SemCoTrip: A variety-seeking model for recommending travel activities in a composite trip

Keywords: Multi-destination trips, leisure activities, diversity, hierarchical clustering, ontology

Abstract: Selecting appropriate activities, especially in multi-destinations trips, is a hard task that many travellers face each time they want to plan for a trip. With the budget and time limitations, travellers will try to select activities that best fit their personal interests. Most of existing travel recommender systems don't focus on activities that a traveller might be interested in. In this paper, we go beyond the specific problem of combining regions in a composite trip to propose a variety-seeking model which is capable of providing travelllers with recommendations on what activities they can engage in when visiting different regions. A semantical hierarchical clustering-based model is proposed to guarantee diversity within the set of recommended activities. Experimental results on a real dataset have shown that the proposed approach helps the traveller to avoid doing the same or similar activities in a composite trip, thus, promoting less popular activities to be selected.

1 INTRODUCTION

Recommendation systems have made a significant difference in people's lives. Being one of the early adoption areas, the tourism industry has taken advantage of the recent advances in recommender systems (RS) to enhance the quality of services offered to travellers and to enrich their travel experiences (Lim et al., 2015; Chen et al., 2016; Lu et al., 2016). One of the potential applications of RS in tourism that has not yet been explored in details is the recommendation of composite trips. Most of the existing approaches dealing with multi-destination trips focus on developing ways to combine single travel items like regions and routes in order to maximize the benefit for the traveller (Maruyama et al., 2004; Cheng et al., 2013; Herzog and Wörndl, 2014). Nevertheless, none of these works has directly tackled the issue of managing activities during the stay at each destination.

This paper provides a substantially extended version of a previous work (Herzog and Wörndl, 2014), in which authors proposed an efficient algorithm for the recommendation of composite trips. Further attempts in this direction will be initiated in order to design complementary strategies to utilize semantic prior knowledge, to improve the diversity of the recommended activities and, most importantly, to enrich visitors' travel experience. The remainder of this paper is organized as follows: Section 2 discusses some related works where the diversity level is considered as a trip constraint. Section 3 gives the necessary background for both hierarchical clustering and ontologies. We then describe in details the SemCoTrip strategy in Section 4. In Section 5, we report experimental results that show the effectiveness of the proposed algorithm. Concluding remarks and future works will be given in Section 6.

2 STATE OF THE ART

In recent years, there has been a continuous line of research focusing on diversifying the recommended lists of activities and destinations to meet tourists' satisfaction. Diversity is commonly defined as the average pairwise distance between recommendations to users (Castells et al., 2015). Authors in (Moreno et al., 2013) proposed to use the K-means algorithm to assign users to clusters that have similar characteristics. When executing their clustering procedure, activities were weighted to ensure that their SigTur recommender system provides diverse recommendations. In (Savir et al., 2013), the diversity level was considered as a trip constraint. To ensure diversity, the authors used a measure of balance between the attractions' categories and the acceptable rating threshold. The work in (Ruotsalo et al., 2013) presents a search result clustering algorithm based on semantic data representation which chooses a set of objects from each cluster to increase the diversity of the proposal made to the visitor of a museum. (Sanchez-Vilas et al., 2015) came up with a surprising result stating that the global error of k-Nearest Neighbours-based recommender systems decreases when a higher diversity is associated to the recommendations.

3 BACKGROUND

In this section, we provide basic concepts related to semantic knowledge-based systems and hierarchical clustering which are essential in understanding the rest of the paper.

3.1 Knowledge-based systems

Knowledge-based systems (KBs) provide domain reasoning frameworks combined with inference engines that usually reason over logical languages. Ontology, which is one of those popular semantic driven knowledge based systems, has received numerous definitions in the literature. The most commonly cited definition was given in (Gruber, 1993). It defines the ontology as an explicit specification of a conceptualization. The "conceptualization", refers to a simplified view of the world by identifying its relevant concepts.The word "explicit" means that all concepts (resp. their specific properties and constraints) must be explicitly defined.

Definition 3.1. *An ontology can be formally expressed as:*

- *A set of concepts* $\mathcal{C}=\{C_1,...,C_n\}$*, which are mainly interrelated by means of taxonomic (is-a) relations in the form of a hierarchy* $\mathcal{H}$*,*
- *A set of properties for each concept,*
- *Semantic (i.e. non-taxonomic) relations between concepts,*
- *A set of instances I (i.e. occurrences of concepts and semantic relations), and*
- *A set of assertions and formal axioms (i.e. constraint-relationships like should, should not, must, must not, etc).*

3.2 Hierarchical clustering

Clustering is a typical unsupervised learning task which aims at grouping together similar objects (with respect to their attribute values) into subsets called *clusters*. A cluster is therefore a collection of objects which are similar to each others and dissimilar to objects belonging to other clusters.

We can distinguish four main categories of clustering methods: (1) *Centroid-based clustering* such as K-means (MacQueen, 1966), (2) *Hierarchical clustering* (Jain and Dubes., 1988) such as single-linkage and complete-linkage clustering methods, (3) *Distribution-based clustering* such as Expectation-Maximization (EM) algorithm (Dempster et al., 1977) and (4) *Density-based clustering* such as DBSCAN algorithm (Kriegel et al., 2011).

Among these categories, we are interested in the hierarchical clustering one which could be either agglomerative or divisive. *Agglomerative methods* are "bottom up" approaches which start by assigning each element to a separate cluster then a merging of the two least distant (most similar) clusters is successively performed leading to larger clusters. However, *Divisive methods* are "top down" approaches in which all objects start in one cluster, and splits are performed recursively as one moves down. In practice, agglomerative techniques were more commonly used.

Distance (or similarity) between two clusters is determined by a *linkage criterion*, which is a function of the pairwise distances between instances one from each cluster. Most popular linkage criteria are: (1) *Single-linkage*: the distance between two clusters is the *minimum* pairwise distance between elements, one from each cluster (i.e. the shortest link between clusters). (2) *Complete-linkage*: the distance between two clusters is the *maximum* pairwise distance between elements, one from each cluster (i.e. the longest link between clusters). (3) *Average-linkage*: the distance between two clusters is the *average* pairwise distance between elements, one from each cluster. Other linkage criteria exist such as the Average group linkage (the sum of all intra-cluster variance), Ward's linkage (the increase in variance for the cluster being merged), V-linkage (the probability that candidate clusters spawn from the same distribution function). A good survey on hierarchical clustering algorithms could be found in (Murtagh and Contreras, 2012).

The agglomerative clustering continues until a stopping criterion is met. We can apply a *distance-based stopping criterion* to stop clustering when the clusters are too far apart to be merged (i.e., distance between the closest clusters to be merged is greater to a user-predefined or computed threshold). A *number of clusters-based criterion* can also be used to stop clustering when there is a sufficiently predefined small number of clusters.

Hierarchical clustering has been mainly used in conjunction with recommender systems to deal with the problem of scalability. In fact, incremental hierarchical agglomerative clustering has been used in (Haruechaiyasak et al., 2005) to handle the large number of user profiles in e-commerce recommender systems. Moreover, in order to better personalize navigational recommendations in social tagging systems, authors in (Shepitsen et al., 2008) applied hierarchical clustering to cluster the wide variety of tags. In (Zheng et al., 2013), an ensemble hierarchical clustering approach has been applied to group users with similar reading profiles and get news hierarchies

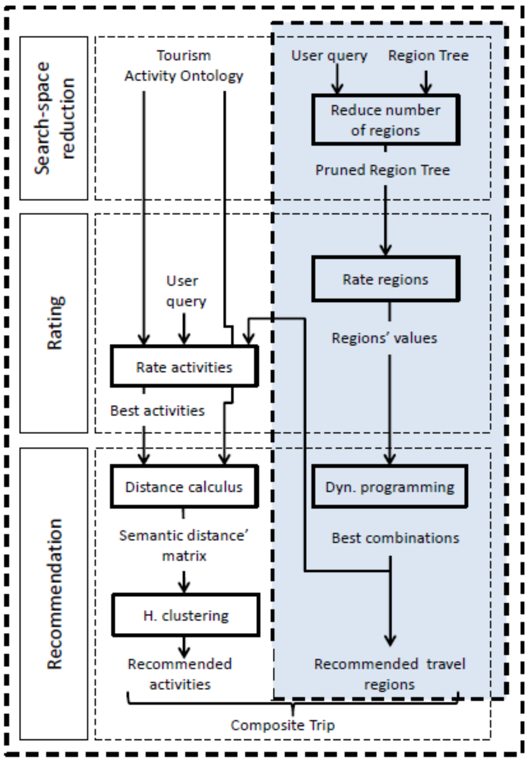

Figure 1: SemCoTrip: Extending the composite trips' RS of (Herzog and Wörndl, 2014) to consider a variety of leisure activities.

which are then used in recommending news articles. More recently, hierarchical clustering has been used in (West et al., 2016) to enhance the relevance of papers to recommend for researchers among a huge number of published papers.

4 SEMCOTRIP: A SEMANTICAL ALGORITHM FOR THE RECOMMENDATION OF COMPOSITE TRIPS

The general overview of the SemCoTrip (Semantical Composite Trip) algorithm is depicted in Figure 1. SemCoTrip inputs are: a travel region dataset and a tourism-activity ontology. We follow approximately the same methodology proposed by authors in (Herzog and Wörndl, 2014) to reproduce the same performance when combining regions and determining the optimal duration of stay per region. A blue shaded area was added in Figure 1 to highlight the differences between the two algorithms and to visualize the recommendation process proposed by (Herzog and Wörndl, 2014).

4.1 Search-space reduction

First, we start by reducing the search space by excluding irrelevant regions to the user query. Using the region tree hierarchy, if a region is removed, all its sub-regions and related activities will be removed as well.

4.2 Rating

The remaining travel activities of the pruned region tree will be then rated. At this level, (Herzog and Wörndl, 2014) used a 5-point Likert scale to rate regions' features depending on the month (season), which could potentially exclude many relevant destinations from the recommendations returned to users. Alternatively, we will simply assume that activities offered in each region are subject to change from season to season. Concepts' attributes in the input ontology will indicate how well the tourism activities match each traveling type group. By doing so, travel region ratings in our scenario will thus depend on their corresponding activities ratings.

Here, the standard rating schema of (Herzog and Wörndl, 2014) that involves user and region dimensions is extended to three-dimensional schema involving activities (Refer to figure 2). Such multidimensional approach is usually used to deal with context in RSs (Adomavicius et al., 2005). For this case, we will further define a rating function R on the recommendation space User$\times$Region$\times$Activity specifying how much user u $\in$ User liked activity a $\in$ Activity in (sub-)region s $\in$ Region, R(u,a,s).

At the end of this step, regions with low ratings will be removed and the remaining ones will be combined in a way to maximize their values for the user while still respecting the budget and the duration constraints.

4.3 Recommendation strategy

The problem, as defined at that level, can still be considered as a variant of the knapsack problem (Burg et al., 1999) which can be efficiently solved by means of dynamic programming (Kellerer et al., 2004). Two objectives are considered here: (1) The value of the composite trip is proportional to the distance between regions and (2) the best combination of regions is obtained based on the optimal duration of the stay per region. The application of the Dynamic programming

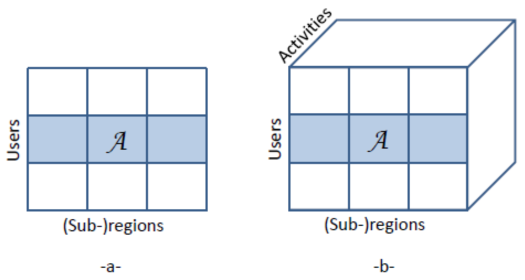

Figure 2: (a) 2-D rating matrix as proposed in (Herzog and Wörndl, 2014) and (b) SemCoTrip multidimensional model for the User×Region×Activity recommendation space.

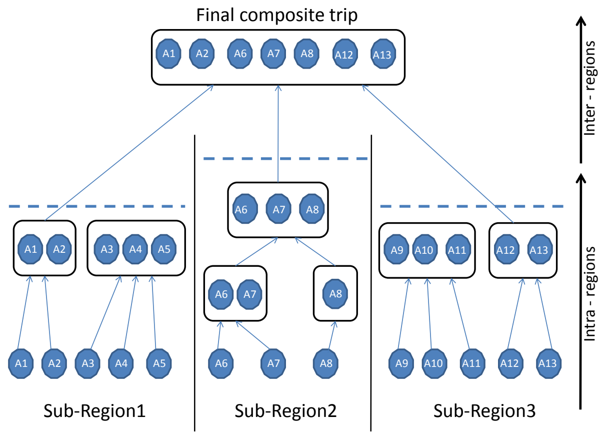

Figure 3: Two-levels hierarchical clustering to optimize the diversity of the activities lists.

approach to our dataset provided a candidate solution which consists of a subset of regions along with the duration time to spend in each of these regions.

The complementary component that we propose in our approach is to select the set of activities to recommend based on the recommended regions. An important criterion that we introduce in selecting the activities is diversity: we want our system to recommend activities which are as dissimilar as possible. To ensure that diversity, we will first use a semantic hierarchical clustering approach which will try to select heterogeneous clusters of activities. Then, a selection algorithm will be applied to find the optimal combination of clusters of activities found in the previous step. Throughout the hierarchical clustering step, we will consider Rada's distance (Rada et al., 1989) as the specific distance for calculating the semantic gain intra-(resp. inter) clusters.

Definition 4.1. *Let C_i and C_j be two concepts in an ontology restricted to taxonomic hierarchy. A measure of the conceptual Rada's distance is expressed as the minimum number of links separating the two concepts.*

We choose this distance because of its simplicity and its broad adoption. Note that Rada's distance can be replaced by any other semantic distance (refer to (Blanchard et al., 2005) for a comparative analysis between semantic distances).

As a first step of our approach, the clustering is performed on the total set of activities of each (sub-) region separately. As shown in Figure 3, for each recommended (sub-)region, the hierarchical clustering will result in a set of one or more clusters. Activities within each cluster are selected based on the maximization of the semantic distance between the activities. The intra-regions activities clustering algorithm is as described in the following:

Algorithm 1 Intra-regions activities clustering algorithm

Require: $A = A_1, .., A_n$: list of all activities related to a given (sub-)region, $Cost$: total cost allocated to the (sub-)region.
Ensure: One or more clusters of dissimilar activities.
1. Assign each activity A_i in A to a separate cluster C_i
2. Calculate the semantic distances between each two clusters.
3. Group together the two closest clusters.
4. Update the cost of that cluster.
repeat
Steps 2. and 3.
until one stopping criteria is met.

The second step of the approach is based on an inter-regions selection algorithm that be used to find out the optimal combination of clusters of activities (no more than one cluster will be selected from each (sub-)region). Clusters that maximize an average semantic distance between clusters of all remaining (sub-)regions will be selected and merged to form the final group of activities. The selection algorithm is described in Algorithm 2.

To illustrate the SemCoTrip outcome, we will use the same example query [1] as in (Herzog and Wörndl, 2014). We adopt almost the same parameters, i.e.,

[1]The estimated cost covers the minimum time and level that we find acceptable to perform each activity.

Algorithm 2 Inter-regions activities selection algorithm

Require: $S = S_1, .., S_k$ where each $S_i = C_1, .., C_n$ is a set of one or more clusters of activities for a sub-region i, k is the total number of sub-regions.
Ensure: F: a final set of activities.
for each S_i in S **do**
1. Calculate the average semantic distance between each cluster C_i in S_i and all clusters in S_j ($\forall j \neq i$)
2. Select the cluster C_i with the highest average distance and add it to F.
end for

a budget of 2000 euro, a maximum time of travelling of eight weeks and a lowest possible crime rate. We also exclude Europe, Asia and their corresponding sub-regions from the search space.

The only modification that we have made is the identification of the user as rejuvenator, in addition to cultural explorer. This will enable SemCoTrip to produce even more diverse activities recommendations without modifying the final output, i.e., sub-regions recommendation order. Here, the best recommendation is composed of three different sub-regions with total costs of 1890 euro.

Then, to recommend the corresponding activities, we declared the user as a first-time tourist. In our dataset, all those recommended activities are declared with at least 3 points on the Likert scales of the two considered tourist' types. As you might have noticed, for this 8 weeks trip, SemCoTrip have embraced three alternative forms of tourism (i.e. eco, food and agro-rural tourism), in addition to proposing the most famous tourist attractions (i.e. mass tourism) in these regions. For future work, we suggest extending the model to consider the sequence in which the activities need to occur (Ibáñez et al., 2016), as we have done for sub-regions.

4.4 Complexity analysis

The SemCoTrip algorithm basically consists of four procedures that take place sequentially, thus the total complexity will be the sum of their respective complexities. The dynamic programming procedure is $O(n_r.b.d)$ where n_r is the number of sub-regions, b is the budget and d is the maximum duration of stay. We suppose that the activities' hierarchy has n_c concepts. For rada distance' calculus, we will need to compute the upper (or lower) triangular matrix. Obviously, the complexity of finding the distance between each single concept pair is O(2). So the complexity of the overall task will be:

$$O(2.1 + ... + 2.(n_c - 1)) = O(\frac{2n_c(n_c - 1)}{2}) = O(n_c^2 - n_c)$$

The time complexity of the intra-regions clustering is $O(n_r^*.n_a^2.logn_a)$, where n_a is the number of activities and n_r^* is the number of sub-regions returned by dynamic programming. The inter-regions clustering procedure can be done in $O(n_r^*.n_l^2)$ in the worst case, where n_l is the total number of clusters in the considered sub-regions.

Table 1: Recommended travel activities for the same example query of (Herzog and Wörndl, 2014).

Regions	Dur.	Cost	Recommended activities
Argentina & Paraguay	4W	790 €	Devil's throat, Plaza 9 Julio, Cachi' wineries, Parque Nacional Los Cardones, Hill of Seven Colors, Cathedral of Salta, Casa Historica de Tucuman, Train to the Clouds, Casa de Gobierno de Jujuy, Loma Bola Vuelo y Aventura Museo Pajcha Arte Etnico Iguazu Falls, Pearl of the south, Tea plantations tour, Reserva Mbatovi, Mercado municipal n4, Wilson Tower skyscraper.
Bolivia	2W	530 €	Yungas world's most dangerous road, Inka trails, Tiwanaku, Solar de Uyuni, Mountain climbing.
Peru	2W	570 €	Sandboarding in Huacachina, ChanChan archeological area, Sacred Valley of the Incas, Trekking Cordillera Blanca Nazca lines flight, Surfing in Mancora, Temple of the Sun Museo Tumbas Reales
SUM	8W	1890 €	

5 EXPERIMENTAL STUDY

5.1 Data Description

5.1.1 Dataset

Our variety-seeking model was tested on an extended version of the dataset used in (Herzog and Wörndl, 2014). The dataset (a region tree-like structure) is composed of a total of 152 regions with 124 leaves. The main difference with the original dataset is that we have assigned a range of seasonal activities to each (sub-)region. These activities are then mapped to their corresponding concepts in the used ontology and a 5-point Likert scale was used to indicate how well the proposed activity matches travellers types that we have categorized in four families according to the Canadian Tourism Commission [2] (See table 2). All other input data (e.g. duration, budget, routing, crime level, etc) is kept the same.

Table 2: Traveler's types classification.

Category	Traveler's type
Learners	Cultural explorer
	Authentic Experiencers
	Cultural History Buffs
	Personal History Explorers
Indulgers	Free Spirits
Familiarity Seekers	Gentle Explorers
	No Hassle Travellers
	Virtual Travellers
Escapists	Rejuvenators

5.1.2 Ontology

The approach that we are proposing builds up on the use of a tourism activities ontology, which is presented as a hierarchy composed of a set of more than 200 concepts taxonomically related by subsumptions. The concepts are formalized into three related (sub-)ontologies, referred to as sport attractions, natural attractions and cultural attractions. Figure 4 shows a small excerpt of our tourism ontology.

5.2 Experimental design

A Java-based prototype was implemented in order to evaluate our approach. We used the prototype to handle a sample of 100 users queries. For each query, we changed input parameters (e.g. traveller's type, budget, total duration, etc.). The prototype executes each query separately and provides the top-rated recommendations; based on the recommendation procedures described in Section IV.

[2]http://en.destinationcanada.com/resources-industry/explorer-quotient

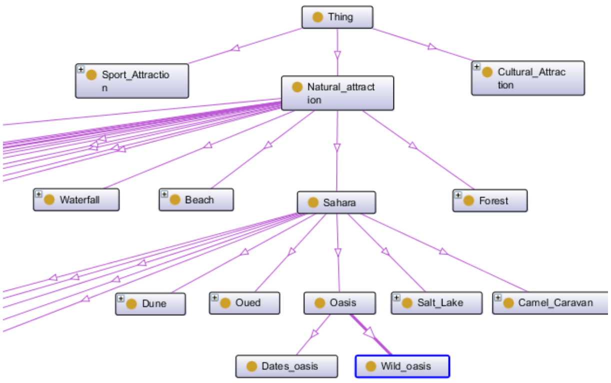

Figure 4: Excerpt of the tourism activities ontology.

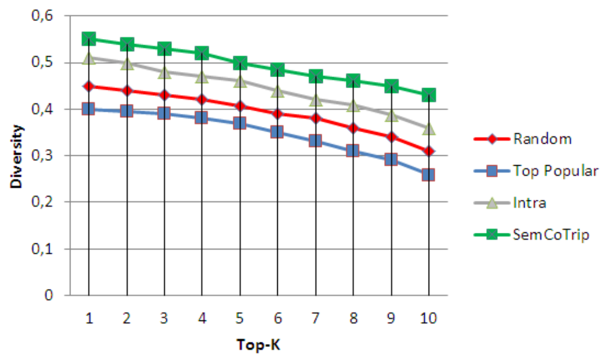

Figure 5: Diversity for the top-K recommendations.

5.3 Results & Interpretations

The main objective of our approach is to ensure diversity when recommending a set of activities. The diversity degree of a set $A = \{A_1, \ldots, A_n\}$ of n recommended activities is measured by:

$$Diversity = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} Rada(A_i, A_j)}{\frac{n*(n-1)}{2}} \quad (1)$$

where $Rada(A_i, A_j)$ is the normalized Rada's distance between two activities A_i and A_j which lies on the unit interval.

Figure 5 shows, for the top-10 recommendations, the diversity level within the set of activities recommended by SemCoTrip as well as three other concurrent algorithms.

The grey line refers to a variant of SemCoTrip which is only using the Intra-regions activities clustering algorithm. The red (resp. blue) curve shows the diversity relative to the baseline method that recommends random (resp. most popular) activities. We stress the fact that all four algorithms are based on the same basic regions recommendation idea proposed in

(Herzog and Wörndl, 2014). Only activities' recommendation strategies have been changed.

For all four algorithms, a list of top-k recommendations is kept and sorted in descending order of the diversity within the activities. As we can observe from the figure 5, the popularity-based strategy gave the worst results, as it drastically reduces the activities' search space for each sub-region.

Surprisingly, the random approach produced competitive results when compared to the variant "Intra". This could be explained by the fact that many sub-regions offer seasonal activities for several periods in the year. Those activities that depend on the weather are, usually, susceptible to be joined together on the corresponding ontology. This is the particular case where the random strategy provides nearly identical results to those provided by SemCoTrip.

The difference in performance between SemCoTrip and its variant "Intra" justifies that the whole clustering process is required to illustrate diversity in recommendations.

6 CONCLUSION

In this paper, we have proposed the SemCoTrip recommender system which seeks to recommend a set of diverse activities for a composite trip. A semantic-based hierarchical clustering approach has been used along with a tourism ontology to ensure diversity. Experimental results on a real dataset have shown that activities recommended by SemCoTrip are better, in terms of diversity, than activities recommended by a variant of SemCoTrip and two baseline approaches.

For future works, we intend to manage the sequence in which the activities need to occur and propose a complementary approach to improve novelty and serendipity whilst maintaining high accuracy of recommendations.

REFERENCES

Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1):103–145.

Blanchard, E., Harzallah, M., Briand, H., and Kuntz, P. (2005). A typology of ontology-based semantic measures. In Missikoff, M. and Nicola, A. D., editors, *EMOI-INTEROP*, volume 160 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Burg, J. J., Ainsworth, J., Casto, B., and Lang, S.-D. (1999). Experiments with the oregon trail knapsack problem. *Electronic Notes in Discrete Mathematics*, 1(B):26–35.

Castells, P., Hurley, N. J., and Vargas, S. (2015). Novelty and diversity in recommender systems. In Ricci, F., Rokach, L., and Shapira, B., editors, *Recommender Systems Handbook*, pages 881–918. Springer.

Chen, D., Ong, C. S., and Xie, L. (2016). Learning points and routes to recommend trajectories. *CoRR*, abs/1608.07051.

Cheng, S.-T., Chen, Y.-J., Horng, G.-J., and Wang, C.-H. (2013). Using cellular automata to reduce congestion for tourist navigation systems in mobile environments. *Wireless Personal Communications*, 73(3):441–461.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistcal Society*, 39(B):1–38.

Gruber, T. R. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N. and Poli, R., editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands.

Haruechaiyasak, C., Tipnoe, C., Kongyoung, S., Damrongrat, C., and Angkawattanawit, N. (2005). A dynamic framework for maintaining customer profiles in e-commerce recommender systems. In *EEE*, pages 768–771. IEEE Computer Society.

Herzog, D. and Wörndl, W. (2014). A travel recommender system for combining multiple travel regions to a composite trip. In Bogers, T., Koolen, M., and Cantador, I., editors, *CBRecSys@RecSys*, volume 1245 of *CEUR Workshop Proceedings*, pages 42–48. CEUR-WS.org.

Ibáñez, J., Sebastia, L., and Onaindia, E. (2016). Planning tourist agendas for different travel styles. In Kaminka, G. A., Fox, M., Bouquet, P., Hullermeier, E., Dignum, V., Dignum, F., and van Harmelen, F., editors, *ECAI*, volume 285 of *Frontiers in Artificial Intelligence and Applications*, pages 1818–1823. IOS Press.

Jain, A. K. and Dubes., R. (1988). *Algorithms for Clustering Data*. Prentice Hall.

Kellerer, H., Pferschy, U., and Pisinger, D. (2004). *Knapsack Problems*. Springer, Berlin, Germany.

Kriegel, H.-P., Kröger, P., Sander, J., and Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):231–240.

Lim, K. H., Chan, J., Leckie, C., and Karunasekera, S. (2015). Personalized tour recommendation based on user interests and points of interest visit durations. In Yang, Q. and Wooldridge, M., editors, *IJCAI*, pages 1778–1784. AAAI Press.

Lu, C., Laublet, P., and Stankovic, M. (2016). Travel attractions recommendation with knowledge graphs. In Blomqvist, E., Ciancarini, P., Poggi, F., and Vitali, F., editors, *EKAW*, volume 10024 of *Lecture Notes in Computer Science*, pages 416–431.

MacQueen, J. (1966). Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability*.

Maruyama, A., Shibata, N., Murata, Y., Yasumoto, K., and Ito, M. (2004). A personal tourism navigation system to support traveling multiple destinations with time restrictions. In *AINA (2)*, pages 18–22. IEEE Computer Society.

Moreno, A., Valls, A., Isern, D., Marin, L., and Borràs, J. (2013). Sigtur/e-destination: Ontology-based personalized recommendation of tourism and leisure activities. *Eng. Appl. of AI*, 26(1):633–651.

Murtagh, F. and Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery*, 2(1):86–97.

Rada, R., Mili, H., Bicknell, E., and Blettner, M. (1989). Development and application of a metric on semantic nets. In *IEEE Transactions on Systems, Man and Cybernetics*, pages 17–30.

Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T., and Hyvönen, E. (2013). Smartmuseum: A mobile recommender system for the web of data. *J. Web Sem.*, 20:50–67.

Sanchez-Vilas, F., Ismailov, J., and Sanchez, E. (2015). The importance of diversity in profile-based recommendations: A case study in tourism. In *TouRS, Workshop on Tourism Recommender Systems, 9th ACM Conf. on Recommender Systems*, pages 43–50. Vienna, Austria.

Savir, A., Brafman, R. I., and Shani, G. (2013). Recommending improved configurations for complex objects with an application in travel planning. In Yang, Q., King, I., Li, Q., Pu, P., and Karypis, G., editors, *RecSys*, pages 391–394. ACM.

Shepitsen, A., Gemmell, J., Mobasher, B., and Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In Pu, P., Bridge, D. G., Mobasher, B., and Ricci, F., editors, *RecSys*, pages 259–266. ACM.

West, J. D., Wesley-Smith, I., and Bergstrom, C. T. (2016). A recommendation system based on hierarchical clustering of an article-level citation network. *IEEE Trans. Big Data*, 2(2):113–123.

Zheng, L., Li, L., Hong, W., and Li, T. (2013). Penetrate: Personalized news recommendation using ensemble hierarchical clustering. *Expert Syst. Appl.*, 40(6):2127–2136.