Data-Driven Destination Recommender Systems

Linus W. Dietz
Technical University of Munich
Garching, Germany
linus.dietz@tum.de

ABSTRACT

Given vast number of possible global travel destinations, choosing a destination has become challenging. We argue that traditional media are insufficient to make informed travel decisions, due to a lack of objectivity, a lack of comparability and because information becomes out of date quickly. Thus, travel planning is an interesting field for data-driven recommender systems that support users to master information explosion. We present unresolved research questions with working packages for a doctoral project that combines the fields of recommender systems and user modeling with data mining. The core contributions will be a framework that integrates heterogeneous data sources from the travel domain, novel user modeling techniques and constraint-based recommender algorithms to master the complexities of global travel planning.

CCS CONCEPTS

• **Information systems → Data mining**; **Recommender systems**; • **Human-centered computing**;

KEYWORDS

recommender systems, data mining, user modeling, tourism

ACM Reference Format:
Linus W. Dietz. 2018. Data-Driven Destination Recommender Systems. In *UMAP '18: 26th Conference on User Modeling, Adaptation and Personalization, July 8–11, 2018, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3209219.3213591

1 INTRODUCTION AND MOTIVATION

Global travel is booming. Airlines offer connections to all continents except Antarctica, and most destinations can be reached within a day. Affordable mobility has increased travel options, and it has become easier to explore the world and learn about foreign cultures. Given endless opportunities, how can the discretionary traveler make informed decisions about possible destinations?

Typically, tourists use websites, blogs, printed travel guides or travel agencies to make their travel plans. These sources are subjective, of uncertain quality, and can become outdated quickly [27]. Moreover, information about destinations that have not yet exploited for tourism is often too limited to form an opinion about their attractiveness.

In this doctoral project, we want to investigate solutions for a variant of the Tourist Trip Design Problem (TTDP) [27] at a regional granularity. Mining and aggregating domain-related data will enable prospective travelers to independently choose their destinations through informed and independent travel recommendations. We assume that it is not obvious to have accurate expectations from far distant regions. Finally, attractions of specific ethno-cultural destinations are often not tangible but have to be experienced [25].

1.1 Scenario

We envision a user planning a multi-week travel comprising of several destinations within a larger query area. This query region could span across continents, or be a set of specific user-selected areas. Based on user preferences, query area, and boundary conditions (e.g., budget, travel season, and duration), the recommender system should compile a set of regions within the query area that considers the user's temporal and monetary constraints.

1.2 Problem Statement and Research Areas

Based on a query region and user preferences, we want to recommend a personalized travel plan comprising a set of destinations with respective durations of stay. Note that the problem space of such a travel recommender is inherently complex, and we have identified three major research topics related to the target problem.

- **Data-driven recommender system framework** (section 3)
  Destination information, such as attractiveness relative to preferences and costs, must be mined continuously from online sources. The data must be stored and made available in aggregated form for the other components of the system.
- **User modeling** (section 4)
  Relevant domain features must be identified and the user's preferences must be elicited effectively without requiring much effort. The interplay of automated solutions and different user interface concepts are to be evaluated.
- **Constraint-based recommendation algorithms** (section 5)
  Travel is constrained by multiple factors, such as time, costs, season, and visa regulations. In addition, the recommendation algorithm should consider diversity and balance the costs and benefits for visiting another region rather than staying at one location for a longer period.

Throughout this doctorate project, we plan to focus on selected aspects within these research topics. The foundation is a modular framework for destination recommender systems that can handle various subsystems, i.e., domain and user modeling, data warehousing, the recommendation algorithms, and the front-end.

UMAP '18, July 8–11, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5589-6/18/07…$15.00
https://doi.org/10.1145/3209219.3213591

2 RELATED WORK

Research into tourist recommender systems has been conducted for more than 15 years [22]. However, due to the complexities of global travel, existing approaches rather focus on urban trips or recommend fixed travel packages.

2.1 Tourist Trip Recommender Systems

Trip recommendations can be performed at different granularities. The least complex are recommendations for single venues and travel packages. Liu et al. [11] proposed the Tourist-Area-Season Topic (TAST) Model to identify traveler interests and the seasonal suitability of travel regions. In a follow-up paper [10], they evaluated their proposed model and augmented it with relationship information to recommend travel packages to groups. A similar approach introduced by Tan et al. [23] focused on feature selection to identify latent user interests. Using a framework of feature-value pairs to represent users and travel packages to calculate distance metrics, their approach employs collaborative filtering methods without any user ratings. Other approaches [7, 29] attempt to solve variants of the TTDP [27] to recommend a series of points of interest (POIs), typically limited to urban areas. The underlying problem for the TTDP is the Orienteering Problem [26]. Most recently, [6] proposed a fast algorithm for multi-day tourist trips.

Herzog and Wörndl [8] developed a region recommender for personalized continental travel. The user is asked to specify her interests, e.g., *nature & wildlife*, *beaches*, or *winter sports*, as well as potential travel regions and monetary and temporal constraints. Respecting these constraints, a recommendation comprises a set of regions that maximizes the user's preference score, while taking the travel season and diversity of the regions into account. Determining the duration of stay per region is done simultaneously, however, the algorithm applies a static reduction of 5–10% to the preference score per week; thus, it is quite coarse-grained. The destination information comes from several online and offline information sources, which must be incorporated and updated manually.

As opposed to these approaches, we propose to replace expert knowledge with automated data mining techniques, because manual maintenance of global travel data is infeasible.

2.2 Mining Traveler Mobility Patterns

A vast amount of spatial-temporal data is collected from GPS modules in mobile phones. Users can choose to publish such data in location-based social networks (LBSNs) whose general adoption has given opportunity for researchers to analyze human mobility combined with social activities. Since the user's locations and social graph reveal significant information about individual preferences such data has been analyzed to improve recommender systems [3], For example, spatial co-occurrences can be used to identify similar users and generate implicit ratings for collaborative filtering algorithms [30]. In a more elaborate approach [2], travelers in a foreign city are matched to local experts based on their respective home behavior when recommending Foursquare venues. Hsieh et al. [9] used past LBSN data to recommend travel routes along urban POIs. They present an approach to derive the popularity, best time to visit, transit time between venues, and the best order to visit POIs. By analyzing past trips, we aim to improve both user modeling side by assessing how certain users prefer to travel, and also improve trip recommendations by determining a destination's typical time and duration of visit.

Working Package 1: Conduct a comprehensive literature survey.

3 A FRAMEWORK FOR DATA-DRIVEN RECOMMENDER SYSTEMS

We aim to create a modularized architecture to establish a destination recommender prototype. We want to compare different approaches for one component; therefore, we propose to establish a microservice architecture with strong encapsulation and standardized message passing protocols. This will enable a dynamic interchange of components, e.g., to conduct A/B experiments with different recommendation algorithms and front-ends.

3.1 Data Warehouse

Providing high-quality recommendations requires rich knowledge about the target domain. Currently, information for tourism purposes can be obtained from various online sources. Nonetheless, obtaining relevant features from heterogeneous sources is a data mining challenge [1]. We propose a data warehouse for data-driven recommendations to store and continuously update information about the items to be recommended, i.e., travel destinations. Having historical data enables the detection of trends and improves recommendations, e.g., by determining seasonal fluctuations.

3.2 Region Model

Recommending destinations requires a set of destinations. In an initial approach [8], regions were structured by hand into a region tree where the Earth was the root node followed by continents, sub-continental regions, countries, and states. The size of the query region determines the granularity of the model, resulting in some countries being combined to larger regions, e.g., the Baltic countries or islands in the Caribbean. The advantage of such a model is that, based on the query region, the depth of the region tree can be adapted dynamically to return destinations of a comprehensible size. An ideal solution would realize a trade-off between the features of GeoTree[1], a strict hierarchical data structure of political and administrative regions and the WikiTravel hierarchy[2], which has relaxed consistency for capturing travel-specific features of regions.

3.3 Region Characterization

With a fine-grained region tree, the regions must be enriched with information about relevant factors, such as costs per day, suitability for certain types of traveler, main attractions and potential activities. Our idea is to calculate these metrics based on the presence of venues in the given areas.

Personal Fit. We are currently investigating how collaborative sites, such as Wikipedia and Wikitravel, and online services, such as Google Places[3] and Foursquare, can be used to gather the features of regions. This information must be analyzed, aggregated, and normalized to obtain useful assessments of the individual suitability

[1] http://geotree.geonames.org
[2] https://wikitravel.org/en/Wikitravel:Geographical_hierarchy
[3] https://developers.google.com/places

of a given region relative to specific user preferences. Here, the basic assumption is that the density of certain venues can be used to derive a good measure of personal suitability.

Budget. The daily costs for a typical tourist usually consist of accommodation, food, and activity costs. Curated lists, such as the Price of Travel[4], which offers a *backpacker index*, are of limited help, because their data are sparse and we require up-to-date, and machine-readable information to estimate a budget. However, such sparse lists can be used as ground truth to validate novel approaches. To determine hotel prices, we calculate the median price of a representative set of hotels obtained via the Google Hotel Prices API[5]. Food prices are more difficult to estimate. A relatively accurate source of information could be Numbeo[6], which provides detailed information (e.g., a meal in a restaurant, cappuccino or transportation costs) about the cost of living in 511 cities in 86 countries around the world. The service offers a paid API; however, this API cannot be used to determine food prices for the other 160 countries.

Duration of Stay. In a previous paper [5], we investigated typical tourist travel patterns with a focus on the duration of stay. Although this initial step lacked generality due to sparse data, it revealed where travelers did or did not spend time. We plan to extend this research with more data and refined granularity.

Transportation Costs. Traveling between regions incurs costs. Although these costs are sometimes negligible, e.g., when driving between two neighboring regions, they can be significant depending on the mode of transportation and the distance between regions. The flight costs between regions can be queried using different APIs, such as Skyscanner[7]; the Google Transit API[8] provides information about public transportation.

WP 2: Create a framework for data-driven recommender systems with a data warehouse that aggregates heterogeneous data sources to be used in other parts of the recommender system.

4 USER MODELING

For personalized, content-based recommendations a good item characterization and detailed understanding of user preferences are fundamental. These are commonly elicited and refined through interaction with a user interface. We also plan to investigate how preferences can be extracted from information provided by the user.

In the original approach [8], users manually provided preferences with binary indications about their interest in certain predefined activities. We plan to improve this to continuous intensities per category whose values are derived automatically based on social media [4]. Here, the idea is that a history of travel destinations or information from posts, images, and other interactions in LBSNs, can be used to obtain a detailed profile of the user's personality [28]. This can be complemented with explicit question-answering about the intended type of travel, such as the seven traveler profiles described by Neidhardt et al. [18].

Travel planning is a process; thus, the user should be able to alter the original recommendations. Therefore, a conversational approach [20] with elements of active learning [19] appears to be quite promising. The possibility of critiquing [15] can help fine-tune algorithms such that the system can *"[...] automatically improve with experience [17, p. xv]"*, which is a key part of Mitchell's definition of machine learning. To understand the benefits of different user modeling techniques, we plan to conduct controlled lab experiments using human-centered research methods. The important dependent variables are the effort for users to provide their preferences and the level of detail of the user model.

WP 3: Develop efficient and effective strategies for eliciting user preferences using available data and user interface concepts.

5 CONSTRAINT-BASED RECOMMENDATION

The core of a recommender system is the recommendation engine that ranks items based on a query. However, in the TTDP, the challenge is not just returning the top *n* items but also to selecting a good combination of such items. Framing the initial ranking as a matrix factorization problem would result in extremely sparse data. We propose to obtain rankings based on weighted features of the items; thus, we plan to employ the content-based recommender systems paradigm [12].

Having computed a ranked list of destinations, the actual set of recommended items must be derived in consideration of various factors, such as temporal and monetary constraints, as well as item diversity. If the user preferences are rather specific and the destination model is fine-grained, the top-ranked items may be very similar. In such situations, the algorithm should skip some destinations in favor of improving the variety in the travel knapsack.

Messaoud et al. [16] proposed a variety-seeking model using semantic hierarchical clustering to establish diversity in a set of recommended activities. Similarly, Savir et al. [21] introduced an additional diversity constraints based on attraction types to ensure that the trip's diversity level is above a threshold. Another idea is to adjust item diversity based on the user's personality profile. For example, Wu et al. [28], assessed the user satisfaction based on the diversity level of the recommendations. That study and other approaches [18, 24], use the Five Factor Model of personality [14] to improve recommendations. Tailoring trip diversity to user preferences appears to be a promising avenue to improve the quality of recommendations.

Finally, given a set of regions that fulfills the basic constraints, we propose to fine-tune recommendations by determining how long each of the destinations should be visited, which depends on the typical time required to visit a specific region and of the type of traveler. Lu et al. [13] presented an approach to find an optimal trip in a city within temporal constraints. However, their data model does now allow adaptation of the duration of stay for each attraction. Thus, it is an interesting challenge to determine how aggregated mobility patterns [5] can improve personalized recommendations relative to the duration of stay.

Having implemented the system, we propose to evaluate it online to maximize the number of test subjects. The framework will support A/B testing of independent variables (i.e., novel algorithms

[4]https://www.priceoftravel.com
[5]https://developers.google.com/hotels/hotel-prices/api-reference
[6]https://www.numbeo.com
[7]https://partners.skyscanner.net/affiliates/travel-apis
[8]https://developers.google.com/transit

and baselines) and will collect the dependent variables derived from explicit and implicit feedback.

WP4: Develop content-based recommender algorithms that solve the TTDP without violating user-defined constraints.

6 CONCLUSIONS

In this exposé we have drafted a collection of open research problems whose solutions comprise a global destination recommender system for independent travelers. The main contributions have been organized into four working packages. First, literature in the corresponding fields must be collected, categorized and written up to form a basis for our own contributions. Then, a framework for data-driven recommender systems will be designed to enable continuous data acquisition for item characterization and domain modeling. User preferences will be elicited using novel data-driven user modeling methods, which are to be compared with traditional approaches. Finally, constraint-based recommender algorithms for multi-destination tourist trips will be designed.

These problems will be approached using adequate user-centered research methods, and by developing a research prototype, which will be evaluated in user studies and presented to the scientific community via peer-reviewed publications. We are confident that the data-driven contributions of this thesis can be generalized from the tourism domain and transferred to other complex domains where *Assistive AI* struggles to establish itself.

REFERENCES

[1] Xavier Amatriain and Josep M. Pujol. 2011. Data Mining Methods for Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, Boston, MA, USA, 227–261. https://doi.org/10.1007/978-0-387-85820-3_2

[2] Jie Bao, Yu Zheng, and Mohamed F. Mokbel. 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12)*. ACM, New York, NY, USA, 199–208. https://doi.org/10.1145/2424321.2424348

[3] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. 2015. Recommendations in location-based social networks: a survey. *GeoInformatica* 19, 3 (Feb. 2015), 525–565. https://doi.org/10.1007/s10707-014-0220-8 arXiv:1502.07526

[4] Philip Bonhard and M. Angela Sasse. 2006. 'Knowing me, knowing you' — Using profiles and social networking to improve recommender systems. *BT Technology Journal* 24, 3 (July 2006), 84–98. https://doi.org/10.1007/s10550-006-0080-3

[5] Linus W. Dietz, Daniel Herzog, and Wolfgang Wörndl. 2018. Deriving Tourist Mobility Patterns from Check-in Data. In *Proceedings of the WSDM 2018 Workshop on Learning from User Interactions*. Los Angeles, CA, USA.

[6] Zachary Friggstad, Sreenivas Gollapudi, Kostas Kollias, Tamas Sarlos, Chaitanya Swamy, and Andrew Tomkins. 2018. Orienteering Algorithms for Generating Travel Itineraries. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*. ACM, New York, NY, USA, 180–188. https://doi.org/10.1145/3159652.3159697

[7] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. A survey on algorithmic approaches for solving tourist trip design problems. *Journal of Heuristics* 20, 3 (June 2014), 291–328. https://doi.org/10.1007/s10732-014-9242-5

[8] Daniel Herzog and Wolfgang Wörndl. 2014. A Travel Recommender System for Combining Multiple Travel Regions to a Composite Trip. In *CBRecSys@RecSys (CEUR Workshop Proceedings)*, Vol. 1245. CEUR-WS.org, Foster City, Silicon Valley, California, USA, 42–48.

[9] Hsun-Ping Hsieh, Cheng-Te Li, and Shou-De Lin. 2012. Exploiting Large-scale Check-in Data to Recommend Time-sensitive Routes. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing (UrbComp '12)*. ACM, New York, NY, USA, 55–62. https://doi.org/10.1145/2346496.2346506

[10] Qi Liu, Enhong Chen, Hui Xiong, Yong Ge, Zhongmou Li, and Xiang Wu. 2014. A Cocktail Approach for Travel Package Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 26, 2 (Feb. 2014), 278–293. https://doi.org/10.1109/tkde.2012.233

[11] Qi Liu, Yong Ge, Zhongmou Li, Enhong Chen, and Hui Xiong. 2011. Personalized Travel Package Recommendation. In *Proceedings of the IEEE 11th International Conference on Data Mining (ICDM '11)*. IEEE, Vancouver, BC, Canada, 407–416. https://doi.org/10.1109/icdm.2011.118

[12] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Recommender Systems Handbook. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, Boston, MA, USA, Chapter Content-based Recommender Systems: State of the Art and Trends, 73–105. https://doi.org/10.1007/978-0-387-85820-3_3

[13] Eric Hsueh-Chan Lu, Chih-Yuan Lin, and Vincent S. Tseng. 2011. Trip-Mine: An Efficient Trip Planning Approach with Travel Time Constraints. In *2011 IEEE 12th International Conference on Mobile Data Management*, Vol. 1. IEEE, Lulea, Sweden, 152–161. https://doi.org/10.1109/mdm.2011.13

[14] Robert R. McCrae and Oliver P. John. 1992. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality* 60, 2 (June 1992), 175–215. Issue 2. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

[15] Lorraine McGinty and James Reilly. 2011. On the Evolution of Critiquing Recommenders. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, Boston, MA, USA, 419–453. https://doi.org/10.1007/978-0-387-85820-3_13

[16] Montassar Ben Messaoud, Ilyes Jenhani, Eya Garci, and Toon De Pessemier. 2017. SemCoTrip: A Variety-Seeking Model for Recommending Travel Activities in a Composite Trip. In *Advances in Artificial Intelligence: From Theory to Practice*. Springer International Publishing, Arras, France, 345–355. https://doi.org/10.1007/978-3-319-60042-0_40

[17] Thomas M. Mitchell. 1997. *Machine Learning* (1 ed.). McGraw-Hill, Inc., New York, NY, USA. https://doi.org/10.1002/0470018860.s00036

[18] Julia Neidhardt, Leonhard Seyfang, Rainer Schuster, and Hannes Werthner. 2015. A picture-based approach to recommender systems. *Information Technology & Tourism* 15, 1 (March 2015), 49–69. https://doi.org/10.1007/s40558-014-0017-5

[19] Neil Rubens, Dain Kaplan, and Masashi Sugiyama. 2015. Active Learning in Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, USA, 809–846. https://doi.org/10.1007/978-1-4899-7637-6_24

[20] Maria Salamó, James Reilly, Lorraine McGinty, and Barry Smyth. 2005. Knowledge Discovery from User Preferences in Conversational Recommendation. In *Knowledge Discovery in Databases*. Springer Berlin Heidelberg, Berlin, Heidelberg, 228–239. https://doi.org/10.1007/11564126_25

[21] Amihai Savir, Ronen Brafman, and Guy Shani. 2013. Recommending Improved Configurations for Complex Objects with an Application in Travel Planning. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13)*. ACM, New York, NY, USA, 391–394. https://doi.org/10.1145/2507157.2507196

[22] Steffen Stabb, Hannes Werther, Francesco. Ricci, Alexander Zipf, Ulrike Gretzel, Daniel R. Fesenmaier, Cécile Paris, and Craig Knoblock. 2002. Intelligent Systems for Tourism. *IEEE Intelligent Systems* 17 (Nov. 2002), 53–64. https://doi.org/10.1109/mis.2002.1134362

[23] Chang Tan, Qi Liu, Enhong Chen, Hui Xiong, and Xiang Wu. 2014. Object-Oriented Travel Package Recommendation. *ACM Transactions on Intelligent Systems and Technology* 5, 3 (Sept. 2014), 1–26. https://doi.org/10.1145/2542665

[24] Marko Tkalcic, Matevž Kunaver, Jurij Tasic, and Andrej Kosir. 2009. Personality Based User Similarity Measure for a Collaborative Recommender System. In *5th Workshop on Emotion in Human-Computer Interaction-Real World Challenges*, Christian Peter, Elizabeth Crane, Lesley Axelrod, Harry Agius, Shazia Afzal, and Madeline Balaam (Eds.). Fraunhofer, 30–37.

[25] UNESCO. 2003. Convention for the Safeguarding of the Intangible Cultural Heritage. https://ich.unesco.org/en/convention. (Oct. 2003).

[26] Pieter Vansteenwegen, Wouter Souffriau, and Dirk Van Oudheusden. 2011. The orienteering problem: A survey. *European Journal of Operational Research* 209, 1 (Feb. 2011), 1–10. https://doi.org/10.1016/j.ejor.2010.03.045

[27] Pieter Vansteenwegen and Dirk Van Oudheusden. 2007. The Mobile Tourist Guide: An OR Opportunity. *OR Insight* 20, 3 (July 2007), 21–27. https://doi.org/10.1057/ori.2007.17

[28] Wen Wu, Li Chen, and Liang He. 2013. Using Personality to Adjust Diversity in Recommender Systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HT '13)*. ACM, New York, NY, USA, 225–229. https://doi.org/10.1145/2481492.2481521

[29] Wolfgang Wörndl, Alexander Hefele, and Daniel Herzog. 2017. Recommending a sequence of interesting places for tourist trips. *Information Technology & Tourism* 17, 1 (Feb. 2017), 31–54. https://doi.org/10.1007/s40558-017-0076-5

[30] Yu Zheng and Xing Xie. 2011. Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology* 2, 1, Article 2 (Jan. 2011), 29 pages. https://doi.org/10.1145/1889681.1889683