Clustering of participants in the MaxBonus loyalty system using Kohonen’s self-organizing maps

M G Dorrer\textsuperscript{1}, A V Fomin\textsuperscript{2} and D A Loginov\textsuperscript{2}

\textsuperscript{1}Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Ave., Krasnoyarsk, 660037, Russian Federation
\textsuperscript{2}Solution Factory, 10c3, Krasnoy Armii street, Krasnoyarsk, 660001, Russian Federation

E-mail: mdorrer@mail.ru

Abstract. The purpose of this article is to investigate customer clustering based on big data of the consumer loyalty system. The object of the research is the retail chains that are the clients of the Maxbonus system. Self-organizing Kohonen maps, implemented by the selforgmap function of the Deep Learning Toolbox module of the Matlab system, were used as a clustering tool. As a result of clustering 990 customers, classes invariant to the partition were identified, and it was shown that their number stabilizes with an increase in the number of classes in the Kohonen self-organizing map. The average indicators of representatives of each class have been determined. This result indicates the efficiency of the approach for clustering customers of customer loyalty systems. The prospect of the work is to include in the number of input parameters clustering the volume of purchases by product categories. Also promising is the transition to clustering retail chains participating in the Maxbonus system.

1. Introduction

This work was carried out within the framework of the project "Development of an intelligent system for analysis and management of the loyalty program - "Maxbonus" [1], carried out with the support of the Foundation for Assistance to Small Innovative Enterprises, FASIE [2] within the framework of a grant agreement for research and development work within the project.

The purpose of R&D on this project is to develop algorithms for analysis and personalized interaction with consumers using the results of processing big data (history of consumer behavior) accumulated in the loyalty program database. The developed algorithms are aimed at solving the problem of small and medium-sized businesses in terms of analyzing accumulated data to improve commercial performance.

The MaxBonus system is implemented to promptly notify the business owner about deviations in the main business processes of customer service and recommendations related to this. Also, the purpose of creating the system is to provide advisory support to the business owner regarding the main indicators of the loyalty program and customer service, informing about the main features of the loyalty program and automatically creating settings for the corresponding processes and promotions.

A fairly large number of scientific works are devoted to solving the problems of researching consumer preferences using big data. For example, in [3], the authors use data from social networks to analyze consumer preferences when making purchases. In [4], Chen analyzes the factors of consumer preferences in mobile shopping. Work [5] is also devoted to the analysis of consumer preferences when...
shopping online. In [6], Aichner and Coletti consider the process of consumer reorientation in online shopping from mass products to personalized ones, which confirms the relevance of the topic of marketing campaigns focused on narrow consumer groups.

In [7], the authors solve the problem of determining the influence of individual factors on overall customer satisfaction for several segments using linear regression and hierarchical clustering. Finally, in [8] Zhang et al. propose a new collaborative filtering algorithm based on clustering user preferences to reduce the impact of data sparseness in developing marketing recommendations. The authors of this work also attempted to analyze consumer preferences according to the data of retail chains using, for this purpose, algorithms for analyzing association rules [9].

Thus, the task of clustering consumers according to their preferences is relevant and in demand. At the same time, it was not possible to find articles describing the use of artificial neural networks without a teacher for solving this problem. So, this work is devoted to the problem of clustering the retail network consumers based on the data of the consumer loyalty system using Kohonen’s self-organizing maps.

2. Methods

2.1. MaxBonus system structure
The MaxBonus system is designed in such a way that any loyalty program with the identification of the user and the composition of the goods he buys can be used as external data. The system has an open architecture that allows you to expand the functionality of the system by adding new components.

System components:

- **External loyalty program.** It is an external DBMS that provides data for the system operation. Typically, the data contains identified purchase receipts with complete data, containing the customer ID, the contents of the receipt with the price of the goods, the partner and store ID, as well as related information contained in the loyalty program, such as the rating of a customer visit or a customer call to support.

- **Loader module.** The module is being finalized for each of the plug-in loyalty programs. The module is adapted to calculate each of the factors - for example, "average bill". The module takes the initial data for calculating the factor for the required period for the required partner of the external loyalty program, calculates the factor and stores it in the storage module in the context of the calculated time interval for each of the partners at specified recalculation time intervals.

- **Storage module.** Carries out storage of data and knowledge in the form necessary to ensure the operation of the system.

- **Automated workstation of an expert.** It is a web page available to an authorized user at the address https://expert.maxbonus.ru and designed to manage the system and train other subsystems to ensure the goals of the system. In the future, the functional of an expert should be replaced by an intelligent system created under the grant. The expert will retain the function of control and consultation.

- **Interaction module.** This subsystem, based on the settings set by the expert / analytical system, controls the dialogue of incoming and outgoing messages in real time between the user and the digital marketer chatbot.

- **User chatbot.** The main point of contact between the user and the system, through the chatbot, the partner can find out the values of the indicators he needs, by clicking on the corresponding menu buttons or a direct question, for example, "What was the revenue yesterday?" It is possible to set up a dialogue, in which the chatbot will ask clarifying questions. Also, the system itself, when it detects problems in the processes, inform the partner about it with the required details, which eliminates the need to regularly review a series of reports and allows you to instantly respond to failures.
Automatic fixing module. This subsystem is the core of the system's operation, based on the "knowledge" from an expert, the system independently, relying on data sets, makes decision chains. Performs checks, recalculations and, if necessary, initiates a dialogue with a partner via the chatbot.

2.2. The problem of clustering and segmentation of consumers

The designed and implemented system uses a loyalty program as external data sources. The information it contains is the digitized buying behavior of customers with a particular partner. The entire set of knowledge that can be obtained from this data is the result of analyzing customer buying behavior. As a result of the analysis of the partners' needs, the project team selected and implemented the following set of factors, for the formation of which groups of algorithms were developed:

General algorithms are used to calculate factors:

- Average check.
- Number of checks.
- Amount of checks.
- Average purchase cycle.
- Percentage of Valued Checks.
- Average purchase rating.

Segmentation algorithms are used to calculate factors:

- Share of new customers;
- Share of customers with free communication channels;
- Share of loyal customers;
- Percentage of lost customers.

Segmentation algorithms and factors that are created in the system based on the results of their work are of the greatest value for the partner of the loyalty program, because obtaining such information in an alternative way requires a lot of analytical work from the partner. It is expected that after the full implementation of the planned functions of the system, it will allow the user through the chatbot to receive recommendations of approximately the following content:

“I noticed that in May you got 147 customers who stopped buying the “meat” category in your store, they continue to visit you, but they spend 24% less. I recommend launching a personal promotion for these customers for this category of products to return them to their original behavior.”

2.3. Application of Kohonen’s self-organizing maps for consumer clustering

The results of loading the data of the loyalty system by the loader module allow for automatic clustering / segmentation of various subjects of consumer behavior analysis. To solve the clustering problem, unsupervised neural network clustering is used using Kohonen’s self-organizing maps [10].

Clustering algorithms were prototyped in the Deep Learning Toolbox [11] neural network module of the Matlab system with subsequent transfer to the core of the system. The selforgmap function of the module was used, which implements the Kohonen self-organizing map model with data migration from the MaxBonus system through Excel files. Figure 1 shows the architecture of the Kohonen map used to cluster the buyers of one of the program partners into 6 classes by 11 criteria.
3. Results
Figure 2 shows the distribution of 990 loyalty program members across 6 classes. It can be seen that the chain has grouped most of the buyers into one large class. Experiments were carried out for the number of classes from 2 to 8, a similar pattern of the partition was repeated constantly, the network singled out two classes that were repeated for any partition, the remaining classes varied. When dividing into more than 4 classes, the network began to allocate "empty" classes, as you can see in figure 2.

In the course of the study, the stability of classes was assessed during clustering, the division of buyers into classes in the process of increasing the number of classes.

The same sample (buyers of a chain of butcher shops) was subjected to automatic clustering by the Kohonen network for the number of classes from 2 to 6. Then, the joint assessment of the same buyer in $N$ classes and in $N+1$ classes for $N = 1..5$ was made.

Tables 1-4 show the redistribution of buyers between classes with an increase in the number of classes. Classes of a smaller subset (by $N$ classes) are listed by rows, by columns - a larger subset (by $N + 1$ classes). At the intersection, the number of buyers in both classes.
Tables 5 to 9 show the average characteristics of the classes. Averaging was carried out over the entire period of operation of the trading network under investigation; in the future, the algorithm will be set to averaging over a typical time range - month, quarter, year. The following database fields were used for averaging:

- Gender - the gender of the buyer.
- Age - the age of the buyer.
- Sum_amount - the total amount of purchases for the period.
- Check_count - the number of checks for the period.
- Avg_amount - average bill for the period.

Table 5. Average characteristics of classes when divided into 2 classes.

| Number of buyers | Distribution in 2 classes | Distribution in 3 classes | Grand total |
|------------------|----------------------------|---------------------------|------------|
|                  | 1  | 2           | 3    |                | 39          |
|                  | 2  | 117         | 834  | 951           |
| Grand total      | 19 | 137         | 834  | 990           |

Table 2. Intersection between classes (clustering for 3 and 4 classes).

| Number of buyers | Distribution in 3 classes | Distribution in 4 classes | Grand total |
|------------------|----------------------------|---------------------------|------------|
|                  | 1  | 2           | 3    | 4    | 19 |
|                  | 2  | 20          | 117  | 137  |
|                  | 3  | 647         | 187  | 834  |
| Grand total      | 19 | 667         | 117  | 187  | 990 |

Table 3. Intersection between classes (clustering into 4 and 5 classes).

| Number of buyers | Distribution in 4 classes | Distribution in 5 classes | Grand total |
|------------------|----------------------------|---------------------------|------------|
|                  | 1  | 2           | 3    | 4    | 5  | 19 |
|                  | 2  | 66          | 601  | 667  |
|                  | 3  | 20          | 97   | 117  |
|                  | 4  | 187         | 187  |      |
| Grand total      | 5  | 34          | 163  | 601  | 990 |

Table 4. Intersection between classes (clustering into 5 and 6 classes).

| Number of buyers | Distribution in 5 classes | Distribution in 6 classes | Grand total |
|------------------|----------------------------|---------------------------|------------|
|                  | 1  | 2           | 4    | 5    | 6  | 5 |
|                  | 2  | 34          | 34   |      |    |
|                  | 3  | 151         | 12   | 163  |
|                  | 4  | 601         | 601  |      |
|                  | 5  | 187         | 187  |      |
| Grand total      | 5  | 151         | 34   | 613  | 187 | 990 |

| Distribution in 2 classes | 1 | 2 | 3 | 4 | Grand total |
|----------------------------|---|---|---|---|------------|
|                            | 19|   |   |   | 19         |
|                            | 20| 117| 834| 951|             |
|                            | 19| 137| 834| 990|             |

| Distribution in 3 classes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Grand total |
|---------------------------|---|---|---|---|---|---|---|---|---|------------|
|                            | 19| 667| 117| 187| 990|    |    |    |    |           |

| Distribution in 4 classes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Grand total |
|---------------------------|---|---|---|---|---|---|---|---|---|------------|
|                            | 5 | 34| 163| 601| 990|    |    |    |    |           |

| Distribution in 5 classes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Grand total |
|---------------------------|---|---|---|---|---|---|---|---|---|------------|
|                            | 151| 12| 163| 601| 990|    |    |    |    |           |

| Distribution in 6 classes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Grand total |
|---------------------------|---|---|---|---|---|---|---|---|---|------------|
|                            | 187| 187| 187| 187| 990|    |    |    |    |           |
### Table 6. Average characteristics of classes when divided into 3 classes.

| Class | Number of customers in the class | Average for the gender field | Average for the age field | Average for the sum_amount field | Average for the check_count field | Average for the avg_amount field |
|-------|---------------------------------|------------------------------|---------------------------|----------------------------------|-----------------------------------|-------------------------------|
| 1     | 39                              | 0.38                         | 39.18                     | 52866.87                         | 156.49                            | 415.49                        |
| 2     | 951                             | 0.32                         | 43.55                     | 4421.17                          | 12.45                             | 431.52                        |
| Grand total | 990                             | 0.32                         | 43.38                     | 6329.64                          | 18.12                             | 430.89                        |

### Table 7. Average characteristics of classes when divided into 4 classes.

| Class | Number of customers in the class | Average for the gender field | Average for the age field | Average for the sum_amount field | Average for the check_count field | Average for the avg_amount field |
|-------|---------------------------------|------------------------------|---------------------------|----------------------------------|-----------------------------------|-------------------------------|
| 1     | 19                              | 0.37                         | 38.74                     | 70987.26                         | 217.11                            | 411.98                        |
| 2     | 137                             | 0.36                         | 43.33                     | 19111.29                         | 50.26                             | 502.09                        |
| 3     | 834                             | 0.31                         | 43.49                     | 2757.00                          | 8.31                              | 419.62                        |
| Grand total | 990                             | 0.32                         | 43.38                     | 6329.64                          | 18.12                             | 430.89                        |

### Table 8. Average characteristics of classes when divided into 5 classes.

| Class | Number of customers in the class | Average for the gender field | Average for the age field | Average for the sum_amount field | Average for the check_count field | Average for the avg_amount field |
|-------|---------------------------------|------------------------------|---------------------------|----------------------------------|-----------------------------------|-------------------------------|
| 1     | 5                               | 0.60                         | 50.80                     | 109920.11                        | 416.60                            | 285.72                        |
| 2     | 34                              | 0.35                         | 37.47                     | 44476.69                         | 118.24                            | 434.57                        |
| 3     | 163                             | 0.32                         | 44.59                     | 14361.78                         | 37.80                             | 496.73                        |
| 4     | 601                             | 0.31                         | 43.76                     | 2938.12                          | 9.12                              | 390.33                        |
| 5     | 187                             | 0.35                         | 41.97                     | 522.75                           | 1.06                              | 507.06                        |
| Grand total | 990                             | 0.32                         | 43.38                     | 6329.64                          | 18.12                             | 430.89                        |
### Table 9. Average characteristics of classes when distributed into 6 classes.

| Class | Number of customers in the class | Average for the gender field | Average for the age field | Average for the sum_amount field | Average for the check_count field | Average for the avg_amount field |
|-------|----------------------------------|------------------------------|----------------------------|----------------------------------|-----------------------------------|----------------------------------|
| 1     | 5                                | 0.60                         | 50.80                      | 109920.11                        | 416.60                            | 285.72                           |
| 2     | 151                              | 0.34                         | 44.59                      | 14803.08                         | 38.63                            | 508.45                           |
| 4     | 34                               | 0.35                         | 37.47                      | 44476.69                         | 118.24                           | 434.57                           |
| 5     | 613                              | 0.30                         | 43.77                      | 3053.04                          | 9.48                             | 389.53                           |
| 6     | 187                              | 0.35                         | 41.97                      | 522.75                           | 1.06                             | 507.06                           |
| Total | 990                              | 0.32                         | 43.38                      | 6329.64                          | 18.12                            | 430.89                           |

As you can see in Table 9 and Figure 2, when clustering into 6 classes, one of the classes (class 3 in Table 9) is empty.

### 4. Discussion

Even without analyzing the distribution of customer parameters by class, we can conclude that the sample contains a significant stable core of copies (600 - 640 customers). This nuclear class (with 4 classes - the second, with 5 classes - the fourth, with 6 classes - the fifth) does not begin to split with an increase in the number of classes, that is, its customers differ in significant similarity (linear proximity) in the feature space.

Analysis of customer parameters by class allows us to draw the following conclusions:

- The "kernel" class includes buyers with about 10 checks and an average check of 390-400 rubles.
- Starting from grade 5, for the first time, a class has been steadily distinguished, including more men than women, and this category is characterized by a very large number of checks - more than 400.
- Customers with 1 check are clearly distinguished into a separate class - these are 187 customers; the class is distinguished starting from the division into 4 classes and remains unchanged during subsequent divisions.

### 5. Conclusion

Assessing the prospects of this work, it seems interesting to supplement the analysis by considering the parameters of purchases of buyers by product group. The use of such data will allow us to analyze how clustering will change, taking into account purchases by product groups and get an idea of the product preferences of buyers of previously identified classes.

Also, it looks promising to conduct clustering not only customers, but also partners. Clustering partners will allow solving the following tasks:

- Better assess the grouping of partners by class, including estimating the number of classes that provide an adequate breakdown given the available data.
- Assign partners newly connected to the loyalty program to one of the previously specified classes, which will ensure the formation of recommendations for them based on their accumulated experience of the system.

### References

[1] Maxbonus  http://maxbonus.ru/
[2] Foundation for Assistance to Small Innovative Enterprises (FASIE)  http://fasie.ru/
[3] Vithayathil J, Dadgar M and Osiri J K 2020 Social media use and consumer shopping preferences Int. J. Inf. Manage 54 102117 doi: 10.1016/j.ijinfomgt.2020.102117

[4] Chen H-J 2018 What drives consumers’ mobile shopping? 4Ps or shopping preferences? Asia Pacific J. Mark. Logist. 30(4) 797-815 doi: 10.1108/APJML-08-2017-0167

[5] Rajamma R K and Neeley C R 2005 Antecedents to Shopping Online: A Shopping Preference Perspective J. Internet Commer. 4(1) 63-78 doi: 10.1300/J179v04n01_03

[6] Aichner T and Coletti P 2013 Customers’ online shopping preferences in mass customization J. Direct, Data Digit. Mark. Pract. 15(1) 20-35 doi: 10.1057/dddmp.2013.34

[7] Schüller D and Pekárek J 2018 Customer Satisfaction Measurement – Clustering Approach Acta Univ. Agric. Silvic. Mendelianae Brun. 66(2) 561-9 doi: 10.11118/actaun201866020561

[8] Zhang J, Lin Y, Lin M and Liu J 2016 An effective collaborative filtering algorithm based on user preference clustering Appl. Intell. 45(2) 230-40 doi: 10.1007/s10489-015-0756-9

[9] Ogurtsov D A and Dorrer M G 2019 Application of association rules learning for studying the store history of a large retail chain in Journal of Physics: Conference Series 1399(3) doi: 10.1088/1742-6596/1399/3/033114

[10] Kohonen T 1990 The self-organizing map Proc. IEEE 78(9) 1464-80 doi: 10.1109/5.58325

[11] Deep Learning Toolbox https://www.mathworks.com/help/deeplearning/