Monitoring Service Adaptation and Customer Churn in the Beginning Phase of a New Service

Teemu Mutanen
VTT Technical Research Center of Finland
Espoo, Finland
Email: teemu.mutanen@vtt.fi

Ville Österlund
Interquest Oy
Helsinki, Finland
ville.osterlund@interquest.com

Risto Kinnunen
Interquest Oy
Helsinki, Finland
risto.kinnunen@interquest.com

***Abstract*—This work focuses on the analytics and metrics of a service adaptation. Adaptation is analysed based on activity and churn behaviour. In a non-binding registration service, churn may not be a permanent phenomenon, but partial churn prediction could be more useful as a metric. Based on the findings service, adaptation remains similar as the customer base broadens. The burst type usage data are challenging for churn predictions, but combinations and merged variables benefit the analysis. As in the future, the findings may be integrated as part of a retention campaign or applied in the development of the service.**

***Keywords*–Data mining; Service adaptation; Customer churn; Customer lifetime; Logistic regression; Cross validation**

I. Introduction

This work focuses on the analytics and metrics of a service adaptation. Data about the temporal adaptation of a service will provide knowledge and decision support for many causes. These are for example customer retention and churn, customer value prediction, and service development actions. Customer churn is not a measure of success itself, but understanding causes for it and taking actions to minimize the amount of churners may produce additional value.

In this work, we have formulated few research questions, while assuming the service in question is not under development during the time period. The hypothesis are based on the visual interpretation of the Figure 1. In Figure 1, is presented customer lifetime distribution of four groups, grouped by registration date. Each customer lifetime is measured from date of registration to the date of last visit. Number of trial customers are restricted to 100 for visual clarity. It seems that the share of active customers remains same and the trial users and churners could be visible in the data already after two months (60 days). In this study, we ask:

- Does the adaptation of the service change as customers base increases to include more users than early adopters?
- What would be a reasonable time horizon to follow customers after the registration?
- How much would additional data add value to the churn prediction accuracy?

The empirical study conducted in this work is based on the usage data of a browser based service received from a Finnish publishing company. The service in question is a publishing platform with various articles, stories and other content offered via browser. The study is based solely on back-end data about article views, browser specifics and reading times. In service adaptation, a fundamental concept is customer churn. This work address the issue by implementing predictions of possible churners and evaluate them over time.

The number of page views and click-through rates are widely used measures of success of web browser based services. There exists several ready made tools, both freemium and commercial-of-the-self type solution, for the analysis of page views and usage data statistics. These are outside the scope of this paper as the focus is on the service adaptation over time and the empirical study is based on the user specific data and not aggregated values.

The customer lifetime and service usage characteristics has been studied before. For example, how innovations spread and take hold [1] and about the nature of e-services and the e-service experience [2]. For this work, it is sufficient to recognize that in any new service implementation and roll out users may behave differently in early stages than later on as the so called early adopters may place higher value on different aspects than other users.

The customer event history has been studied from time window perspective before, *c.f.* [3]. The work by Poel *et al.* [3] focus on an already existing service and the aim is to predict churners with customer lifetime spanning for years. This work focus on new service and how the service adaptation develops over time. Jahromi *et al.* have studied the churn behaviour in business-to-business (B2B) context [4] and found out that similar methodology is suitable as in business-to-consumer (B2C) context.

This paper makes the following contributions: (1) problem description on a practical research topic, (2) empirical experiments of tools and methods for the analysis of service adaptation, (3) presents key findings how analytics benefits the selected domain, and (4) discuss possible future research aspects and implementations.

The paper is organized as follows: in Section I we discuss related research. In Section II we presents tools and methods in more detail which are used in this study. The empirical study is presented in Section III. Section III presents source data, pre-processing and results. In Section IV we state our conclusions and discuss about future research topics.

 ISBN: 978-1-61208-423-7

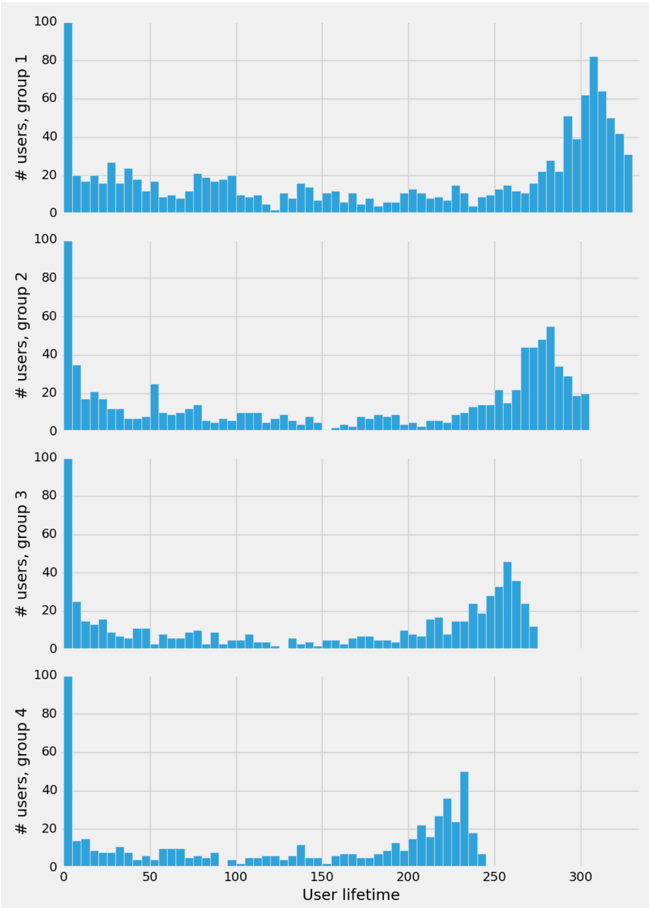

Figure 1. Customer lifetime distribution of four groups, grouped by registration date.

II. METHODOLOGY

A. *Modelling customer churn*

Customer churn has been studied and modelled before in various context. Understanding churners and possible causes for it is a way to analyse service adoption over time. Table I presents examples of customer churn studies. In the table, there are business sector and the amount of empirical data presented. As can be seen from the Table I, customer churn is highly interesting in sectors such as banking, insurance and telecommunication. Why interest is high in general on these sectors? Although few companies and services may have individual drawbacks, high churn rates are present in services and sectors where competing products are very similar and complement products are available, switching costs are low and customers need only one of them, for more see *c.f.* [5]. Current study focuses on churn from service adoption point of view in publishing sector.

In short, churners are predicted to increase overall satisfaction of a service. The process can be described in two phases, *c.f* [4] or [15]. In first phase the customer base is analysed and possible churners are identified. Then the churners are targeted with retention campaigns. Two widely used criteria for detection are either the most probable churners with the highest probability of defection or to weight future profit from lifetime value with the probability of defection.

In sectors such as publishing churn modelling and customer relationship management (CRM) are challenging in part because of rare event data. Rare event data may include individual events or numerous events, but which happen in bursts and may have long time periods in between. The case with rare event data has been studied by Ali *et al.* in [6], where rarity is addressed by sampling and observations from different time periods. Imbalanced data where churners are few in number may also be called as rare event data [6].

In publishing sector, the share of attention time is important concept. Electronic articles as well as all reading is in part entertainment and the amount of time customers spend in various channels is a zero-sum game. This implies that customers may be partial churners. Partial churners have been studied previously by Miguéis *et al.* [16] in grocery retail sector. Miguéis *et al.* [16] have analysed the importance of first impression as a measure of future customer loyalty. Instead of already established business or service, this work focus on the beginning phase of a service and how adaptation develops over time.

B. *Predicting customer churn*

In Table I, there are also presented methods applied to churn analysis or churn prediction. In literature more often than not the churn is either predicted by probability or classified in binary classes. Two of the most widely applied techniques are logistic regression and decision trees [15]. Other techniques, such as bayesian inference, partial least square, or support vector machines may produce additional value (*c.f.* [17]). In addition, dimension reduction and feature selection could also be applied if the amount of data or frequency is large (*c.f.* [18]).

In customer activity analysis, a complete churn is usually a rare event. This implies that class imbalance may create problems. Weiss *et al.* [19] names several challenges which may arise from class imbalance. These are for example, improper evaluation metrics, lack of data and noise. Class imbalance in churn prediction has been studied by Poel *et al.* [20]. Boosting techniques could be used in case of lack of data, *c.f.* [17].

C. *Definition of customer churn*

Defining customer churn in a non-binding registration service is complex and challenging. As described above, the publishing industry and its services compete with the other forms of entertainment, and with partial churners it is difficult to detect the exact time for churn. More importantly the partial churn maybe even more important from business perspective than the actual churn.

ID	m 1	m 2	m 3	m 4	m 5	m 6	Total	
#####	15	35	21	0	0	7	78	Churner
#####	10	12	3	7	15	24	71	Non -churner
#####	35	5	0	0	0	0	78	Churner
#####	76	32	41	92	102	90	435	Non -churner

Figure 2. Examples of derivation of partial churning variable and reading habits.

TABLE I. EXAMPLES OF DATA SETS APPLIED TO THE CHURN PREDICTION IN THE LITERATURE.

Authors	Year	Market sector	Source data	Methods used	Temporal coverage
Ali *et al.* [6]	2014	Banking	7204 customers	survival analysis	2008-2009
Amin *et al.* [7]	2014	Telecommunications	public data sets	rough set	n/a
Coussement *et al.* [8]	2008	Newspaper	90000 subscriptions	svm, random forests	Jan 2002 - Sep 2005
Gunther *et al.* [9]	2011	Insurance	160000 customers	logistic regression	Nov 2007 - May 2009
Jahromi *et al.* [4]	2014	B&B	11021 customers	decision tree	Sep 2011 - Sep 2012
Karahoca *et al.* [10]	2011	Telecommunications	24900 customers	fuzzy clustering	n/a
Lee *et al.* [11]	2012	Telecommunications	114000 customers	nn-classification	n/a
Mutanen *et al.* [12]	2006	Banking	151000 customers	logistic regression	2002 -2005
Xia *et al.* [13]	2008	Telecommunications	two public data sets	factor analysis, svm	n/a
Xie *et al.* [14]	2009	Banking	20000 users	Random forests	n/a

In this study churn is defined as a form of inactivity. There are numerous other ways to define churn, for example based on monetary value [16] or cancellation of an account [9]. We choose first six months of 2014 as the data set for the churn analysis. In Figure 2, examples of the derivation of partial churn are presented. We define churn based on article views. Leaving bounce visits unnoticed, the churners are customer who are inactive during the month in focus.

III. EMPIRICAL ANALYSIS

As mentioned in the introduction, our research questions were formulated based on the visual interpretation of customer lifetimes and frequency of visits. In Figure 1 lifetime distribution of four customer groups were presented. It seems that the churners could be detected already early on the duration of the customer relationship. However, before we can predict churners we will find out if the customer population changes during the time period.

A. Source data

The service in question was launched in the second half of 2013. For this study we selected customers who registered in the first half of 2014. In Figure 3 is shown the amount of customers registered on weekly basis during time period. The selection of customers from early 2014 allowed us to monitor the status of each customer for at least six months since the registration.

The data set in this study included 5956 customers in total.

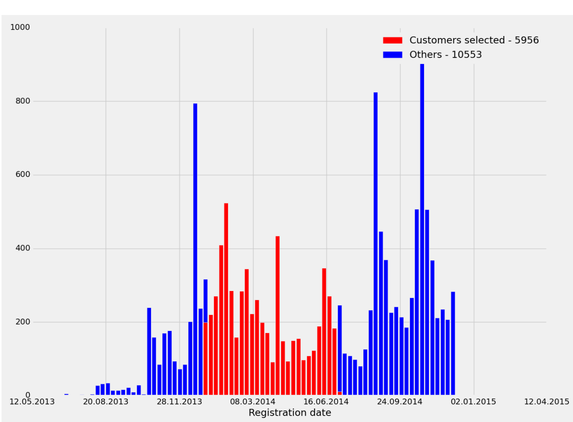

Figure 3. Selected set of customers based on the registration date.

It is worth noting that we selected our data set based on customer IDs and sorted them based on the registration date. However, we used different data sets for reading habit and activity analysis and churn prediction.

For reading habits and activity, we collected and analysed data from each customer of the first six months of their customer lifetime. Examples of this type aggregated values are presented in Figure 2. In the figure data from each customer has been aggregated on monthly basis from six months but for example different months (m1-m6) might not refer exactly to the same period in the calendar between customers.

For customer churn prediction, we selected customers who registered between January and April. Each group consisted of customers registered in a given month. From these customer groups we formulated seven set of customers for churn prediction. Illustration of the source data and churn prediction time period is presented in Figure 4. The division of sets aims to provide information about the reasonable time horizon after registration and would the additional data provide value in customer churn context.

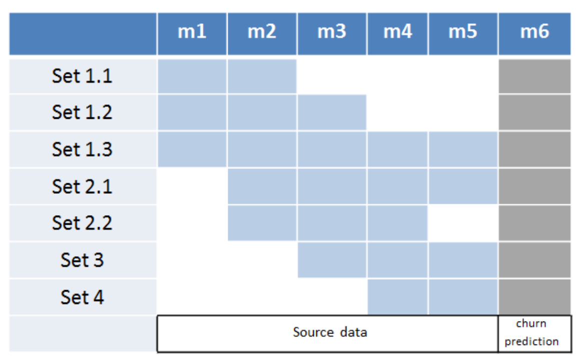

Figure 4. Illustration of the collection period of source data from customers for the prediction of customer churn.

B. Reading habits

One of our research questions was how adaptation of the service change during the beginning of phase of the service. For this we calculated information about frequency of visits for each customer. We also calculated monthly data table for each customer about the articles viewed. In the service for each article view information about the channel was saved. The channel is unique for each piece and describes source of the content. In Figure 5 examples of the data collected by channels are presented; the figure presents data from one customer. There were 14 different channels in total.

month	'c1'	'c2	'c3'	'c4'	...	'c14'
1	15	0	0	0	0	0
2	5	0	0	0	18	2
3	0	0	30	0	0	0
4	3	0	0	2	0	0
5	1	11	5	2	1	0
6	1	3	7	0	3	0

Figure 5. Examples of the derivation of reading habits variables in various content channels (c1-c14) of one customer.

Based on the reading data and frequency of visits, we classified customers to five activity segments. The five segments were formed to give information about the share of customers registered in a given month which are active after six months from registration. Out of the five segments one included trial users and one passive users. The share of each of these segments are presented in Figure 6. In Figure 6 is visible that the share of trial, passive and different types of active customers remain closely aligned. The customer base can be assumed to be similar among the customers who registered in the beginning of 2014.

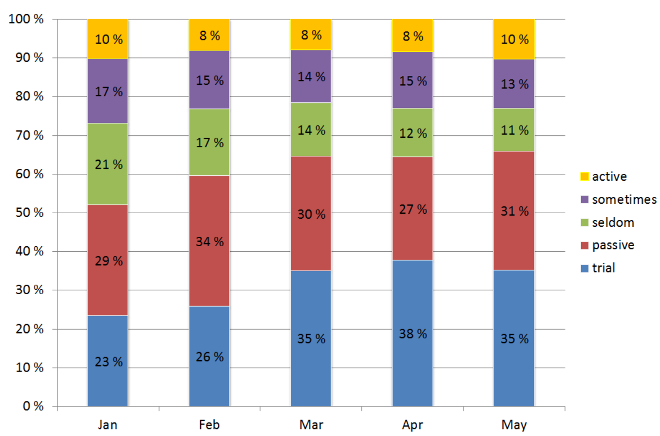

Figure 6. Customers classified to five activity segments based on their reading habits and visit frequency. The segments describe how active a set of customers are after six months from the registration.

C. Customer churn

1) Variables in training and test data: To predict customer churn we selected 24 variables for each customer. The aggregation of source data was done according to the principle presented in Figure 4. Table II presents description of the variables selected. Variables were formulated based on expert knowledge and available data. We selected only customers with lifetime over five days.

For the division of training and test set data, we selected all active customers and twice as many churners. We assigned customers randomly to either training or test set. In Table III is presented size of each data set. Reason for selecting only part of the churning customers was simply to avoid over estimation of one class.

In addition to the crisp division between training and test sets, we applied five-fold cross validation [21] to the churn prediction model fit.

TABLE II. SELECTED VARIABLES FOR CHURN PREDICTION.

	Description
1	visit score (average)
2	visit total pages (average per month)
3	number of visit (average per month)
4	number of bounce visits (average per month)
5	visit duration (average)
6	days between visits (average)
7	page view time, hour of the day before noon (average)
8	page view time, hour of the day after noon (average)
9	number of morning pages views
10	number of evening page views
11-24	number of views per channel / active months

TABLE III. SIZE OF TRAINING AND TEST SETS IN CHURN PREDICTION. NOTE, THAT AS THE CLASSES WERE NOT BALANCED, ONLY PART OF THE CHURNERS WERE SELECTED.

	Total	Training set *(non-churn/churn)*	Test set *(non-churn/churn)*
Set 1.1-1.3	1166	(143 / 300)	(53 / 94)
Set 2.1-2.2	799	(83 / 169)	(29 / 55)
Set 3	574	(60 / 151)	(34 / 37)
Set 4	483	(83 / 164)	(27 / 56)

2) Prediction: For customer churn we applied logistic regression, *c.f* [17]. Each of the data sets described in Tables II and III were trained separately. Results are presented in Table IV. In Table IV average model fit from five-fold cross validation is presented along with total values for precision, recall and area under the curve (AUC) score. As the churners were the majority population prediction could have been targeted as activity prediction as well. Thus, in Table IV total precision and recall are presented. Two examples of the resulting confusion matrix is presented in Table V. Confusion matrix present the actual values of classification results.

TABLE IV. CLASSIFICATION RESULTS FOR EACH SET. PRECISION AND RECALL ARE TOTAL VALUES FOR THE CLASSIFICATION RESULT.

	Area Under the Curve (AUC) score	Precision total	Recall total	model fit avg, 5-fold cv
Set 1.1	0.66	0.69	0.68	0.70
Set 1.2	0.67	0.70	0.70	0.72
Set 1.3	0.77	0.69	0.70	0.74
Set 2.1	0.62	0.66	0.68	0.71
Set 2.2	0.64	0.66	0.68	0.71
Set 3	0.49	0.56	0.55	0.69
Set 4	0.76	0.78	0.77	0.66

TABLE V. CONFUSION MATRIX OF THE SET 1.3 AND SET 4.

	predicted churn	set 1.3 active	predicted churn	Set 4 active
churn	82	12	54	2
active	32	22	17	10

IV. Conclusion

In this study, service adaptation and customer churn were analysed. We asked three questions in the study. The initial hypothesis were made based on visual interpretation of the data, see Figure 1. For each of these questions, we formulated suitable metrics and data based support. Classificaiton and logistic regression techniques were applied in the analysis.

Based on the findings, it seems that the user characteristics and behaviour does not change significantly. The so called early adopters do not have notable differences in usage behaviour. Thus, we can compare customer behaviour from different time periods to those of another. The adaptation overall is tilted towards short customer lifetimes. Majority of the customers have tried the service and thus we classified them as trial users.

As the service was novel at the time of data collection and the service has non-binding registration, we analysed the time period for how long it is reasonable to follow inactive customers. We are aware that all customers are important and no service provider will cancel any account on their part. However, as the trial and test users were significant population, for the service provider it is beneficial to have understanding of the lifetimes of the recently registered customers.

From the prediction results can be seen that the best results are found in either from the very recent data set or from the most extensive data set. For our research question, it is reasonable to conclude that more data produces higher accuracy for predictions.

A. Future research directions

In this work, churn was defined based on customer activity. Future research could focus on other systematic methods to formalize churning customers and active customers. The non-binding registration is challenging for the service provider as the exact moment of churn is difficult to detect. However, as registration already exists, the service development could benefit from a closer user studies in the future. Another beneficial direction for the future research could be to include and research usefulness of other variables.

It is unclear for us how these results can be generalized over a wider range of services. The churn studies are already extensive in academic literature. It would be beneficial to extent and validate these findings with other databases in other services. As the churn prediction is usually combined with customer retention campaign, it would be beneficial to conduct comprehensive studies with retention campaign and a follow-up. In [22] is presented directions and ideas how analytics could benefit the CRM overall.

From the methodology point of view, this study applied only logistic regression. Future studies may utilize other methods and scoring variations. For example, decision trees might provide value and additional information about the source data and variables. In case the amount of usage data and possibly the number of users will increase, feature selection methods could be applied before the classification and prediction results are computed.

References

[1] E. M. Rogers, Diffusion of innovations. Simon and Schuster, 2010.

[2] J. Rowley, "An analysis of the e-service literature: towards a research agenda," Internet Research, vol. 16, no. 3, 2006, pp. 339–359.

[3] M. Ballings and D. V. den Poel, "Customer event history for churn prediction: How long is long enough?" Expert Systems with Applications, vol. 39, no. 18, 2012, pp. 13 517 – 13 522.

[4] A. T. Jahromi, S. Stakhovych, and M. Ewing, "Managing {B2B} customer churn, retention and profitability," Industrial Marketing Management, vol. 43, no. 7, 2014, pp. 1258 – 1268.

[5] J. Lee, J. Lee, and L. Feick, "The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in france," Journal of services marketing, vol. 15, no. 1, 2001, pp. 35–48.

[6] Özden Gür Ali and U. Aritürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking," Expert Systems with Applications, vol. 41, no. 17, 2014, pp. 7889 – 7903.

[7] A. Amin, C. Khan, I. Ali, and S. Anwar, "Customer churn prediction in telecommunication industry: With and without counter-example," in Nature-Inspired Computation and Machine Learning, ser. Lecture Notes in Computer Science, A. Gelbukh, F. Espinoza, and S. Galicia-Haro, Eds. Springer International Publishing, 2014, vol. 8857, pp. 206–218.

[8] K. Coussement and D. V. den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert Systems with Applications, vol. 34, no. 1, 2008, pp. 313 – 327.

[9] C.-C. Günther, I. F. Tvete, K. Aas, G. I. Sandnes, and O. Borgan, "Modelling and predicting customer churn from an insurance company," Scandinavian Actuarial Journal, vol. 2014, no. 1, 2014, pp. 58–71.

[10] A. Karahoca and D. Karahoca, "{GSM} churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system," Expert Systems with Applications, vol. 38, no. 3, 2011, pp. 1814 – 1822.

[11] Y.-H. Lee, C.-P. Wei, T.-H. Cheng, and C.-T. Yang, "Nearest-neighbor-based approach to time-series classification," Decision Support Systems, vol. 53, no. 1, 2012, pp. 207 – 217.

[12] T. Mutanen, J. Ahola, and S. Nousiainen, "Customer churn prediction - a case study in retail banking," in ECML/PKDD 2006 Workshop on Practical Data Mining: Applications, Experiences and Challenges, 2006.

[13] G. en XIA and W. dong JIN, "Model of customer churn prediction on support vector machine," Systems Engineering - Theory & Practice, vol. 28, no. 1, 2008, pp. 71 – 77.

[14] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Systems with Applications, vol. 36, no. 3, Part 1, 2009, pp. 5445 – 5449.

[15] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, "Defection detection: Measuring and understanding the predictive accuracy of customer churn models," Journal of Marketing Research, vol. 43, no. 2, 2006, pp. 204–211.

[16] V. Miguéis, D. V. den Poel, A. Camanho, and J. F. e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," Expert Systems with Applications, vol. 39, no. 12, 2012, pp. 11 250 – 11 256.

[17] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[18] A. Rajaraman and J. D. Ullman, Mining of Massive Datasets. New York, NY, USA: Cambridge University Press, 2011.

[19] G. M. Weiss, "Mining with rarity: a unifying framework," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, 2004, pp. 7–19.

[20] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," Expert Systems with Applications, vol. 36, no. 3, 2009, pp. 4626–4636.

[21] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Ijcai, vol. 14, no. 2, 1995, pp. 1137–1145.

[22] A. Tuzhilin, "Customer relationship management and web mining: the next frontier," Data Mining and Knowledge Discovery, vol. 24, no. 3, 2012, pp. 584–612.