Personalization Approaches for Cultural Heritage Study

Emanuela Mitreva[1], Alexandra Nikolova[1[0000-0002-2225-5077]], Vladimir Georgiev[2[0000-0002-1194-1098]], Ani Gigova[3]

[1] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] Computer Science Department, American University in Bulgaria, Blagoevgrad, Bulgaria
[3] Private Primary School "Pitagor", Sofia, Bulgaria
emitreva@gmail.com, alxnikolova@gmail.com, vgeorgiev@aubg.edu, annett909@gmail.com

Abstract. In this paper, we review different approaches of providing personalization for cultural heritage content – several static and adaptive approaches, and recommendation systems. The purpose of our research is to compare these methods, evaluate their feasibility for certain types of applications in the field. We outline the needs for personalized experience involving cultural heritage content and describe how those needs are addressed by the different approaches, and point the challenges, benefits, and disadvantages of each of these methods.

Keywords: Personalization, Adaptive Methods, Web Usage Mining, Content-based Filtering, Cultural Heritage.

1 Introduction

Personalization is becoming increasingly important in many industries, including e-commerce, entertainment, and education and business is reacting to the needs of their users more rigidly than the rest of the areas. Being able to see a tailored content based on one's needs and interests is a must, considering that we are living in the era of big data - having too many resources that can be useful for a learner. Moreover, due to the rapid digitalization of cultural content, there is an enormous amount of data that is in a variety of formats and not properly categorized which makes the need for personalized approach crucial. A system that allows users to access cultural content based on their context (e.g.: background, needs or social status) is no longer impossible as the advancements in technology and data analysis have made it possible to provide customized experiences to each consumer of content. An essential component for implementing personalized cultural experience for users is the recommendation system, as it helps match learners with the right materials and resources that fit their needs and interests. However, as mentioned the enormous amount of resources raises two problems - proper management and organization of the content and providing a personalised flow of data that could help with obtaining the right resources more easily. In this paper, we are discussing what approaches and how can the flow be optimized for a better learning curve (Liqiang & Quan, 2019).

Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. Vol. 13, Sofia, Bulgaria: Institute of Mathematics and Informatics – BAS, 2023. ISSN: 1314-4006, eISSN: 2535-0366

2 The Need for Personalized Experience

Traditional academic systems often rely on a one-size-fits-all approach, which can lead to a lack of engagement and motivation for learners. With the advent of technology and the increasing availability of data, personalization has become a paramount aspect of modern systems (Fayyaz, Ebrahimian, Nawara, & Ibrahim, 2020). The benefits of personalization for learning and processing content are numerous. These systems can provide learners with materials that are tailored to their specific learning needs and interests, improving their engagement and overall performance (Liqiang & Quan, 2019). By providing learners with content that is tailored to their specific needs, they are more likely to engage with the material and retain it for longer periods. Furthermore, learners who receive personalized instruction are more likely to develop a sense of ownership over their learning, which can lead to greater motivation and a stronger desire to learn.

In the next sections, we are going to present static and dynamic approaches for a personalized experience. Static personalization can be fulfilled through user settings, and a dynamically tailored personalized experience can be achieved through recommendation systems, which are based on machine learning algorithms that analyse large amounts of data to identify patterns and relationships. In education, these systems can use data from a variety of sources, such as learners' prior performance and their behaviour, to recommend content that is most likely to be of interest and benefit to them. Moreover, personalized learning of cultural content refers to the adjustment of the educational experiences to the unique needs, interests, and abilities of each learner. This approach recognizes that each learner has a different learning style and pace and that a one-size-fits-all approach is no longer effective.

The algorithms used in these systems can also take into account a range of other factors, such as learning style, preferred learning pace, and subject matter expertise, to provide more personalized and accurate recommendations. This data can be used to create a unique learning path for each learner, ensuring that they receive content that is both relevant and challenging.

There are several benefits to using recommendation systems for cultural heritage content, including:

- Improved engagement: Personalized experience can provide learners with materials that are specifically designed to match their interests, which can increase their engagement and motivation to learn.
- Increased learning outcomes: By providing learners with content that is tailored to their individual needs, recommendation systems can help improve their academic performance and overall learning outcomes.
- Data-driven decision-making: Recommendation systems can provide valuable insights into learners' performance and engagement, which can be used to inform instructional decisions and improve the learning experience.

In the next section, we are going to summarize the different approaches to recommending the correct content and we are going to state the differences, advantages and disadvantages of those approaches.

3 Personalization Approaches

Providing recommendations of the correct cultural content to learners can be done in various ways - statically by directly asking the users for their preferences and adding keywords in their profiles that correspond to their interests (Stefanov, Boychev, Stefanova, & Georgiev, 2011), using simple statistical methods - generating a list of cultural objects that are of interest of most of the users or that reflect a certain area of interest. The more advanced way to provide personalized content is by using adaptive methods - most of which involve a recommendation system. In this section we are providing an overview of the different approaches.

3.1 Static Methods

We refer to the first group of methods that we are going to present as static methods because the information is either directly provided by the user or gathered through some initial surveys and used afterwards without any dynamic adjustments. In our previous research (Christozov, & Mitreva, 2020; Mitreva, Nikolova, & Georgiev, 2021) we have extensively discussed the options to have user settings that could be updated by the users at any time and will provide information about their interests. Those settings can be viewed as static personalization - once set they could be changed, but are not dynamically changed based on the behaviour of the user. An improvement of that static approach is the availability of interests of the users stated in the profile and if the system gathers information about each cultural object that was accessed by the users, this can generate the data needed by any statistical method to provide a list of mostly accessed objects and offer to the user that list. Even further improvement that can view more as an adaptive rather than static method will be if semantic relationships can be found (Stoikov, 2021) and the lists of most interesting objects are grouped per topics and matched to the interests of the users that are stored in the profile.

The other likewise static method can be based on an initial survey (Deng, Li, Zhang, Ding, & Lam, 2022) or a test that is done by a user to collect useful data about the personal approach of the user.

Although those methods are not dynamic and efficient, they can be especially useful in the early stages of the use of a system, before any dynamic data is accumulated. Furthermore, as suggested in (Liqiang & Quan, 2019) each learner can provide details about their learning style so that the recommendation can be based on those properties. Those techniques can be combined with the adaptive ones after there is enough data to include the recommendation system because most adaptive algorithms work well with a lot of data, but they are lacking when sparse information is available.

3.2 Adaptive Methods

In this section we are discussing adaptive methods most of which involve a machine learning algorithm or even a few algorithms. Personalized content can be offered by developing a recommendation system and they work in three phases - collect information about a user, apply the method to derive meaningful insights and finally recommend content to the user (Narke & Nasreen, 2020). One common approach is to use data and analytics to track learner progress and adapt the learning experience accordingly. This can in-volve using algorithms to analyze data such as behaviour patterns, learner's interactions with the learning environment, and objects that were accessed or even rated. Based on this analysis, the system can recommend new content or activities that are best suited to each learner's needs and abilities.

Another approach to personalization is to give learners more control over their learning. This can involve providing learners with a range of content and activities to choose from or allowing them to create their learning paths. In these cases, learners can select the content and activities that best suit their learning style and pace, which can lead to a more engaged and motivated learning experience.

Profiles and Web Usage Mining. Personalization via profiles can be included in both the static methods and the adaptive methods because we could have different techniques. The static approach was discussed in the previous section - profile set-tings can be set by the user and the content or the system can be altered based on those settings. The dynamic approaches to using profiles could be either web usage mining or a clustering algorithm (e.g. k-nearest neighbour) to determine that the pro-file is similar to a group and provide content that is "interesting" for the group of profiles.

The web usage mining approach was discussed in previous research (Christozov, & Mitreva, 2020; Mitreva, Nikolova, & Georgiev, 2021). The process of web usage mining is the extraction of useful patterns from the weblogs (Ju, & Wang, 2021), thus if the recommendation of learning content should be done in a web environment this technique can be useful. This process consists of several steps - gathering logs, cleaning the logs from all errors and not beneficial information and finally extracting useful rules and patterns of those (Srivastava, Garg, & Mishra, 2015). However, it is only applicable if we have a web system and sometimes the weblogs do not provide enough information that could result in an effective recommendation system. However, this approach could be used as an auxiliary method.

Content-Based and Collaborative Filtering. E-commerce is effectively using recommendation systems - product recommendations based either on previous users' orders or search history (Liao, Sundar, & Walther, 2022). This approach can be applied to learning content in systems storing cultural content by using content-based filtering or collaborative filtering. Content-based filtering is based on similarities of items and the accumulated information about a specific user (Liao, & Sundar, 2022). So, content-based filtering requires a lot of data about one user so that some insight about the user can be generated. Also, the algorithm has data only about what the user has searched for or accessed, which usually is limited by the knowledge of the user about certain items. This means that in a system with a lot of cultural objects, the user should have enough knowledge about the content and the relationships of the items in order for the algorithm to be successful. Although if semantic relations are found (Stoikov, 2021) and if the metadata of the items is properly stored, this algorithm can find useful similar objects for the key interest of a learner that is using this system. However, content-based filtering focuses on the content and if ratings are not provided this approach will not yield good results (Schafer, Frankowski, Herlocker, & Shen, 2007), (Kapembe & Quenum, 2019)- an example of this is an object that is accessed because it is similar to other objects, but it is not what the user is searching for. As a result of the lack of ratings objects similar to the object that is considered not useful are offered. If the system is developed in a way to store not just the preferences of the user, but also their rating, this could make content-based filtering more effective, because large weight can be attributed to items that are useful or match completely the user's needs and small and even negative weights for items that are not interesting for the learner.

Nevertheless, content-based filtering is considered inferior to collaborative filtering, because collaborative filtering is a method of making recommendations based on the preferences of a group of people, rather than relying on a single user's preferences (Schafer, Frankowski, Herlocker, & Shen, 2007), (Liao, & Sundar, 2022). The idea behind collaborative filtering is that people who have similar likings in the past are likely to have similar choices in the future. There are two main types of collaborative filtering techniques: user-based and item-based (Kapembe, & Quenum, 2019). User-based collaborative filtering looks at the preferences of similar users to make recommendations to a given user. Item-based collaborative filtering looks at the preferences of users for similar items to make recommendations.

Collaborative filtering compares the preferences of users to find similarities and uses these similarities to make recommendations. In practice, this involves calculating a similarity score between users or items and using this score to determine which items are likely to be of interest to a given user. One of the key challenges in implementing collaborative filtering is handling the large amounts of data required to calculate similarity scores and make recommendations. This data can include user preferences, ratings, and demographic information, and can come from a variety of sources, such as surveys, website clicks, and purchase history.

Based on Schafer et al. (2007) collaborative filtering is considered the most effective approach to provide personalized experience and considering the rest of the methods that we have presented, it is indeed the technique that we think will yield the best results despite the fact that even that approach has some challenges and disadvantages. This is be-cause if enough data is collected from different users, adequate recommendations can be provided for things that similar users are interested in. For example, if we are interested in certain cultural items that are of some category and other users are interested in the same category, there is a high chance another category that is interesting for those users to be interesting to us. Of course, the proper data needs to be collected about the behaviour of the users in order for that approach to work. Similar to what was done in another system - ShareTEC (Stefanov, Boychev, Stefanova, & Georgiev, 2011) - the data for the actions and behaviour of the users should be stored separately and some calculations must be done asynchronously because running algorithms on large amounts of data sometimes take more time than acceptable and the result should be provided immediately.

Although we consider collaborative filtering as the best approach among all the methods we have researched and presented here in the next section we consider all of the challenges and benefits of them all and suggest the creation of a hybrid approach - a combination of both static and adaptive methods.

4 Challenges and Benefits of the Different Approaches

The static methods for personalization are quite rudimentary they are not providing much personalization of content rather than predefined options in the profile of the user. Web usage mining has limited effectiveness, because of the tiresome process of gathering the logs and the process of cleaning and processing them. That approach can definitely be used and will provide adaptive content, but it can be done on a web-based system, which logs we could collect and process. Sometimes the actual gathering of the logs might not be trivial, thus the process might not that straightforward.

The adaptive methods, on the other hand, can provide unsatisfactory results in the early stages of the collection of the information and at the same time if there is too much data it can be also a problem because the recommendation should be calculated and produced in real-time (Fayyaz, Ebrahimian, Nawara, & Ibrahim, 2020). Another issue with those approaches is the fact that users tend to rate a limited number of items, so the method cannot differentiate be-tween positive and negative examples. However, based on our research admittedly collaborative filtering can produce the best results, especially if there is enough data about the users so that the proper group of users with similar interests can be identified.

As mentioned in Fouad et al. (2022) just one recommendation approach probably will provide good results only for specific data, but not in all cases. Based on some things we have researched in ShareTEC (Stefanov, Boychev, Stefanova, & Georgiev, 2011) a good approach will be a hybrid approach. As a first step before there is enough data is collected to use collaborative filtering, we can use some of the static approaches - user preferences, keywords for interests, etc. Based on the user preferences and the interests of the users we could offer a list of items that are mostly accessed - either globally or grouped by a certain area or interest - this could be done by only using simple statistical methods. This could offer enough initial personalization so that a learner can start exploring different cultural objects and data. Of course, cultural objects need to be stored with as much metadata and additional data as possible so that the objects are not just big data, but something that could be useful. Also, as much information about the behavior of the users and ratings for the different objects needs to be stored and processed in a way that the data is easily used for real-time recommendations. After enough data is collected preferably with ratings for the content or enough user information so that the user can be allocated to the proper cluster of users, then collaborative filtering to be used.

5 Future Work and Conclusion

Providing personalization is an important tool for cultural heritage content due to the different needs and interests of the users. To improve the engagement of the users, data analysis and machine learning algorithms are used to provide learners with content that is tailored to their individual needs. With the continued growth of technology and data analysis, the use of personalization in all areas is likely to become more widespread. As a next step in our research, we are planning to implement that hybrid approach and test it with the available data.

Acknowledgements.

This work was partially supported by the Bulgarian Ministry of Education and Science under the National Research Programme "Young scientists and postdoctoral students - 2" approved by DCM 206 / 07.04.2022.

References

Christozov, D., & Mitreva, E. (2020). Trust in learning from big data: the two sides of the same coin. *Issues in Information Systems, 21*(1), 147-152, https://doi.org/10.48009/1_iis_2020_147-152

Deng, Y., Li, Y., Zhang, W., Ding, B., & Lam, W. (2022). Toward Personalized Answer Generation in E-Commerce via Multi-perspective Preference Modeling. *ACM Transactions on Information Systems, 40*(4), 1-28, https://doi.org/10.1145/3507782

Fayyaz, Z., Ebrahimian, M., Nawara, D., & Ibrahim, A. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Sciences, 10*(21), Article 7748. https://doi.org/10.3390/app10217748

Fouad, M., Hussein, W., Rady, S., Yu, P. S., & Gharib, T. (2022). A Hybrid Recommender System Combining Collaborative Filtering with Utility Mining. *International Journal of Intelligent Computing and Information Sciences, 22*(4), 13-24. https://doi.org/10.21608/ijicis.2022.145103.1192

Ju, H., & Wang, H. (2021). Application analysis of computer web data mining technology in E-commerce. In *5th International Conference on Electronic Information Technology and Computer Engineering* (pp. 1233-1238). https://doi.org/10.1145/3501409.3501626

Kapembe, S., & Quenum, J. (2019). A Personalised Hybrid Learning Object Recommender System. In *11th International Conference on Management of Digital EcoSystems* (pp. 242-249). https://doi.org/10.1145/3297662.3365810

Liao, M., Sundar, S., & Walther, J. (2022). User trust in recommendation systems: A comparison of content-based, collaborative and demographic filtering. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)* (Article 486, pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3491102.3501936

Liao, M., & Sundar, S. S. (2022). When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *Journal of Advertising*, *51*(2), 256-267, https://doi.org/10.1080/00913367.2021.1887013

Liqiang, H., & Quan, L. (2019). Design of resource recommendation model for personalized learning in the era of big data. *Annual Meeting on Management Engineering* (pp. 181-187). https://doi.org/10.1145/3377672.3378054

Mitreva, E., Nikolova, A., & Georgiev, V. (2021). Web Mining Techniques Applicable for Cultural Heritage Observations. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *11,* 253-260. https://doi.org/10.55630/dipp.2021.11.22

Narke, L., & Nasreen, A. (2020). A comprehensive review of approaches and challenges of a recommendation system. *International Journal of Research in Engineering, Science and Management*, *3*(4), 381-384, https://www.ijresm.com/Vol.3_2020/Vol3_Iss4_April20/IJRESM_V3_I4_91.pdf

Schafer, J., Frankowski, D., Herlocker, J., & Shen, S. (2007). Collaborative filtering recommender systems. In: P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web. LNCS, vol. 4321* (pp. 291–324). Springer. https://doi.org/10.1007/978-3-540-72079-9_9

Srivastava, M., Garg, R., & Mishra, P. (2015). Analysis of data extraction and data cleaning in web usage mining. *International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)* (Article 13, pp. 1-6). https://doi.org/10.1145/2743065.2743078

Stefanov, K., Boychev, P., Stefanova, E., & Georgiev, A. (2011). Digital Libraries in Teacher Education. *Fortieth Jubilee Spring Conference of the Union of Bulgarian Mathematicians* (pp. 120-135). https://research.uni-sofia.bg/handle/10506/762

Stoikov, J. (2021). Using Conditional Probability for Discovering Semantic Relationships be-tween Named Entities in Cultural Heritage Data. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *11,* 77-88. https://doi.org/10.55630/dipp.2021.11.7

Received: March 15, 2023
Reviewed: April 06, 2023
Finally Accepted: May 17, 2023