6th International Scientific Conference ITEMA 2022 – Proceedings
DOI: https://doi.org/10.31410/ITEMA.2022.17

Dynamic Generation of Website Content Based on User Segmentation Using Artificial Intelligence

Nenad Kojić[1]
Mladen Petrović[2]
Natalija Vugdelija[3]

Received: November 16, 2022
Accepted: January 11, 2023
Published: June 12, 2023

Keywords:
Dynamic website;
User segmentation;
Artificial intelligence;
Web 4.0;
Kohonen neural network

Abstract: *The aim of this work is the segmentation of website users on the basis of artificial intelligence with the aim of dynamically modifying the content of the website for users, in accordance with the objectives of Web 4.0, and in this way enabling quick and optimal display of content following their needs. User classification will be based on click events on categories/subcategories and articles. Based on that information, using Kononen's neural network, the user will be classified into one of the n categories to which the neural network was initially trained. Based on the detected type of the user's classification, the content of the site is dynamically changed to the user, and the categories and products for which the majority of users of that type of classification have expressed greater interest are initially displayed and offered. The goal is to adapt the content of the site to the needs of the user and in this way the user can easily and quickly find the desired product.*

1. INTRODUCTION

Modern Internet users are increasingly basing their needs on searching for information and products as well as purchasing products through websites and web applications. In the process of searching for desired products or services, we often find and buy those that we were not initially looking for. The goal of every site that offers services or products to the user is to attract the user, keep them on the site, place as many different products as possible and finally sell its product to the user (Gutama, 2021). To that end, websites often have large amounts of pages, with as many products as possible, often the same product presented in multiple images, in multiple colors, from multiple angles, multiple variations, and all with the aim of finding the best solution for the user's needs. Such large amounts of information, data, pages, texts, and images often lead to the complete opposite effect, which is burdening users with a large number of unnecessary details, which can make users feel uncomfortable and leave the site. This effect is even more emphasized with sites that have a large number of interactive elements (sliders with a large number of images, large-sized images with a large number of objects in them, pop-ups that appear frequently and in large numbers, advertisements, imposing the opinions of other customers, etc.). Although both developers and website owners have created their applications to offer the user a product or service, and in addition for the user to come, find and buy that product, the great competition on the market and the desire for dominance can often drive users away from certain websites due to excessive information burden on users (Rosário, 2021).

As both sellers and buyers want to be in a win-win situation, websites intended for advertising and sales have been constantly changing and adapting to the specific needs of users. Thus, a

[1] Academy of Technical and Art Applied Studies Belgrade, Department School of Applied Studies for Information and Communication Technologies in Belgrade, 16 Zdravka Celara St., Belgrade, Serbia
[2] Quadro Consulting, 6 Vladimira Popovića St., Belgrade, Serbia
[3] Academy of Technical and Art Applied Studies Belgrade, Department School of Applied Studies for Information and Communication Technologies in Belgrade, 16 Zdravka Celara St., Belgrade, Serbia

significant number of websites began to remember the user's previous behaviour and to offer the user goods from the same or similar product categories in advance, with the desire to increase the user experience. The second group of sites started by including different algorithms for recognizing the needs of users and defining a certain group of products as a suggestion to the user. In many different ways, sellers try to segment the items they offer to the user with the aim that the user finds the desired products as soon as possible and buys them (Punj, 2012).

Research has shown that user behaviour, user needs, expectations and wishes can be classified in certain ways. Psychology, as a science, has developed a large number of specific disciplines that analyse user behaviour in different ways and the various effects that lead to the final opinion, attitude and impression of the final product. So today we have a large number of works and researchers dealing with mass psychology, psychology of design, psychology of colour, psychology of optimal experience (Sharpe, 1974; Csikszentmihalyi, 2014). All this research shows that user experiences, behaviours and characteristics can be grouped and classified. In addition, the correct choice of colours, designs, shapes, etc. increases user experience, which contributes to better placement and sale of goods and services.

Psychology also indicates that users can be classified into certain logical groups that gather users with similar interests, desires and needs. This classification of users enables web applications to perform user segmentation. The goal of segmentation is to recognize the type of user and offer them products that correspond to their expected needs (Tu, 2010). In this way, the users will not be overwhelmed by a large amount of information and can quickly find what they need.

The era of Web 4.0 indicates that the goal of future web applications is based on the implementation of artificial intelligence and finding information that suits each user and their specific needs (Cioffi, 2020). Starting from that, in this paper, a solution for the segmentation of site users is proposed, based on clicks on individual target parts, ads/products, and based on artificial intelligence. For the purposes of artificial intelligence, self-organizing artificial neural networks were used, namely its representative Kohonen's neural network (KNN) (Kohonen, 2012). KNN aims to dynamically classify users (Müller, 2022), to provide the website code with information that is further used to dynamically modify the content of the website and display products to the user following their expectations. Such a dynamic system has a constant phase of learning and training, and with each new input it further improves the way of classifying the user experience.

This paper is organized into five chapters. After the Introduction, the basic principles of Kohonen's neural network are given. In the third chapter, there are suggestions for the use of KNN for the purposes of user classification and the proposed logic for changing the content of the site. In the fourth chapter, a conclusion is given with guidelines for further research and in the end, a list of used references is given.

2. KOHONEN NEURAL NETWORK

There are a large number of data processing methods that are based on the clustering of the set Q, which involve dividing and grouping the data into clusters Q_i, i = 1, . . . ,N so that data distances within the same cluster are minimal, while data distances between different clusters are maximal. This type of classification requires that the observed data can belong to only one cluster and must be classified into a certain cluster. Historically, artificial neural networks have shown good results in terms of input data classification and clustering (Roohi, 2013).

Starting from three primary learning paradigms, artificial neural networks can be divided into three main categories: supervised, feedback-reinforced and independent-unsupervised. Self-organizing maps (SOM), also known as Kohonen networks after their inventor Teuvo Kohonen, belong to the class of unsupervised networks (Kohonen, 2012). They determine the representation of internal weights for presenting input data without any user supervision. This is an extremely important feature in the case when there is a large dynamic of changes and fluctuations in the input data. Self-organizing maps are a visualization technique for data representation that reduces data dimensions through the use of self-organizing neural networks (Pal, 1993). Self-organizing feature maps (SOFM) learn to classify input vectors according to how they are grouped in the input space. The input space can be any set of data with the most diverse types of data contained in it (Kohonen, 2012).

A self-organizing map consists of elements called nodes or neurons, which can be visualized as a neural network in the form of a matrix. The usual arrangement of nodes is a regular arrangement in the form of a hexagonal or rectangular grid. Each node is associated with a weight vector of the same dimension as the input vectors representing the data, and is also associated with a position within the map structure (Kohonen, 2012). The procedure of placing a vector from the data space on the map starts by finding the node with the closest weight vector to the vector taken from the data space. Then the coordinates in the map of this node are set to the value of the vector. The distance between neurons is calculated based on their positions using the classic mathematical distance function. When input samples are presented to the network, a search is performed to select a winning neuron *c*. The input vector *x* at time *t* is compared with each of the weight vectors mi belonging to the SOM, and the minimum of the Euclidean distance between the input signal and the neuron's weight coefficients determines the best match, i.e. belonging to the same cluster.

$$\| x(t) - m_c(t) \| = \{ \|x(t) - m_i(t)\| \ \| x(t) - mc(t) \| \} \tag{1}$$

In this case, the weight coefficients of the nodes are updated taking into account the environment in the form of a circle around the winning unit with a gain function α(t) (values of this function are $0<\alpha<1$) as follows:

$$m_i(t+1) = m_i(t) + \alpha(t)[x_i(t) - m_i(t)] \tag{2}$$

Outside this area, the weight coefficients remain unchanged:

$$m_i(t+1) = m_i(t) \tag{3}$$

The gain function is linear, with the highest values in the winning neuron and the lowest at the region boundaries. The arrangement of neurons in the SOM builds a discrete approximation of the distribution of samples used to train the neural network (Kohonen, 2012).

More neurons indicate regions with a high concentration of samples and fewer in regions where the samples are diluted. Figure 1 shows sample data as well as nodes of the SOM in the final state in the coordinate system. The final state of the map depends on three main conditions: the initial values of the weight vectors in the network, the data used for training and the parameters (characteristics) of the map. Map configuration parameters such as the number of network nodes, gain function, degree of learning and node distances are the main elements that determine the final result. A self-organizing map represents a topological organizer in the sense of ordering data, but not a clustering procedure itself (Pal, 1993; Kohonen, 2012). The final weight vectors in the map are used for clustering, which allows the formation of clusters with specific network links.

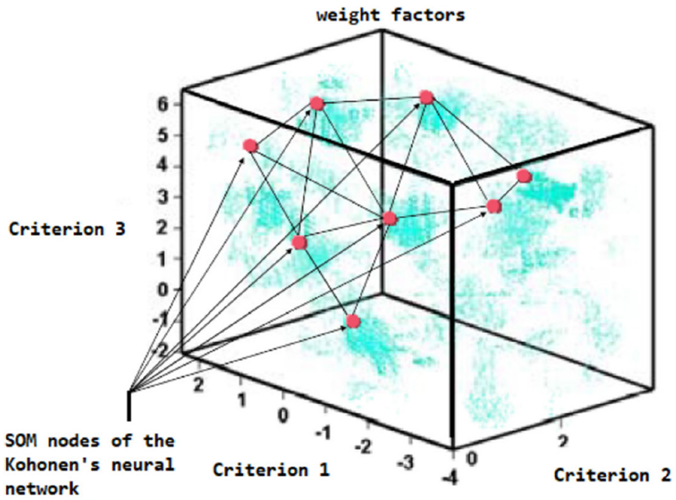

Figure 1. Distribution of nodes within the SOM based on input data.

3. PROPOSED ALGORITHM

The goal of every website user is to find the appropriate content as soon as possible, and the goal of every website owner dealing with sales is for the user on their website to be satisfied and make a purchase. When the concept of creating websites is based on marketing and advertising products that are primarily based on the content that is marketed to the user, we call it Contextual advertising (Zhang, 2012). The ideal mutual benefit scenario is that the user quickly finds the desired content/product and buys it. The psychology of customer behaviour defines the targeting of users in accordance with their experience, behaviour and habits as behavioural targeting (Ozcelik, 2019). The key problem is how to automatically classify users and products on the website and how to match them with each other. One of the methods presented in the professional literature is the Click-through rate (Zhou, 2019).

In this way, it is possible to monitor the user's movement within the website based on the monitoring of click events and thus try to understand the user's expectations in relation to their needs. On the other hand, this approach may require a fairly large sample of users in order to recognize their common characteristics. Another approach can be seen through the analysis of the strings that the user enters in the search field and in this way more clearly understand what exactly a specific group of users is looking for, in what order, priority, etc. This concept shows good results through the structure of the Vector space model (VSM) (Singh, 2015).

However, relying only on entered search terms can greatly reduce the quality of user classification because some users do not use text search, so their activity on the website cannot be detected. Some form of combination of these approaches should have the best results. During the last decade, one of the statistical models that shows exceptional results in modelling a large amount of data that can be more or less connected, i.e. correlated, is Latent Dirichlet Allocation (LDA) (Jelodar, 2019). In this way, grouping can be realized by analysing a large amount of information, by looking for previously unnoticed details that can create logical groups with a clear explanation of why some parts of the initial data are similar (Jelodar, 2019).

Starting from these scientific achievements, this paper proposes a solution for user classification based on artificial neural network (ANN) and in accordance with the principles of behavioural targeting, simultaneously combining the principles of click-through rate, VSM and the principle of LDA. The methodology used in this paper is based on the use of one of the web platforms for advertising and selling drinks in the Republic of Serbia. The activities of the user x_i, are stored in terms of each of their clicks, array **C**, and the entered search terms, array **S**. Also, for each user, certain pieces of information are stored: the time spent on a page - array **T**, the set of pages visited - array **P**, the items purchased - array **A**. On the basis of the results collected over several months, the detection of users who accessed the site multiple times was realized using client cookies in correlation with the identifiers of logged-in users who went through the authorization and authentication process. In this way, a clear distinction was made between users who visit the site while they are not authenticated among themselves and connecting an authenticated user with them while they were unauthenticated.

Considering that KNN can work with more criteria, for the purposes of this work, a total of 3 criteria were used: the total number of clicks on individual products, the terms that were searched and the items that were finally purchased. All three criteria are viewed as multidimensional parameters because they are based on the number of users, the time the user spent implementing the observed action and taking into account the distinction/pairing of the same users who are authorized or unauthorized.

Samples defined in this way are considered as inputs to KNN. The input vector x is compared at each moment t with each of the weight vectors m_i belonging to the SOM. The minimum value of the Euclidean distance determines the best match, i.e. belonging to a certain cluster. The number of clusters is not necessarily limited and does not have to be formally numbered and described. It presents the conclusion of KNN based on the entered results and groups users into N clusters in accordance with recognized common characteristics of user behaviour.

When a new user comes to the website, and KNN has already created its own clusters, the new user's activities will either contribute to assigning them relatively quickly to the existing cluster (if their behaviour is in accordance with the previously recognized behaviour of the user), or will contribute to the new organization of KNN (which may result in a change in the weight vectors m_i, or a change in the number of clusters and the logic used to group the nodes of the network into new clusters in the n-dimensional space).

In the first phase of work, the training of the network was done after the user left the site until the minimum amount of data needed for the training of KNN was formed. In the testing phase, corrections of input parameters and weight coefficients are dynamically applied in each of the three observed user activities. In this way, the first and key phase was realized, which represents the classification of users in accordance with the principles of behavioural targeting in correlation with the observed multicriteria principles of LDA. Based on the dynamic user classifications created in this way, the second phase of implementation of the proposed algorithm is entered. The second phase aims to dynamically create the content of the website in accordance with the conclusion of the first phase of user classification.

For these needs, the website was realized as a dynamic website, using the server programming language PHP and the MySQL database, as a server-side rendering method for displaying data to the end user. The initial view is defined by predefined content that is displayed to each user.

This content is stored in the database and displayed to the user for whom no previous activity is found in the client cookie. Based on that content, the user can either leave the site, which is not recorded, or click on one of the links/products, i.e. specific items, or enter search content. Regardless of which of the mentioned two activities is implemented, that activity changes the input vector of the KNN and the user tries to classify it in relation to the previous clusters of the neural network. With each subsequent activity, this process is repeated and the user can only be more precisely assigned to a cluster that is more suitable for the user experience. In relation to the calculated type of cluster to which the user belongs, the content that is dynamically retrieved and displayed from the database in the form of a group of articles and the articles themselves. They are displayed to the user on the page and in the block for recommended articles. This mechanism is presented in Figure 2.

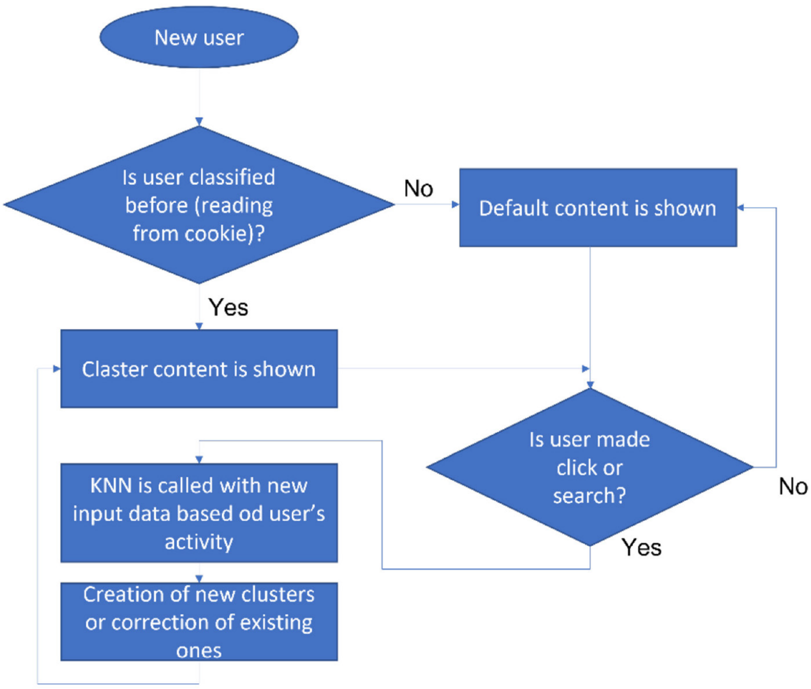

Figure 2. Proposed algorithm for user's activity in relation to the clicks

In the proposed way, each activity of the user can either confirm the previously juxtaposed group of clusters and their position increases, or the change appears if the weight coefficients and distances of the network nodes led to the change of the cluster by their recent changes. In each case, the dynamics of changing activities of each user are taken into account and entered equally into the total knowledge available to the neural network. Based on the change of cluster membership, the dynamic content that is further displayed to that user is directly changed.

The described activities were analysed over one year. Over 20,000 entries participated in the final form. In order to compare the results, a lot of different analyses were carried out, of which we singled out four special phases for the purposes of this work. Each phase lasted 3 months. In the first phase, the same content is used without changes based on user experience (Phase 1). This phase was taken as the reference phase. In the second phase, only the user's activity in relation to the click was observed (Phase 2). In the third phase, only the use of search keywords was observed (Phase 3). In the final phase, the simultaneous effects of user clicks and the use of search terms were combined, as described in this paper (Phase 4). The effect was measured in relation to the time spent on the viewed pages (*TC*), the total number of clicks that resulted in a purchase (*CC*) and the value of the purchase (*VC*). These values are scaled to relative total values of all phases in the interval 1-100%.

The KNN was implemented in the form of an external API that referred to the described user activities and which activated the stored procedures in the primary database, which recorded all the user activities from before that were used as input vectors for the KNN.

The obtained results show that the measurement of the TC parameter by phases was 75%, 49%, 58% and 41%, Figure 3. In correlation with the measurement of the *CC* parameter, which has values of 81%, 74%, 69% and 52%, a difference is observed in Phase 2 and Phase 3. Such results indicate that the observed group of users does not have the same behavioural characteristics when compared in relation to the number of clicks and retention time, Figure 3.

However, the analysis of the *VC* parameter gives a more uniform situation, with values of 75%, 77%, 78% and 82%, and with a more clearly profiled function that indicates that the phase of integration of several different parameters, was the most effective from the point of view of users and website owners. These results are also shown graphically on the diagrams, Figure 3.

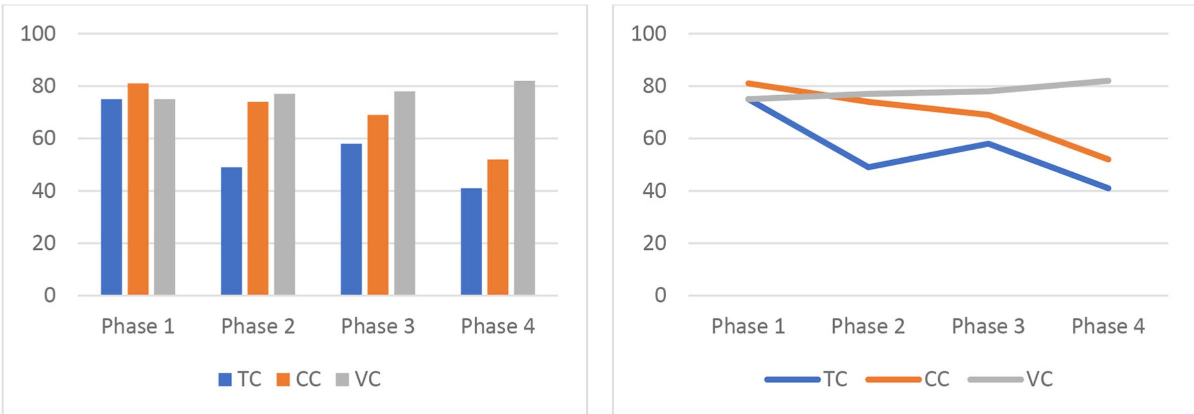

Figure 3. Values of parameters TC, CC and VC for Phase 1 to Phase 4 (two different types of charts).

The aforementioned analysis indicated the need for multicriteria analysis and contributed to the LDA concept. In order for the results to be even more credible, it is necessary to increase the sample number of users and include various factors that influence the decisions and attitudes of consumers. The mentioned testing was temporarily interrupted after the testing period because it was shown that the time required to update the KNN weight statuses became longer and longer with a larger number of data, so a different way of implementing the mentioned code must be found if it is maintained as a condition that every action of the user is carried out contacting KNN. This is certainly the subject of further work on improving the proposed algorithm.

4. CONCLUSION

This paper presents a solution for the dynamic generation of website content intended for advertising and product sales. Dynamic content generation is based on the dynamic creation of clusters of users in relation to their user behaviour on the site and their expectations. Clustering was realized using artificial neural networks. The results show that the application of the proposed algorithm increased the efficiency of the initial website solution and proved to be very useful. Further research will be directed towards changing the way clustering is realized due to the speed of the system's response and changing the number of input parameters, the size of the dimensions of the matrix of the input set, and the optimization of the speed of calculating the weight coefficients of the proposed neural network.

References

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & De Felice, F. (2020). Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, 12(2), 492.

Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. In Flow and the foundations of positive psychology (pp. 209-226). Springer, Dordrecht.

Gutama, D. H., Umami, I., & Saputro, P. H. (2021). Analysis of the effect of website sales quality on purchasing decisions on e-commerce websites. *International Journal of Informatics and Information Systems*, 4(1), 71-81.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.

Kauffmann, J., Esders, M., Ruff, L., Montavon, G., Samek, W., & Müller, K. R. (2022). From clustering to cluster explanations via neural networks. IEEE Transactions on Neural Networks and Learning Systems.

Kohonen, T. (2012). Self-organization and associative memory (Vol. 8). Springer Science.

Kohonen, T. (2012). Self-organizing maps (Vol. 30). Springer Science & Business Media.

Ozcelik, A. B., & Varnali, K. (2019). Effectiveness of online behavioral targeting: A psychological perspective. Electronic Commerce Research and Applications, 33, 100819.

Pal, N. R., Bezdek, J. C., & Tsao, E. K. (1993). Generalized clustering networks and Kohonen's self-organizing scheme. *IEEE Transactions on Neural Networks*, 4(4), 549-557.

Punj, G. (2012). Consumer decision making on the web: A theoretical analysis and research guidelines. *Psychology & Marketing*, 29(10), 791-803.

Roohi, F. (2013). Artificial neural network approach to clustering. Int. J. Eng. *Sci.*(IJES), 2(3), 33-38.

Rosário, A., & Raimundo, R. (2021). Consumer Marketing Strategy and E-Commerce in the Last Decade: A Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(7), 3003-3024.

Sanborn, F., & Harris, R. J. (2019). A cognitive psychology of mass communication. Routledge.

Sharpe, D. T. (1974). The psychology of color and design. Nelson-Hall.

Singh, V. K., & Singh, V. K. (2015). Vector space model: an information retrieval system. *Int. J. Adv. Engg. Res. Studies*/IV/II/Jan.-March, 141(143).

Tu, S., & Lu, C. (2010, November). Topic-based user segmentation for online advertising with latent dirichlet allocation. In International Conference on Advanced Data Mining and Applications (pp. 259-269). Springer, Berlin, Heidelberg.

Zhang, K., & Katona, Z. (2012). Contextual advertising. *Marketing Science*, 31(6), 980-994.

Zhou, G., Mou, N., Fan, Y., Pi, Q., Bian, W., Zhou, C., ... & Gai, K. (2019, July). Deep interest evolution network for click-through rate prediction. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 5941-5948).