Chou P.H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications* 10(3): 331–341

Ye, L., Qiuru, C., Haixu, X., Yijun, L., & Guangping, Z. (2013). Customer Segmentation for Telecom with the k-means Clustering Method. *Information Technology Journal*, *12*(3), 409-413.

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a Service and Big Data. *Proceedings of the International Conference of Advances in Cloud Computing. 8: 21-29*

Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). Large-scale parallel collaborative filtering for the 19etflix prize. In *Algorithmic Aspects in Information and Management* (pp. 337-348). Springer Berlin Heidelberg.