Working Papers

Working Paper No. 28

Challenges of Data Management and Analytics in Omni-Channel CRM

Heike Trautmann, Gottfried Vossen, Leschek Homann, Matthias Carnein, Karsten Kraume

ISSN 1614-7448

cite as: Heike Trautmann, Gottfried Vossen, Leschek Homann, Matthias Carnein, Karsten Kraume: Challenges of Data Management and Analytics in Omni-Channel CRM. In: Working Papers, European Research Center for Information Systems No. 28. Eds.: Becker, J. et al. Münster 2017.

Contents

List of Figures

Working Paper Sketch

Type

Research Report

Title

Challenges of Data Management and Analytics in Omni-Channel CRM.

Authors

Heike Trautmann, Gottfried Vossen, Leschek Homann, Matthias Carnein, Karsten Kraume
contact via firstname.lastname@ercis.de and Karsten.Kraume@uni-muenster.de

Abstract

Data Management and Data Analytics are of huge importance to Business Process Outsourcing Providers in Customer Relationship Management (CRM) in order to offer tailor-made CRM Solutions to their business clients during presales, sales and aftersales. These solutions support business clients to improve their internal processes, as well as their customer service in a variety of communication channels (including e-mail, chat, social media, private messages, etc.) to reach out to end customers in an efficient way. As customer interactions may happen via various channels basically at any time, a crucial challenge is to efficiently store and integrate the data of the various channels in order to obtain a unified customer profile. This paper abstracts from the underlying platforms and instead considers the requirements to CRM solutions for the various communication channels, in order to devise a uniform and corporate-wide data architecture for an omni-channel customer view to maximize the business clients' value in customer retention and customer centric analytics. Especially, online customer segmentation integrating channel usage and preferences is presented as a very promising means for constructing a self-energising information loop which will lead to highly improved customer service along the whole customer journey.

Keywords

Omni-Channel CRM, Big Data, Customer Segmentation, Data Architecture

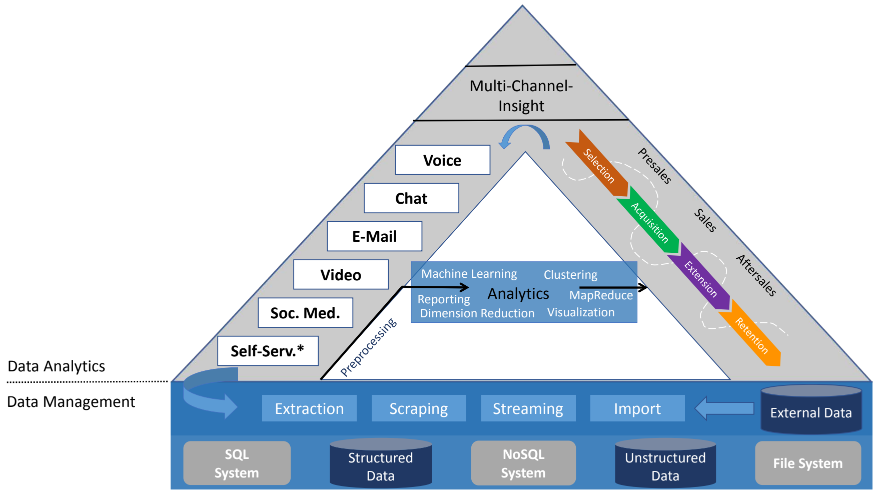


Figure 1: Analytics Framework

1 Introduction

Nowadays customers can be connected to the web basically anytime, anywhere and as long as they wish. Interaction with companies happens via different channels with varying proportions and frequency. Therefore, customer relationship management (CRM) faces specific challenges in order to meet customer expectations with regard to ensuring transition-free communication across all channels, ideally with unlimited availability [11]. Companies' organization structures have to be adapted in that traditional models of handling each channel separately in terms of personnel, management and strategy must be reassessed. Efficient interweaving of channels supplemented by a sophisticated usage of customer data has become a key company USP. One focus of this paper is on the requirements of a corporate-wide data architecture for a comprehensive omni-channel customer view to maximize the business clients' customer insight leading to improved customer service and business clients' success. We will further focus on the information resp. data perspective in terms of a specific regulation framework for analytics (see Figure 1).

Customer information is now distributed across different, possibly channel-specific platforms, in an asymmetric and dynamic fashion. Moreover, brand-specific marketing messages have to be delivered consistently across all different channels.

An overall company-specific reference model has various facets and has to focus on client as well as internal matters. Of course, client-owned systems and platforms are an important part, including sophisticated platform management, both externally as well as internally. Due to the fact that most tailor-made customer CRM solutions only cover a subset of communication channels the end customer related data is stored isolated on each platform. This results in the challenging task of combining the data of the various solutions in order to obtain a unified customer profile.

Efficiently stored, curated and analysed customer data offers huge potential for improving omni-channel CRM. For example, identified homogeneous customer segments enable sophisticated routing strategies to service agents, both inside a channel and across channels. Analytics play a key role in the whole framework by extracting relevant information from the present 'big data', as well as storing and analysing it efficiently in order to achieve maximal insight into customers' behaviour and needs along the whole customer journey [20]. By these means presales, sales and aftersales activities are heavily supported in a data-driven way from an omni-channel perspective.

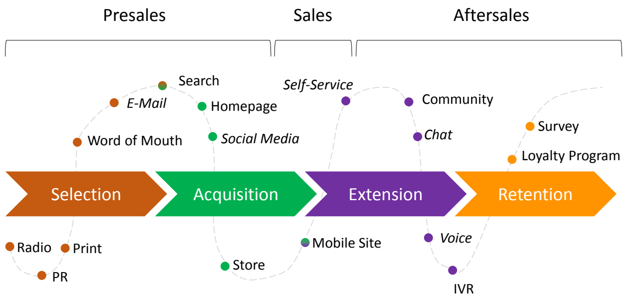


Figure 2: Customer Journey

According to Gartner [1], from a BPO's perspective, a custumer journey comprises the four phases: customer selection, customer acquisition, customer extension and customer retention. Throughout this paper we use a more general approach of three phases, which does not contradict Gartner's categorisation but subsumes it as illustrated in Fig. 2. More specifically, presales involves customer selection and acquisition, sales involves customer acquisition and extension, and aftersales involves customer extension and retention.

Presales represents the first step to interact with a potential customer. In doing so the tasks of presales are comprehensive: E.g., customer advisory service, feasibility studies for potential clients and proactive marketing activities. Social media engagement CRM solutions could be an important means in this setting. Undecided customers could be supported in their purchase decision by offering a proactive chat solution to eliminate nescience and dispel any doubts. Additionally, newsletters via mail could be sent or potential customer called. The *sales* phase includes the development and signing of a contract between a customer and the sales department. In case this step is outsourced it primarily happens remotely in one or several of the multiple channels. From a data management point of view this step is crucial and technically requires transactional guarantees.

Based on the experience that it is easier and cheaper to retain a customer than to acquire a new one, *aftersales* primarily aims at customer retention. Although essentially the same communication channels are used as in the presales and sales steps, the main difference is that customer data is available. In other words information, e.g., contact, used product, and customer domain are at hand to create individual retention (or win-back) strategies. Essential instruments of a retention strategy include offering new products or suggest additional services. To benchmark customer satisfaction within the social media network social media monitoring can be used to measure values like response times or response rates [7]. In this study on the airline industry, we have collected several million customer service requests from Twitter and Facebook and automatically analyzed them for this purpose. During win-back of a customer, it is necessary to understand why the customer decided to stop the collaboration [25], e.g., by interviews, web forms or personal phone calls. Another aim of aftersales is the support of the customer in case of questions, complaints, or issues, mostly by voice service.

The remainder of this paper is structured as follows: In Section 2 the challenges to achieve an omni-channel data management system are described by high level requirements which the desired data architecture has to meet. Section 3 introduces the analytics framework and exemplary illustrates the

[1] http://www.gartner.com

potential of customer segmentation. Moreover, the respective underlying cluster analysis and streaming techniques are described. Section 4 concludes the paper and provides an outlook on further research perspectives.

2 Data Architecture and Data Management

The overall basis for any meaningful analytical technique is formed within the data storage level which, in principle, is not limited to any specific kind of database or file system, but rather depends on the type, availability and frequency of the considered data. Company-internal as well as external data have to be made available upon request, regardless of its origin, leading to maximum information load across channels. The kind of data import heavily depends on the type, size and dynamics of candidate data. Data can be structured in terms of 'spreadsheet data' containing customer information stored in a column-based fashion. Especially external data, such as crawled social media communication, i.e., text-based unstructured data, might have to be treated with streaming technology, which has a direct impact on the kind of analytical methods used later on. Moreover, it could be appropriate to solely save aggregated instead of the original raw data, due to limited storage capacity.

In the following, requirements to modern data management systems for the CRM solutions offered by a customer service provider in the omni-channel context are discussed without considering the underlying platforms themselves. This abstraction will help us to exhibit the appropriate characteristics for devising a uniform, corporate-wide data architecture. In general, solutions support clients in improving their internal processes, as well as their customer service along the customer journey during the phases presales, sales and aftersales. Examples of such solutions are social media monitoring, social media engagement, chat, video-assisted sales, voice service or e-search and virtual assistance for customer service in various communication channels. A CRM solution itself thus can be considered as a compilation of human resources and platforms.

At first glance these solutions seem to be a suitable approach for business clients to adapt or extend their customer service as needed, but on the other hand hinder implementing a comprehensive omni-channel customer view because of the following reasons: First, a CRM solution does not have to consist of the same platforms across countries. For example, this means that a solution for social media engagement can be offered with Software A in Germany, but with Software B in France. Moreover, a solution in most cases only covers a subset of possible communication channels, which increases the problem to utilize and exploit data in the most adequate way in every context. Moreover, Software A may be capable to support the channels social media, chat and email, but the business client may only be interested in social media. Thus platform integration and management is a central issue.

Challenging are also the usually different data representations across the solutions, which lead to a variety of formats, ranging from highly structured (relational) data, say, from the customer's purchase history, to entirely unstructured data that stems from a chat or a voice or text message; intermediate formats, e.g., in an XML dialect, are also possible. These formats need to be properly integrated and be made available to analytics. Typically, such an integration requires data cleansing, since not all data delivered may be of acceptable or of similar quality. Indeed, the data integration steps needed resemble the ETL (extraction, transformation, loading) process commonly used for data warehouse applications. Additionally, data need to be processed by a suitably chosen architecture which is capable of handling the various types and formats in such a way that ultimately the CRM goals can be met. Besides that it must be mentioned that most business clients have their own solutions for data management of their business context, e.g., for storing transactional data or user data.

Summarising, specific data management poses a variety of challenges, both from a technical, economic, organizational, and sometimes even from a legal point of view. Data must be up-to-date, must ultimately provide (or at least contribute to) a 360-degree-view of the customer, and must be readily available during any form of customer contact. In addition, data should be processed in such a way that not only a vastly complete customer profile can be established but that would also allow predictive as well as prescriptive measures for the near to medium-term future.

2.1 Data Management Requirements

A crucial goal is to provide comprehensive analytics in the aftersales phase, yielding insight into areas such as customer value, win-back concepts, or even internal processes as captured in the internal regulation framework. Since the same channels can be used within the different sales phases of a customer journey, a generalization from aftersales by including the other two phases is required. The various channels used in the execution process produce digital data in the form of voice recording files, chat recordings, e-mails, mail scans, social media posts, tweets, or conversations, short messages, or even contact forms into which a user can write a message. All these data need to be associated with the single customer to which it pertains, and the data ideally give rise to a customer profile into which past, current, and future data (predictions) can be seamlessly integrated. This way, the customer service representative of the customer journey will be in the position to recognize the customer efficiently, to know about his or her concerns, issues from the past, and how best to treat that customer during current contact (next best action / next best channel). From this, a number of requirements result that the underlying data management system has to meet:

Data persistency The various channels suffer from isolated data storage, which hinders a uniform data management approach across all steps of the customer journey, although the processes per channel are similar. Therefore, it is of high importance for a data management system to be capable of storing the metadata that appears on each communication channel, as well as the channel specific information, e.g., mail attachments, voice recordings, chat history.

Unification of various data formats Albeit the data management system could store the data in a uniform way, the problem to integrate external sources with various data formats does not disappear. For example, Facebook messages and posts must be collected separately. Other examples are metadata inside an SaaS application or external systems owned by the business clients. This could be, for instance, airline booking information in a central booking system when the service provider only handles support for their clients, resp. their client's customers.

High data quality The information from the various channels and formats must run through an ETL process to receive high quality data. The transformation process is critical at this point. Data Preprocessing will be discussed in more detail in Section 3.

Backup and access The large number of channels leads to a huge amount of data which must be saved in a fail-safe and time-efficient manner. Hence, the data management system must support replication to guarantee access even when one system fails. Additionally, partitioning must be supported to achieve fast access to the data that is shared across multiple machines as disjoint portions.

Fast retrieval Fast retrieval is important to query data from the data sources. Especially during the communication with the end customer the customer service representative must receive all data in an appropriate time interval by the data management system. To communicate directly with the data management system an adequate query language must be available.

Long-time archival Long-time archival offers the possibility to store all kind of customer information in a centralized way for later analytics, e.g., reporting or long-term customer development observations. Due to the size of historical data HDFS seems to be an appropriate solution for this requirement. In combination with MapReduce it is a powerful framework.

High availability High availability is a requirement that is critical for 24/7 customer support. Each step of the execution model needs access to the customer information at any time. In other words the database must be failsafe as discussed previously and therefore has to be in the CA or AP classes of the CAP-theorem (consistency, availability, partition tolerance).

Transaction safety In order to assure that all systems have the same information instantly available, it is very important to guarantee transaction safety. Especially in the sales phase process it is crucial to store user related information, e.g., when a contract was finished.

Data security and privacy Another important requirement poses data security and privacy. This includes data storage as well as data access. Regarding data storage it must be transparent to the client where the sensitive customer data is hosted, e.g., using a cloud service or on premise. Additionally, it must be guaranteed that only privileged services are allowed to access or process the data.

These high-level data management requirements offer the perspective for identifying technical data management requirements as well as systems able to meet them. Thereby regarding storage requirements distributed storage of large amount of data and relational, as well as in-memory databases systems for advanced analytics can be considered and data management systems for processing distributed data and traditional reporting have to be taken into account. The integration of all these systems is reflected in the regulation framework for analytics in Fig. 1.

3 Regulation Framework Analytics

Analytical techniques require sophisticated data preprocessing. While this is true for any kind of analysis, it becomes increasingly important when dealing with huge volumes of data from different sources, as simple visual techniques will fail to detect irregularities in varying dimensions and one will definitely loose track on irregular data due to the dynamic environment. Incorrect or missing data will lead to biased results, following the well-known 'garbage in – garbage out' principle. Therefore, outlier detection technique [17] and methods capable of handling and imputing missing data [30] are essential in this context. Dimensionality reduction techniques play a central role in reducing the data volume in a systematic and meaningful way focusing on the most relevant information.

Unstructured data has to be preprocessed in a sequential manner. In order to retrieve information about customer characteristics and behaviour, text-based communication has to be converted into analysable features, e.g., using natural language processing or specific techniques from communication studies such as bag-of words approaches, topic modelling, document clustering or semi-automated content analysis (see, e.g., [19]).

In general, customers can be characterised using different perspectives, such as [26]

- *Geographic:* Country, City, Region, Climate, City size, ...
- *Demographic:* Age, Sex, Marital status, Income, Education, Occupation, ...
- *Psychological:* Personality, Risk perception, Attitude, ...

- *Psychographic:* Lifestyle, Interests, Opinions, Motivation, Activities, ...
- *Socio-Cultural:* Culture, Religion, Ethnicity, Social class, ...
- *Use-related:* Usage rate, Brand loyalty, Buying History, ...
- *Usage Situation:* Time, Location, Objective, ...

These variables can either be continuous (e.g., age or income), discrete (e.g., marital status or religion) or even binary (e.g., new or old customer). While some characteristics are easily quantifiable, others, such as personality or lifestyle can be difficult to determine or quantify.

An omni-channel CRM framework provides potential to supplement these characteristics by *channel-specific aspects*, such as channel preferences, both generally as well as product-related. Moreover, dynamics over time should be addressed as well as customer satisfaction and complexity of customer personality extracted from channel communication. Storage and integration of channel-related data are crucial to meet customers needs in the most appropriate way. Similar to storing external data, it has to be distinguished between data streaming and traditional data import.

Data analytic techniques can be applied with different foci in mind and influence all stages of the customer journey. While *customer segmentation* plays a key role and will be exemplary detailed later on, *customer value*, *customer retention* and *win-back* analyses are also central issues in analytical CRM. Of course, these aspects are interrelated, as e.g., win-back potential is based on customer potential derived from the corresponding customer value.

Another aspect which must not be underestimated is the possibility of systematically *analysing company-internal processes*, such as the process and the quality of routing customers across channels and to specific call-center agents having the required expertise. Moreover, also quality and potential of omni-channel CRM platforms can be addressed as well as human resources planning. These kinds of analyses can be conducted on different hierarchical levels, such as client-specific, or call-center or country-specific.

3.1 Omni-Channel Insights related to Customer Journeys by Means of Customer Segmentation

Since its introduction in the late 1950s, market and customer segmentation have received considerable attention in practice and research. Since then, it has grown to become an essential tool in marketing, i.e., especially presales, customer value analysis and potential (after sales) and CRM in general. The initial concept of market segmentation has been introduced by [28], who states that 'market segmentation [...] consists of viewing a heterogeneous market [...] as a number of smaller homogeneous markets in response to differing preferences among important market segments'. In other words, market segmentation is the separation of a heterogeneous market with diverse preferences and behaviour into distinct subsets of homogeneous groups that share similar interests and behaviour. The acknowledgement and identification of customer segments makes it possible to target each group with appropriate value propositions and marketing strategies [6].

After the identification and characterization of different market segments, it is important to identify which of the segments are attractive for the company. The most attractive segments should be selected and targeted with a specific value proposition or marketing mix. Often, six criteria are used that are necessary for an effective targeting of a market segment [32, 26]:

- *Identifiability* Marketers should be able to identify relevant characteristics that can reveal marketing segments. Additionally, it should be possible to associate customers with a given segment using these characteristics.

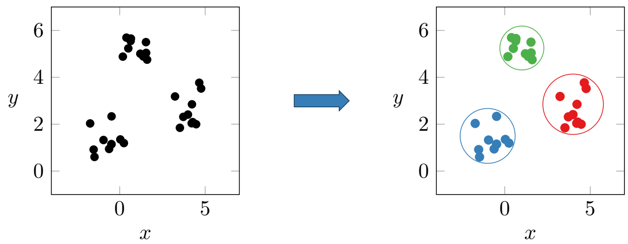


Figure 3: Cluster Analysis on customer characteristics x and y.

- *Sustainability* Targeted marketing for the segments should be profitable enough to sustain the segmentation efforts. Specifically, a segment should comprise enough customers to warrant targeted value propositions.

- *Accessibility* It should be possible to reach segments with offers and promotions. This requires that there are channels and platforms where relevant customers can be addressed and reached.

- *Stability* Segments should be relatively stable over time such that segments do not change during the development or deployment of marketing efforts.

- *Responsiveness* A segment should respond uniquely to marketing efforts. This requires that the different segments are distinct and do not overlap.

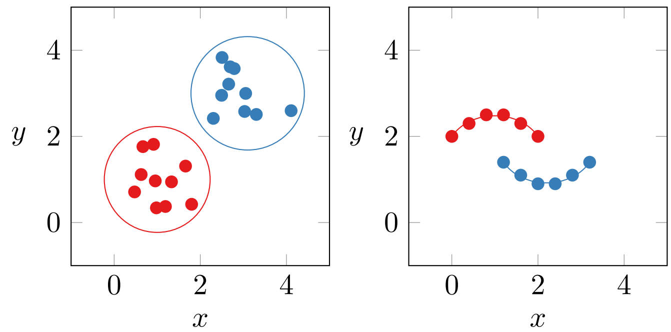
- *Actionability* Segments should be in line with the companies objective and resources. For some companies, certain segments can be out of scope, despite being attractive market groups.

Market resp. customer segmentation virtually exists in every large company. However, often this process is highly informal, and the marketing team will determine customer segments based on experience and intuition [6]. Nevertheless, more sophisticated and theoretically founded approaches have been developed to find market segments, sometimes called data-based market segmentation [6]. Among the most popular approaches are cluster analysis and mixture models [32], possibly combined with specific approaches to data streaming.

3.2 Cluster Analysis and Streaming Techniques

Cluster analysis is a technique in statistical data analysis which aims at identifying subsets or groups of objects resp. customers in this case, also called *clusters*. The goal is to find clusters for which the similarity of objects within the cluster is high, whereas similarity of objects across clusters is low. The number of clusters can either be fixed a-priori (e.g., based on domain knowledge) or determined post-hoc (e.g., based on the result of the partitioning) [32]. Figure 3 visualizes the main idea behind clustering. In the example, three separate clusters can be identified.

To determine the similarity or dissimilarity of objects, a distance measure between different customers and their associated variables is required in order to assess the degree of similarity. Popular distance measures for continuous variables are the Euclidean distance or the Manhattan distance. A problem of cluster analysis is that there is no objective performance criterion, which makes it difficult to assess the performance of different solutions so that clustering is referred to as an unsupervised learning technique. Instead, a variety of different objectives and criteria is available. Clustering algorithms determine the clusters by minimizing one of these objectives for all feasible clusterings. An example of such an objective is to minimize the within-cluster sum of squares, i.e., to minimize the variance within clusters. This approach aims to find a compact solution and is visualized in Figure 4a. An alternative goal is to find connected data points, as shown in Figure 4b. Popular clustering algorithms are the k-means [21] algorithm and agglomerative clustering algorithms [31].



a Compactness b Connectedness

Figure 4: Example: different types of clustering objectives [13]

From an optimization perspective, clustering can be considered an NP-hard grouping problem [16]. For this reason, many algorithms have been developed to find an approximation of the 'optimal solutions'. In particular, evolutionary algorithms [9] have been applied in the context of cluster analysis. Evolutionary algorithms apply a 'survival of the fittest' strategy that is population-based and inspired by nature and human evolution. Different clustering criteria are often contradicting which makes it difficult to find a single best clustering solution. For this reason, multi-objective clustering algorithms have been developed. These algorithms aim to find the best trade-off solutions between multiple criteria. Not that the set of trade-off solutions to a multi-objective clustering problem always comprises the optimal solutions to the single-objective clustering problems. One of the most popular multi-objective clustering algorithms is `MOCK` [13] which is based on a multi-objective evolutionary optimization algorithm.

To improve clustering quality, the number of selected variables containing customer-related data should be limited. This is necessary since almost all pairs of points will be at a similar distance in a high-dimensional space. In addition, it can be useful to remove correlated variables. To address both of these issues, the data could be transformed using a Principal Component Analysis (PCA) [24, 15] which expresses possibly correlated variables as a set of linearly uncorrelated variables.

Streaming Techniques Traditional cluster analysis assumes a fixed data set to be analysed. In practice, however, new customers are regularly added to the system and new information for existing customers becomes available dynamically. In addition, external information extracted from the Web or social media tools might be streamed into the data base. This requires that customer segmentation is regularly repeated to account for changes in the underlying data set. Unfortunately, this analysis might be computationally costly and a more suitable approach is to update the existing clusters as new data points arrive. This is the goal of stream clustering [27, 22, 3], where data are assumed to arrive in a constant stream of new examples where the order of examples cannot be influenced. Since this stream is possibly unbounded, many algorithms also assume that the data cannot be stored in its entirety and examples can only be evaluated once. This may be especially relevant due to data privacy and security issues mentioned in Section 2.1, not only in the social media context.

Such data streams are a challenging task since they require to build models than can be incrementally updated when new data points arrive. To solve this problem, stream clustering algorithms utilize a two-step process. First, an online component constantly evaluates the data stream and captures summary statistics that describe the examples in the stream. This can be seen as a data abstraction step where only a summary of the entire stream is generated. Typically, the result of this analysis is a number of *micro-clusters* that contain the summary statistics for many preliminary groups in the data. The number of micro-clusters is much smaller than the number of data points in the stream but considerably larger

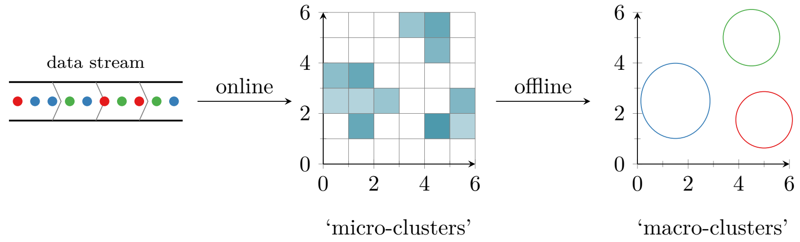


Figure 5: Two-phase stream clustering approach

than the final number of clusters. Whenever necessary, an offline component can use the summary statistics to derive the final clusters, called *macro-clusters*. Since micro-clusters are much smaller and often of bounded size, it can be assumed that this data fits into main memory such that any existing clustering algorithm can be used to generate the clustering. Fig. 5 visualizes the stream clustering concept.

One of the earliest stream clustering algorithms that still receives considerable attention today is `BIRCH` [33]. Since then, many algorithms have been developed to find clusters in data streams, most notably `STREAM` [23, 12], `ClusTree` [18], `CluStream` [2] and `StreamKM++` [1].

3.3 Customer Profiles – Characterization of Clusters

A crucial step is to characterise the individual clusters in order to understand the reasons why the grouped customers are perceived as homogeneous in their characteristics and behaviour. For this purpose, statistical and machine learning techniques [14] generate a model predicting the membership of a customer to a specific group by means of customer characteristics reflected by the variables considered for clustering (e.g., demographic or channel-specific information). Random Forests [5] or Support Vector Machines [29] are candidate approaches for which also online variants exist [10, 4, 8] for efficiently handling huge data volumes. By means of integrated feature selection approaches, the most important characteristics distinguishing the different clusters can be detected. This ideally leads to customer profiles per cluster which enables addressing and handling specific types of customers appropriately at milestones of their customer journey. Also, routing of customer requests to most adequate service agents is facilitated.

Summarizing, within the omni-channel CRM context, *channel-specific insights and channel usage of customers can be integrated into market segmentation which supplements traditional analyses in an extremely informative manner*. Thus, omni-channel CRM functions as an interface to customer segmentation in two directions. On the one hand, having the knowledge of a specific segment the customer belongs to enables an efficient routing within the channel and agent architecture. On the other hand, channel usage analysis is used for updating and improving the existing customer segmentation leading to *a self-energising information loop*. By means of clustering streaming techniques information can be constantly upated. Moreover, marketing activities can be specifically tailored to specific customer segments using the most appropriate channel(s).

4 Summary

Related to the customer journey, high-level data management requirements for business outsourcing providers in CRM are derived ensuring that the huge potential of omni-channel CRM solutions can be efficiently exploited. Based on this, technical data management requirements as well as appropriate systems can be identified. Distributed storage of large amount of data and relational, as well as non relational database systems for advanced analytics have to be considered. Along with that, data management systems for processing distributed data and traditional reporting have to be taken into account.

Insight into the dynamics and characteristics of omni-channel customer relationship management can be substantially increased by means of appropriate analytical techniques. Systematically and efficiently applied, a self-energising information loop is obtained having a direct impact on the generation and understanding of customer profiles and segments which can directly and efficiently used along the customer journey, i.e., in presales marketing activities, sales and aftersales.

In future work high level requirements for data architecture and data management have to be substantiated into requirements for the various communication channels to achieve an appropriate omni-channel architecture. This includes, e.g., the investigation and evaluation of the generated data volume and structure across the various channels. Customer segmentation will be conducted on real world customer data specifically taking into account customer channel preferences.

References

[1] Marcel R. Ackermann, Marcus Märtens, Christoph Raupach, Kamil Swierkot, Christiane Lammersen, and Christian Sohler. Streamkm++: A clustering algorithm for data streams. *J. Exp. Algorithmics*, 17:2.4:2.1–2.4:2.30, 2012-05.

[2] Charu C. Aggarwal, Jiawei Han, Jianyong Wang, and Philip S. Yu. A framework for clustering evolving data streams. In *Proceedings of the 29th International Conference on Very Large Data Bases*, volume 29 of *VLDB '03*, pages 81–92. VLDB Endowment, 2003.

[3] Amineh Amini, Teh Ying Wah, and Hadi Saboohi. On density-based data streams clustering algorithms: A survey. *Journal of Computer Science and Technology*, 29(1):116–141, 2014.

[4] A. Bordes, S. Ertekin, J. Weston, and L. Bottou. Fast Kernel Classifiers with Online and Active Learning. *Journal of Machine Learning Research*, 6:1579–1619, September 2005.

[5] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[6] Francis Buttle. *Customer Relationship Management: Concepts and Technologies*. Elsevier Butterworth-Heinemann, 2009.

[7] M. Carnein, L. Homann, H. Trautmann, G. Vossen, and K. Kraume. Customer Service in Social Media: An Empirical Study of the Airline Industry. In *Big Data Management Systems in Business and Industrial Applications 2017, accepted*, 2017.

[8] E. Y. Chang, K. Zhu, H. Wang, H. Bai, J. Li, and Z. Qiu. PSVM: Parallelizing Support Vector Machines on Distributed Computers. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

[9] Agoston E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. SpringerVerlag, 2003.

[10] T. Glasmachers and C. Igel. Second-Order SMO Improves SVM Online and Active Learning. *Neural Computation*, 20(2):374–382, 2008.

[11] Beatrice Grimonpont. The impact of customer behaviour on the business organization in the multichannel context (retail). *Journal of Creating Value*, 2(1):56 – 69, 2016.

[12] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O'Callaghan. Clustering data streams: Theory and practice. *IEEE Transactions on Knowledge and Data Engineering*, 15(3):515–528, 2003-05.

[13] Julia Handl and Joshua Knowles. An evolutionary approach to multiobjective clustering. *IEEE Transactions on Evolutionary Computation*, 11(1):56–76, 2007-02.

[14] T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, 2nd edition, 2009.

[15] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417 – 441, 1933.

[16] Eduardo R. Hruschka, Ricardo J. G. B. Campello, Alex A. Freitas, and André C. P. L. F. de Carvalho. A survey of evolutionary algorithms for clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 39(2):133–155, 2009-03.

[17] A. Koufakou and M. Georgiopoulos. A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. *Data Mining and Knowledge Discovery*, 20(2):259–289, 2010.

[18] Philipp Kranen, Ira Assent, Corinna Baldauf, and Thomas Seidl. Self-adaptive anytime stream clustering. In *Ninth IEEE International Conference on Data Mining (ICDM '09)*, pages 249–258, 2009-12.

[19] K. Krippendorff. *Content Analysis: An Introduction to its Methodology*. Sage, Thousand Oaks, 2. edition, 2004.

[20] Katherine N. Lemon and Peter C. Verhoef. Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6):69–96, 2016.

[21] James B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[22] Hai-Long Nguyen, Yew-Kwong Woon, and Wee-Keong Ng. A survey on data stream clustering and classification. *Knowledge and Information Systems*, 45(3):535–569, 2015.

[23] L. O'Callaghan, N. Mishra, A. Meyerson, S. Guha, and R. Motwani. Streaming-data algorithms for high-quality clustering. In *Proceedings of the 18th International Conference on Data Engineering (ICDE)*, pages 685–694, 2002.

[24] Karl Pearson. LIII. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

[25] Doreén Pick, Jacquelyn S. Thomas, Sebastian Tillmanns, and Manfred Krafft. Customer win-back: the role of attributions and perceptions in customers' willingness to return. *Journal of the Academy of Marketing Science*, 44(2):218–240, 2016.

[26] Leon G Schiffman, Håvard Hansen, and Leslie Lazar Kanuk. *Consumer behaviour: A European outlook*. Pearson Education, 2008.

[27] Jonathan A. Silva, Elaine R. Faria, Rodrigo C. Barros, Eduardo R. Hruschka, André C. P. L. F. de Carvalho, and João Gama. Data stream clustering: A survey. *ACM Comput. Surv.*, 46(1):13:1–13:31, 7 2013.

[28] Wendell R. Smith. Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21(1):3–8, 1956.

[29] I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, 2008.

[30] S. van Buuren. *Flexible Imputation of Missing Data*. Chapman & Hall, CRC Press, Boca Raton, 2012.

[31] Ellen Marie Voorhees. *The Effectiveness and Efficiency of Agglomerative Hierarchic Clustering in Document Retrieval*. Cornell University, PhD Thesis, 1986, 1986.

[32] Michel Wedel and Wagner A. Kamakura. *Market Segmentation*. Springer US, 2 edition, 2000.

[33] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, SIGMOD '96, pages 103–114, New York, NY, USA, 1996. ACM.

Working Papers, ERCIS

Nr. 1 Becker, J.; Backhaus, K.; Grob, H. L.; Hoeren, T.; Klein, S.; Kuchen, H.; Müller-Funk, U.; Thonemann, U. W.; Vossen, G.; European Research Center for Information Systems (ERCIS). Gründungsveranstaltung Münster, 12. Oktober 2004.
Nr. 2 Teubner, R. A.: The IT21 Checkup for IT Fitness: Experiences and Empirical Evidence from 4 Years of Evaluation Practice. 2005.
Nr. 3 Teubner, R. A.; Mocker, M.: Strategic Information Planning – Insights from an Action Research Project in the Financial Services Industry. 2005.
Nr. 4 Gottfried Vossen, Stephan Hagemann: From Version 1.0 to Version 2.0: A Brief History Of the Web. 2007.
Nr. 5 Hagemann, S.; Letz, C.; Vossen, G.: Web Service Discovery – Reality Check 2.0. 2007.
Nr. 6 Teubner, R.; Mocker, M.: A Literature Overview on Strategic Information Management. 2007.
Nr. 7 Ciechanowicz, P.; Poldner, M.; Kuchen, H.: The Münster Skeleton Library Muesli – A Comprehensive Overview. 2009.
Nr. 8 Hagemann, S.; Vossen, G.: Web-Wide Application Customization: The Case of Mashups. 2010.
Nr. 9 Majchrzak, T.; Jakubiec, A.; Lablans, M.; Ükert, F.: Evaluating Mobile Ambient Assisted Living Devices and Web 2.0 Technology for a Better Social Integration. 2010.
Nr. 10 Majchrzak, T.; Kuchen, H: Muggl: The Muenster Generator of Glass-box Test Cases. 2011.
Nr. 11 Becker, J.; Beverungen, D.; Delfmann, P.; Räckers, M.: Network e-Volution. 2011.
Nr. 12 Teubner, A.; Pellengahr, A.; Mocker, M.: The IT Strategy Divide: Professional Practice and Academic Debate. 2012.
Nr. 13 Niehaves, B.; Köffer, S.; Ortbach, K.; Katschewitz, S.: Towards an IT consumerization theory: A theory and practice review. 2012
Nr. 14 Stahl, F., Schomm, F., & Vossen, G.: Marketplaces for Data: An initial Survey. 2012.
Nr. 15 Becker, J.; Matzner, M. (Eds.).: Promoting Business Process Management Excellence in Russia. 2012.
Nr. 16 Teubner, R.; Pellengahr, A.: State of and Perspectives for IS Strategy Research. 2013.
Nr. 17 Teubner, A.; Klein, S.: The Münster Information Management Framework (MIMF). 2014.
Nr. 18 Stahl, F.; Schomm, F.; Vossen, G.: The Data Marketplace Survey Revisited. 2014.
Nr. 19 Dillon, S.; Vossen, G.: SaaS Cloud Computing in Small and Medium Enterprises: A Comparison between Germany and New Zealand. 2015.
Nr. 20 Stahl, F.; Godde, A.; Hagedorn, B.; Köpcke, B.; Rehberger, M.; Vossen, G.: Implementing the WiPo Architecture. 2014.
Nr. 21 Pflanzl, N.; Bergener, K.; Stein, A.; Vossen, G.: Information Systems Freshmen Teaching: Case Experience from Day One (Pre-Version of the publication in the International Journal of Information and Operations Management Education (IJIOME)). 2014.
Nr. 22 Teubner, A.; Diederich, S.: Managerial Challenges in IT Programmes: Evidence from Multiple Case Study Research. 2015.
Nr. 23 Vomfell, L.; Stahl, F.; Schomm, F.; Vossen, G.: A Classification Framework for Data Marketplaces. 2015.
Nr. 24 Stahl, F.; Schomm, F.; Vomfell, L.; Vossen, G.: Marketplaces for Digital Data: Quo Vadis?. 2015.
Nr. 25 Caballero, R; von Hof, V.; Montenegro, M.; Kuchen, H.: A Program Transformation for Converting Java Assertions into Control-flow Statements. 2015.
Nr. 26 Foegen, K.; von Hof, V.; Kuchen, H.: Attributed Grammars for Detecting Spring Configuration Errors. 2015.
Nr. 27 Lehmann, D.; Fekete, D.; Vossen, G.: Technology Selection for Big Data and Analytical Applications. 2016.