Personalized Context-Aware Recommender System for Travelers[1]

(Discussion Paper)

Mahsa Shekari, Alireza Javadian Sabet, Chaofeng Guan, Matteo Rossi,
Fabio A. Schreiber and Letizia Tanca

Politecnico di Milano - Dipartimento di Elettronica, Informazione e Bioingegneria

Abstract

Nowadays, traveling has become more convenient thanks to many recent technological advancements. However, the main problem is that, with non-customized offers, the risk for the travelers is to waste their time looking for the most appropriate one. Consequently, travelers need a system capable of understanding their contextual preferences for ranking travel offers accordingly. In this work, we propose The Hybrid Offer Ranker (THOR) as a possible solution: on the one hand, we employ various classification algorithms to learn the individuals' contextual preferences; on the other hand, to help new users the system has no information on (cold users), we employ unsupervised algorithms to identify clusters of users with similar preferences and build group preference models accordingly.

Keywords

Context-Awareness, Data Mining, Recommender System, Mobility, User Modeling, Personalization

1. Introduction

Recent technological advancements have made traveling more convenient and more personalized (think, e.g., of sharing mobility, or demand-responsive schemes). However, non-customized travel offers can cause travelers to waste a lot of time in finding the most suitable ones [1]. Recommendation Systems (RS) can help in this regard, as they aim to understand the user's "preference" towards an item [2] and guide them in the offer selection process. Travel preferences may be expressed by *the context in which the traveler interacts with the system* [3, 4], where *context* is any information that can be used to characterize an entity's situation (e.g., person, place, physical environment) [5]. A system is context-aware if it uses context to provide the user with relevant information and services, where relevance depends on the user's task.

Unlike many works which focused on recommending only touristic destinations [6] to travelers, in this work we rank a list of complete travel offers for the travelers according to

[1]This work was supported by Shift2Rail and the EU Horizon 2020 research and innovation programme under grant agreement No: 881825 (RIDE2RAIL)

SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy

mahsa.shekari@polimi.it (M. Shekari); alireza.javadian@mail.polimi.it (A. Javadian Sabet); chaofeng.guan@mail.polimi.it (C. Guan); matteo.rossi@polimi.it (M. Rossi); fabio.schreiber@polimi.it (F. A. Schreiber); letizia.tanca@polimi.it (L. Tanca)

0000-0002-5191-2961 (M. Shekari); 0000-0001-9459-2411 (A. Javadian Sabet); 0000-0002-9193-9560 (M. Rossi); 0000-0002-5237-0455 (F. A. Schreiber); 0000-0003-2607-3171 (L. Tanca)

their context. This work has been carried out within the Ride2Rail (R2R)[1] project, which is part of the Innovation Programme 4 (IP4) of the Shift2Rail (S2R) initiative[2], which aims at advancing the European transportation domain. The so-called Travel Companion (TC), which is one of the contributions of the S2R IP4 ecosystem, is an application that can operate on various devices to assist travelers before, during, and after each journey. To connect rural areas with transportation hubs, the S2R ecosystem, through the TC, allows drivers to share their rides.

In this work, we continue the implementation of the recommender system introduced by Javadian et al. [4]. The system employs the Context Dimension Tree (CDT) methodology [7, 8] to design a contextual preference model presented in [3]. More precisely, this work aims at answering the following questions:

1: How can we create a personalized preference model for ranking travel offers according to the contextual preferences of the traveler using their historical records?

2: How can we create an initial preference model for new travelers using other travelers' data with similar characteristics?

To implement a personalized Context-Aware RS for Travelers, we propose The Hybrid Offer Ranker (THOR). THOR models the problem as a *binary classification problem* which predicts if the traveler will buy any of the available travel offers or not. Using each traveler data, a unique *contextual preference model* is built for each traveler. A ranking mechanism is designed in case of multiple travel offers, which assigns the probabilities to rank the offers according to the user's personal preferences. In addition, the proposed system utilizes clustering algorithms to build a preference model for each group by recognizing users with similar desires. Finally, by generating an extensive dataset considering some precise rules, we test the system's performance (in terms of both computation time and accuracy). The developed RS showed promising results in terms of the accuracy of its predictions.

The rest of the work is organised as follows. Sec. 2 presents the related work. Sec. 3 details the proposed system and validation procedure. Finally, Sec. 4 concludes and outlines some future work.

2. Background and Related Work

RSs typically fall into one of the following categories [9]: (*i*) *Content-based*, which provide recommendations based on the user's past purchases; (*ii*) *Collaborative*, which recommend based on other users with similar preferences; and (*iii*) *Hybrid*, which combine (*i*) and (*ii*).

The framework developed by Sebasti et al. [6] collects users' information and general preferences by asking them to enter their data and introduce their specific preferences for the current visit; then, it generates a list of activities that are likely of interest to the user. A hybrid RS first classifies users into different classes; then, according to the user's records, a content-based approach recommends places, and a filter selects the fitting places considering the current request. Javadian et al. [10] propose a data-mining-based RS to rank the offers by considering a score according to the characteristics of the user's mobility request. The proposed method by Fang et al. [11] automatically generates temporal feature vectors from a

[1]https://ride2rail.eu
[2]http://www.shift2rail.org/

collection of documents based on Wikipedia and Twitter. Additionally, Coelho et al. [12] employ travel-related tweets to personalize recommendations regarding POIs for the user. In this work, historical buildings are considered for identifying POIs and classification models. Consonni et al. [13] collect travel-centered mobility data and characterize the time spent traveling in terms of the activities performed, i.e., *fitness*, *enjoyment*, and *productivity*; understanding the travelers and their choices according to the nature of their travel time can help the system learn the preferences more accurately. In [14] the sorting and selection actions are provided as perspectives related to trip search options. This allows the system to rapidly learn the users' preferences while they select different offers.

A content-based RS needs to access a sufficient number of users' records in order to select the user's preferences. As a result, a new user might not receive accurate recommendations if they have very few records. The "cold start problem" [15] occurs when there is not enough historical data. Some techniques mentioned in [16] employ strategies based on each item's popularity and entropy on the user profile to determine the best recommendation items for new users. Another approach is recommending to a new user the top popular offers. This can increase the user's purchase likelihood, but also decreases the level of personalization. Finally, collaborative methods can be helpful from the personalization point of view, but the recommendations' precision might remain quite low.

In this work, we combine collaborative and content-based methods to develop a novel *hybrid* approach. We consider a content-based method to find appropriate travel offers for new users. Therefore, we use a collaborative approach to build recommender models for various groups of users with similar characteristics and use them to provide recommendations to new users.

3. The Hybrid Offer Ranker (THOR)

We call the proposed RS The Hybrid Offer Ranker (THOR). In this section we provide the reader with some details of THOR's implementation and functions.

The main function offered by THOR allows for the ranking of travel offers presented to TC users according to their contextual preferences. Figure 1 provides an overview of the THOR system. The main procedure is as follows. As soon as a TC submits a desired source and destination, as well as potential contextual preferences (mobility request), the Travel Service Provider (TSP) returns a list of non-ranked travel offers satisfying the user's request. In the next step, for each of the travel offers, the Offer Categorizer module (OC) computes category scores (e.g., *environmentally friendly* and *comfortable*). The resulting *enriched travel offer* will be used by THOR for learning and ranking. Then, the enriched offers are combined with the user's most recent profile information which will be used as the Ranker's input data.

In order to provide **Context-Aware** recommendations to the users, THOR employs the aspects defining the users' choice criteria. To identify these criteria, THOR employs the CDT methodology which models the context as a rooted tree (See Fig. 2). The idea behind this modeling approach is that it builds the model with the most general context as the top-level nodes and gradually breaks the context into atomic characteristics while traversing the tree. In our case, the main context nodes are as follows.

- **User Profile** contains the static information of the user. For example, the personal data,

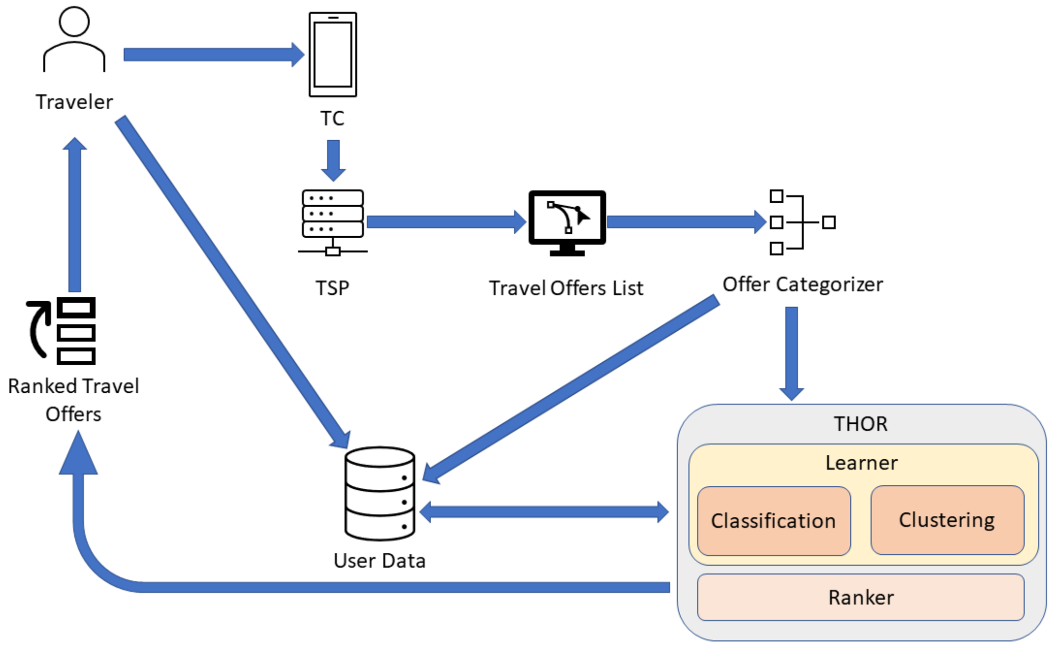

Figure 1: High-level representation of THOR.

profile type, disability information, etc.

- **Context** contains the momentary information related to each of the mobility requests. For example, the time (e.g., day vs. night), if the user was accompanying anyone, etc.
- **Search Options** represent specific explicit requests by the user while submitting a mobility request. For example, if they prefer specific means of transportation (e.g., airplane).
- **Available Travel Offers** encompasses the list of travel offers characterized by their offer category scores as well as their leg-level information. This node helps us to understand for a given mobility request, which travel offers were presented to the user and what was the user choice among them.

Combining the mentioned context information enables THOR to better learn the contextual preferences of each user separately. To exemplify, please consider a female user who mostly travels for leisure with her partner in economy class flights. After giving birth to her child, she changes her preference to train because of the comfort for her child. THOR learns such contextual preferences and when she sends a mobility request accompanying only her partner, probably the economy flights will be ranked higher, while if she accompanies her child, offers containing train could be ranked higher. It should be noted that, in this example, for brevity we simplified the contextual preferences. In reality, the user choice is affected by a non-linear combination of the mentioned features.

Learner module is in charge of regularly updating each user's personal preference model; it utilizes various classification algorithms (KNN [17], SVC [18], DT [19], RF [20], and LR [21]) to build classifiers which take as input the historical data of a single user and outputs the user's

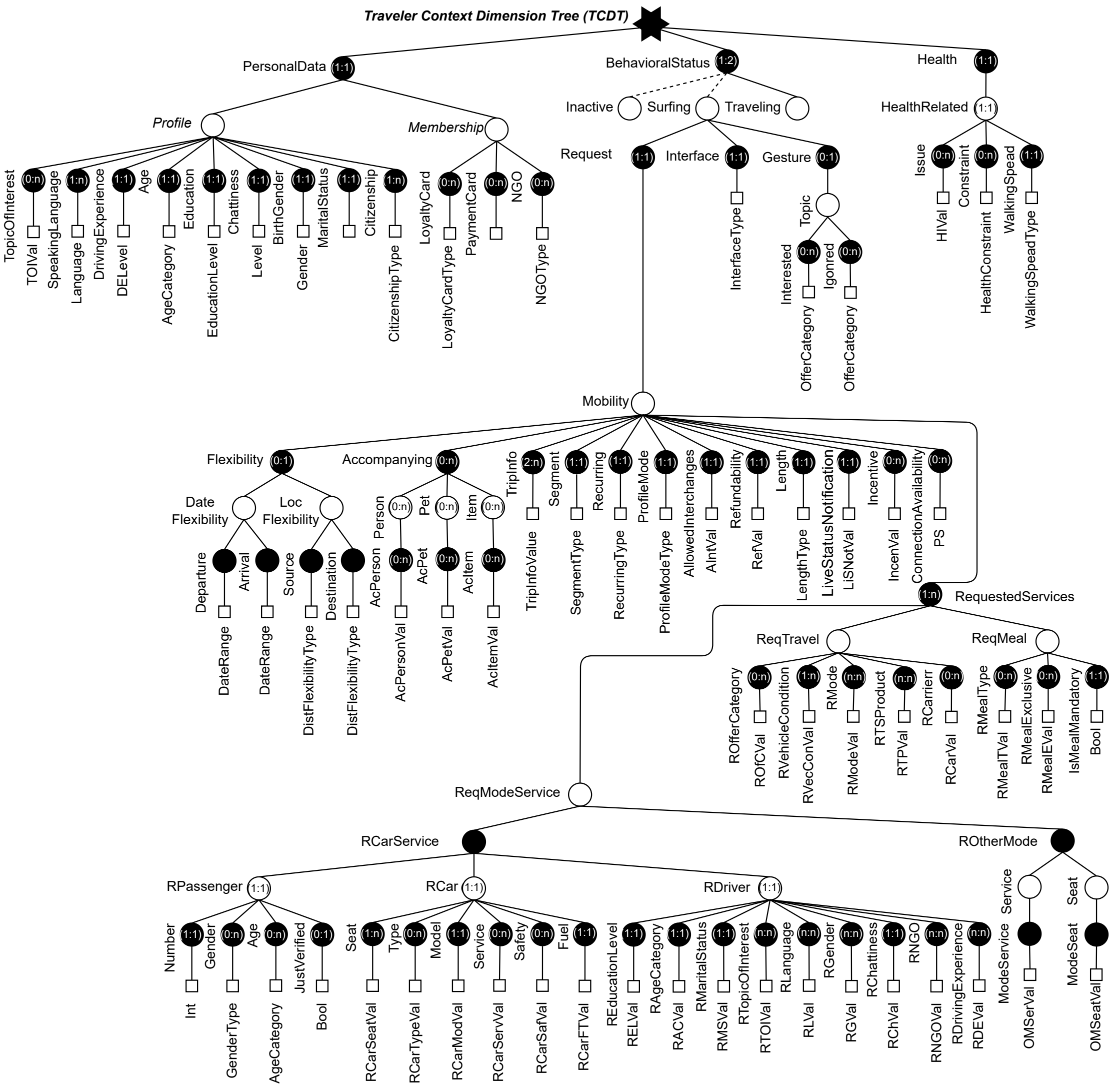

Figure 2: Proposed Traveler Context Dimension Tree. Black circles, white circles, and white squares represent, respectively, dimensions, concepts and parameters.

contextual preference model. Concerning the newly registered users for whom the system has no historical data, the Learner module invokes its clustering block. A clustering block is designed using *K-means* [22] and *Density-based spatial clustering of applications with noise (DBSCAN)* [23] algorithms. The idea is that, at first, using the user profiles of all the users, THOR identifies clusters of users with similar profile information. In the next step, for each cluster, it collects the historical data of all the users belonging to that group. Then, using the classification block mentioned earlier, it learns a group-level contextual preference model. Having a clustering model and group-level contextual preference models for groups in hand, as soon as a new user registers to the system, the Learner module identifies the group which the user belongs to. Finally, the group-level contextual preference model will be used as the user's personal preference model. The model will be in use until the system collects enough historical data from the user for building a pure personal contextual preference model.

The **Ranker module** is the main interface for computing the rankings of enriched travel offers. At first the Ranker retrieves the corresponding recommender model of the user as well as the enriched travel offers. In the next step, the Ranker predicts if the user will buy or not any of the offers and saves the results. Finally, the system computes the ranking of the list of travel offers through the confidence score of the predictions and presents them to the user.

Since the TC is going to be used by the users in the near future, we do not have access to any real data. As a result, to test the performance of the system, both from the computation time and accuracy points of view, we generated a dataset according to pre-defined distributions and rules. Having a rule-based generated dataset enabled us to compare the performance of the various classifiers and the system as a whole. On a fully controlled setting in which user's historical data were carefully inserted using human supervision, THOR reaches 95% accuracy.

4. Conclusion and Future Work

In this work, we tackled the issue of personalization of the user experience in the transportation domain by designing and implementing THOR, a system that: (i) is able to learn individuals contextual preferences using their historical records (RQ1); and (ii) employs the wisdom of crowds to tackle the so-called cold start problem (RQ2). THOR ranks the travel offers for each individual using the learned models.

In future works, we plan to use appropriate feature selection methods [24] to reduce the complexity of the models. We also plan to build the social media (SM) core proposed in [4] as a tool to characterize urban mobility patterns [25]. Having an appropriate SM core can help decision-makers understand the user preferences during events [26, 27] that bring many travelers to specific European cities.

References

[1] X. Chen, Q. Liu, X. Qiao, Approaching another tourism recommender, in: 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2020, pp. 556–562. doi:10.1109/QRS-C51114.2020.00097.

[2] F. Ricci, L. Rokach, B. Shapira, Introduction to Recommender Systems Handbook, Springer US, Boston, MA, 2011, pp. 1–35. doi:10.1007/978-0-387-85820-3_1.

[3] A. J. Sabet, M. Rossi, F. A. Schreiber, L. Tanca, Context awareness in the travel companion of the shift2rail initiative, in: Proceedings of the 28th Italian Symposium on Advanced Database Systems, volume 2646 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 202–209. URL: http://ceur-ws.org/Vol-2646/15-paper.pdf.

[4] A. Javadian Sabet, M. Rossi, F. A. Schreiber, L. Tanca, Towards learning travelers' preferences in a context-aware fashion, in: Ambient Intelligence – Software and Applications, Springer International Publishing, Cham, 2021, pp. 203–212. doi:10.1007/978-3-030-58356-9_20.

[5] A. Dey, Understanding and using context, Personal and Ubiquitous Computing 5 (2001) 4–7. doi:10.1007/s007790170019.
[6] L. Sebastia, I. Garcia, E. Onaindia, C. Guzman, e-tourism: A tourist recommendation and planning application, in: 2008 20th IEEE International Conference on Tools with Artificial Intelligence, volume 2, 2008, pp. 89–96. doi:10.1109/ICTAI.2008.18.
[7] C. Bolchini, E. Quintarelli, L. Tanca, Carve: Context-aware automatic view definition over relational databases, Inf. Syst. 38 (2013) 45–67.
[8] P. Agnello, S. M. Ansaldi, E. Lenzi, A. Mongelluzzo, D. Piantella, M. Roveri, F. A. Schreiber, A. Scutti, M. Shekari, L. Tanca, Reckon: a real-world, context-aware knowledge-based lab, in: 29th Italian Symposium on Advanced Database Systems, SEBD 2021, volume 2994, CEUR-WS, 2021, pp. 466–473.
[9] Y. Cai, H. F. Leung, Q. Li, H. Min, J. Tang, J. Li, Typicality-based collaborative filtering recommendation, IEEE Transactions on Knowledge & Data Engineering 26 (2014) 766–779.
[10] A. J. Sabet, S. Gopalakrishnan, M. Rossi, F. A. Schreiber, L. Tanca, Preference mining in the travel domain, in: 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2021, pp. 358–365. doi:10.1109/ICAICA52286.2021.9498231.
[11] G.-S. Fang, S. Kamei, S. Fujita, Automatic generation of temporal feature vectors with application to tourism recommender systems, in: 2016 Fourth International Symposium on Computing and Networking, 2016, pp. 676–680. doi:10.1109/CANDAR.2016.0121.
[12] J. Coelho, P. Nitu, P. Madiraju, A personalized travel recommendation system using social media analysis, in: 2018 IEEE International Congress on Big Data (BigData Congress), 2018, pp. 260–263. doi:10.1109/BigDataCongress.2018.00046.
[13] C. Consonni, S. Basile, M. Manca, L. Boratto, A. Freitas, T. Kovacikova, G. Pourhashem, Y. Cornet, What's your value of travel time? collecting traveler-centered mobility data via crowdsourcing, 2021. arXiv:2104.05809.
[14] L. Boratto, M. Manca, G. Lugano, M. Gogola, Characterizing user behavior in journey planning, Computing 102 (2020). doi:10.1007/s00607-019-00775-8.
[15] A. I. Schein, A. Popescul, L. H. Ungar, D. M. Pennock, Methods and metrics for cold-start recommendations, in: Proc. of the 25th ann. int. ACM SIGIR conf. on Research and dev. in IR, 2002, pp. 253–260.
[16] A. Sang, S. K. Vishwakarma, A ranking based recommender system for cold start data sparsity problem, in: 10th IC3, 2017, pp. 1–3. doi:10.1109/IC3.2017.8284347.
[17] H. b. Jaafar, N. b. Mukahar, D. A. Binti Ramli, A methodology of nearest neighbor: Design and comparison of biometric image database, in: 2016 IEEE Student Conference on Research and Development (SCOReD), 2016, pp. 1–6. doi:10.1109/SCORED.2016.7810073.
[18] L.-S. Lan, M-svc (mixed-norm svc) - a novel form of support vector classifier, in: IEEE ISCAS, 2006, pp. 4 pp.–3264. doi:10.1109/ISCAS.2006.1693321.
[19] Y. Yu, F. Zhong-liang, Z. Xiang-hui, C. Wen-fang, Combining classifier based on decision tree, in: WASE ICIE, volume 2, 2009, pp. 37–40. doi:10.1109/ICIE.2009.12.
[20] Z. Bingzhen, Q. Xiaoming, Y. Hemeng, Z. Zhubo, A random forest classification model for transmission line image processing, in: 15th ICCSE, 2020, pp. 613–617. doi:10.1109/ICCSE49874.2020.9201900.
[21] X. Zou, Y. Hu, Z. Tian, K. Shen, Logistic regression model optimization and case analysis,

in: IEEE 7th ICCSNT, 2019, pp. 135–139. doi:10.1109/ICCSNT47585.2019.8962457.
[22] S. Lu, H. Yu, X. Wang, Q. Zhang, F. Li, Z. Liu, F. Ning, Clustering method of raw meal composition based on pca and kmeans, in: 2018 37th Chinese Control Conference (CCC), 2018, pp. 9007–9010. doi:10.23919/ChiCC.2018.8482823.
[23] A. Smiti, Z. Elouedi, Dbscan-gm: An improved clustering method based on gaussian means and dbscan techniques, in: 2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES), 2012, pp. 573–578. doi:10.1109/INES.2012.6249802.
[24] A. Brankovic, M. Hosseini, L. Piroddi, A distributed feature selection algorithm based on distance correlation with an application to microarrays, IEEE/ACM TCBB 16 (2019) 1802–1815. doi:10.1109/TCBB.2018.2833482.
[25] M. Manca, L. Boratto, V. Morell Roman, O. Martori i Gallissà, A. Kaltenbrunner, Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study, OSNEM 1 (2017) 56–69. doi:https://doi.org/10.1016/j.osnem.2017.04.002.
[26] A. Javadian Sabet, M. Brambilla, M. Hosseini, A multi-perspective approach for analyzing long-running live events on social media. a case study on the “big four” international fashion weeks, Online Social Networks and Media 24 (2021) 100140. doi:https://doi.org/10.1016/j.osnem.2021.100140.
[27] M. Brambilla, A. Javadian Sabet, M. Hosseini, The role of social media in long-running live events: The case of the big four fashion weeks dataset, Data in Brief 35 (2021) 106840. doi:https://doi.org/10.1016/j.dib.2021.106840.