Enhancing Personalized Travel Recommendations: Integrating User Behavior and Content Analysis

Karolina Alicja Selwon
Gdansk University of Technology and QTravel.AI
Gdańsk, Poland *karselwo@pg.edu.pl*

Julian Szymański
Gdańsk University of Technology
Gdańsk, Poland *julian.szymanski@pg.edu.pl*

Abstract

In the research presented in this paper, we focus on overcoming the obstacles of delivering personalized travel recommendations in the tourism sector. The paper introduces a three-part contribution: initially, it delves into the distinctive challenges of making recommendations in tourism and presents a framework to improve the ranking of trip options in tour operators' search engines. We also propose an innovative method that utilizes the behaviors of tourists and the descriptive content of travel offers to compile a dataset rich in insights about the travel industry. Furthermore, we prove that enhancing listwise learning-to-rank algorithms with an attention mechanism for selecting features significantly boosts the effectiveness of the model beyond traditional probabilistic ranking methods. The research concludes by assessing these ranking models and shedding light on the intricacies of recommending travel offers in the tourism industry.

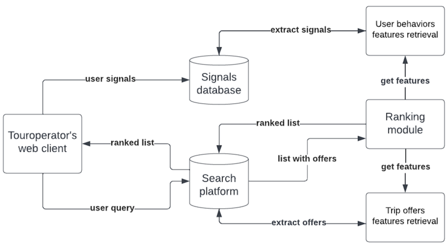
Keywords: learning to rank, travel recommendations, feature engineering

1. Introduction

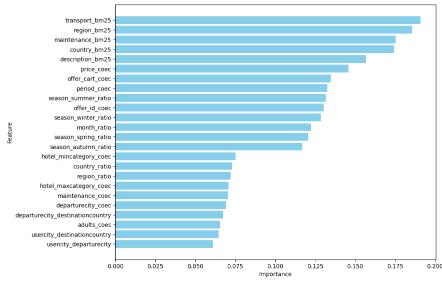
Travel search engines face the challenge of providing effective travel recommendations. The key to implementing a recommendation system in the tourism industry is continuously gathering data for the tourist offer search engine and understanding the customer's reasoning. To face these requirements, in this paper, we propose a new search engine-based adaptable framework, which manages to retrieve travel tour offers and uses user behaviors to train an effective algorithm to re-ranking travel offers. Moreover, we propose an approach to creating a dataset consisting of valuable information related to the features of travel offers and travel users' behavioral data. We reveal that obtaining more appropriate travel user characteristics and performing the most critical features for trip offer recommendations is essential for market understanding. We performed extensive experiments and compared the proposed feature selection model for re-ranking against multiple robust LTR baselines over our dataset. The framework has been developed based on the analysis of commercial data obtained from one of the leading tour operators in the Polish market. The objectives of implementing this framework are to increase the tour operator's competitiveness in the market and encourage tourists to use a search engine.

2. Methodology

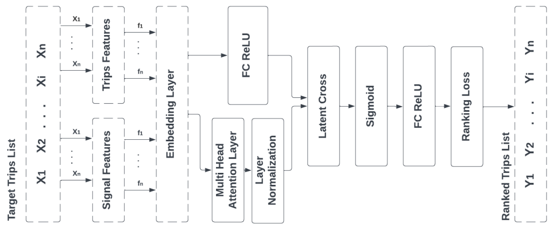
Our framework addresses the challenge of anticipating the optimal ranking of offers for a given search query in tour operators' search engines by leveraging user behavior data. To achieve this objective, the framework utilizes both textual features of travel offers and data retrieved from user's click behaviors. We used collected data to select relevant features during the feature engineering stage, evaluate ranking models, compare the model's performance, and assess the impact of various features on the model results. The framework includes the modules that perform the trip offers features retrieval, the user behavior features retrieval, and the ranking module that delivers ranked offers lists for search engine platforms. The user activities are tracked and stored in a real-time database, while data on tour operators' offers are collected in a large-scale platform for data processing and searching. In the proposed travel offers re-ranking model architecture, we first extracted features into vector embedding and encoded features using an embedding layer. We used a transformer block consisting of a multi-head attention model and latent cross idea [1] to distinguish the influence of each feature. The probability of selecting each feature was obtained by taking a sigmoid function. For the reranking problem, we used the listwise approach with regularization.



(a) Schema of the proposed framework.



(b) Permutation feature importance of created trip-related features.



(c) An illustration of the model architecture.

2.1. Features retrieval

In this experiment, the content relevance is measured by the BM25 scores [2]. Various click-related features were computed for retrieved queries. To address the position bias problem, we used a position-normalized statistic COEC [3]. Through the process of feature engineering, we were able to establish a set of features. Specifically, we extracted 24 features from the user's query and behaviors as follows: the popularity of destination countries based users departure city; popularity of destination countries based users home city; popularity of departure city for users home city; COEC scores of: maximum hotel rating expected by user, minimum hotel rating expected by user, based on the number of adults, based on the departure cities, based on the maintenance regarding meal plans, the popularity of destination regions within countries, the popularity of destination countries, the popularity of months, the popularity of four seasons, popularity of the offers, reservations of the offers, price ranges and range of days; BM25 scores of: description, maintenance, transport, region and country for given search query.

2.2. Travel dataset description

The dataset built for this research is a real-world learning-to-rank dataset created from travel offers and activities of search engine users. The dataset is a representative sample of over a year period of time. Queries are randomly sampled as a representative sample of 20000 user searches, where at least one user click was recorded. Each query related to a user in a given

Table 1. Comparison of LTR algorithms. The best performance is highlighted in boldface.

	nDCG@5	nERR@5	AP@5	P@5
RankNet	0.6706	0.6621	0.4916	0.3887
ListMLE	0.6858	0.6836	0.5040	0.3926
ListNet	**0.6916**	**0.6932**	0.5083	0.3925
STListNet	0.6857	0.6857	0.5013	0.3906
LambdaRank	0.6910	0.6928	0.5074	0.3921
IRGANPointwise	0.6638	0.6744	0.4838	0.3827
IRGANPairwise	0.6676	0.6721	0.4913	0.3896
Ours	0.6759	0.6673	**0.5423**	**0.4165**

session has a unique identifier. The collected dataset contains a sample of 1240 unique trip offers. We used only those offers that were displayed by users. The lists of offers in the dataset have unequal lengths, where the average number of documents per query is 10. The minimum and maximum number of documents per query is 2 and 49 respectively. The relevance label is assessed on a scale from 0 (less significant) to 3 (most significant). For each query-related offer trip, we generated a vector representation, containing real values, along with an associated relevance label. Standardization was performed before introducing the features to the learning algorithm.

2.3. Baseline Methods and Evaluation Metrics

We conduct experiments on the benchmarking platform PT-Ranking library [5] to ensure the reproducibility of the results. We have compared our method with multiple existing learning-to-rank models: pointwise wersion of IRGAN [12] approach; RankNet [6] and IRGAN [12] pairwise methods; the listwise methods include ListNet [4], ListMLE [7], ST-ListNet [8], LambdaRank [9]. We consider popular evaluation metrics to assess both ranking order and precision of finding relevant offers due to the small ratio relevance labels different than 0 and 1. In particular, we NDCG [10], ERR [11], Precision and Mean Average Precision.

3. Results

Research question 1: Does our method achieve the best performance compared to other baselines? After comparing several models, ListNet was the top-performing model across grade-sensitive evaluation metrics. However, our model demonstrated competitive performance and improved precision-related metrics. Given that both models yielded satisfactory results, this may suggest that we should focus more on building features and acquiring relevant signals rather than on improving the model's architecture.

Research question 2: What features affect model performance, and what is the significance of its features in the context of the tourism industry? By examining the assigned weights to each feature, we identified critical features relevant to travel offers and users' preferences, which gave us valuable insights. We utilized the permutation feature importance technique while considering the correlation between features. By comparing the critical values associated with each feature type, we can observe COEC and BM25 features, emphasizing the synergy between click-through features and textual information in shaping the model's decision-making process.

Ablation study: To assess the impact of different hyperparameters on the model's performance, we experimented with a different number N of encoder blocks, H attention heads, dropout rate p_{drop}, and size d_h of hidden dimensions. We experimented with the following ranges for the hyperparameters: $N \in (1,\ 2,\ 3)$, $H \in (2,\ 4)$, $p_{drop} \in (0.0, 0.1, 0.3)$, $d_h \in (64, 128, 256, 512)$. Best performing configurations observed for AP@5 achieved with $H = 4$,

$N = 2$, $p_{drop} = 0.3$, $d_h = 256$. Best performing configurations for NDCG@5 achieved with $H = 4$, $N = 2$, $p_{drop} = 0.1$, $d_h = 256$. A dropout rate of 0.1 balances underfitting and overfitting, providing robust performance across various configurations.

4. Conclusions

The project aimed to create a model that employs user behavior on a tourism website and the features of tourist offers provided by tour operators. Based on the provided analysis of the study results, we can draw the following conclusions about the tourism industry in the examined market. The study has showcased opportunities for further work utilizing multimodality data for a tourist offer recommendation model. The attention-based model demonstrated in listwise learning to rank problems improves the model's performance compared to the base probabilistic model for ranking. Future research on LTR algorithms in the travel industry should focus on incorporating relevant features that reflect travelers' preferences. Additionally, more ranking methods and industry-related features will be considered in further research.

5. Acknowledgements

The work has been supported partially by funds of the Department of Computer Architecture of the Faculty of Electronics, Telecommunication and Informatics of Gdańsk University of Technology and project CAISE, no. KPOD.05.10-IW.10-0005. The data and research environment was provided by Qtravel.AI company under NCBiR Grant POIR.01.01.01-00-0798/19-00.

References

1. Qin, Z., et al.: Are neural rankers still outperformed by gradient boosted decision trees? In: International Conference on Learning Representations (2020)
2. Robertson, S E., Walker, S., Hancock-Beaulieu, M.: Okapi at TREC-7: Automatic Ad Hoc, Filtering, VLC and Interactive. In Proc. of the 7th Text Retrieval Conference, TREC-7, pp. 199–210 (1998)
3. Cheng, H., Cantú-Paz, E.: Personalized click prediction in sponsored search. Proceedings of the third ACM international conference on Web search and data mining (2010)
4. Cao, Z., et al.:Learning to rank: from pairwise approach to listwise approach. Proceedings of the 24th international conference on Machine learning (2007)
5. Yu, H.-T.: PT-ranking: A benchmarking platform for neural learning-to-rank. arXiv preprint arXiv:2008.13368 (2020)
6. Burges, C., et al.: Learning to rank using gradient descent. Proceedings of the 22nd international conference on Machine learning (2005)
7. Xia, F., et al.: Listwise approach to learning to rank: theory and algorithm. Proceedings of the 25th international conference on Machine learning (2008)
8. Bruch, S., et al.: A stochastic treatment of learning to rank scoring functions. Proceedings of the 13th international conference on web search and data mining (2020)
9. Burges, C. Ragno, R., Le, Q.: Learning to rank with nonsmooth cost functions. Advances in neural information processing systems 19 (2006)
10. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS) 20 (4), 422-446 (2002)
11. Chapelle, O., et al.: Expected reciprocal rank for graded relevance. Proceedings of the 18th ACM conference on Information and knowledge management (2009)
12. Wang, J., et al.: Irgan: A minimax game for unifying generative and discriminative information retrieval models. Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (2017)