

Multi-Criteria Journey Aware Housing Recommender System

Elizabeth M. Daly, Adi Botea, Akihiro Kishimoto, Radu Marinescu
Smart Cities Technology Centre
IBM Research - Ireland
{dalyeliz, adibotea, akihirok, marinrad}@ie.ibm.com

ABSTRACT

Recommender systems can be employed to assist users in complex decision making processes. This paper presents a multi-criteria housing recommender system which takes into account not just features of a home, such as rent, but also the transportation links to user specified locations. First, we describe an efficient multi-hop journey time calculator. Second, we introduce a mechanism to find the optimal solutions for multi-criteria evaluation, where a balanced trade-off between the target goals is found. Finally, we present a user study to demonstrate the potential of such a system.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Algorithms, Decision-Support

Keywords

Recommender Systems; Multi-criteria; Public Transportation

1. INTRODUCTION

Recommender systems may assist users in the decision-making process and have been employed in a wide variety of domains from product recommendations [9], to music and movies, to social connections [2] and city events [7]. More recently, recommender systems have been applied to more complex tasks such as trip planning [8] and journey planning [10] in the area of tourism.

Recommender systems can be extremely valuable in domains which require a great deal of computation in order to evaluate candidate solutions and assist a user in decision making in the face of an abundance of options. Housing selection presents an interesting challenge, due to a combination of two reasons. First off, there can be many candidate houses to consider. Secondly, evaluating one house needs to consider multiple factors. Besides traditional factors, such as cost, number of bedrooms, and housing type, one needs to take into account where they travel to and from (i.e. their

geographic requirements), when they travel (e.g. rush-hour or day time) and how frequently they intend to travel. These many dimensions may make it difficult for users to reason about, when evaluating a home in order to find the optimum location depending on their individual needs.

Journey planning and route recommendations have become increasingly popular in the tourism domain. Tumas and Ricci present a personalised route recommendation system using a combination of walking and public transportation to support tourists taking into account user preferences [10]. Garcia et. al. integrate public transportation and real-time scheduling and in order to design tourist trails [3]. Lathia and Capra leverage user travel patterns to recommend public transportation tickets to minimise the cost to the passenger [5].

The deployment of recommender systems for the purpose of home finding is less studied due to the increased complexity and the higher risk associated with such decisions [4]. However, recent work has begun to explore this application domain. Wang et. al. propose a collaborative filtering approach to recommending housing using a person's attributes such as gender, age and profession [11]. Yang et. al. developed a real-estate recommendation system employing case-based reasoning to match user wish lists to homes on the basis of location, attributes such as number of bedrooms and bathrooms and property price [12]. Our work focuses on evaluating the utility of providing users with transportation related information when selecting a home and presenting options that optimize their choice across multiple criteria.

2. HOUSING RECOMMENDER

Our JOURA (Journey-Aware) Housing Recommender System combines multi-modal journey planning with a multi-objective constraint optimization engine in order to recommend available homes. The map interface allows users to view available homes based on their geographic location. Users can input a number of target locations that they wish to be accessible through public transportation from their home. For instance, the work place can be one of the selected target locations. Once users have selected their target locations, the journey planner is queried to calculate the travel times from each target location to each home over the course of the day. For the purposes of this paper the chosen times of day are 8am, 12pm, 4pm and 8pm.

2.1 System Overview

The JOURA Housing Recommender System consists of three main components:

- **User Profile.** Users specify target locations that they are interested in travelling to on a regular basis and therefore would like to be considered when recommending a home.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

RecSys'14, October 6–10, 2014, Foster City, Silicon Valley, CA, USA.

Copyright 2014 ACM 978-1-4503-2668-1/14/10 ...\$15.00.

<http://dx.doi.org/10.1145/2645710.2645764>.

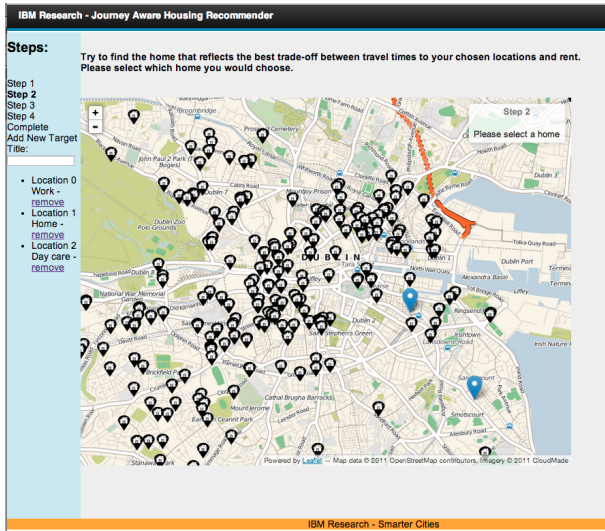


Figure 1: JOURA Recommender System

- **Journey Planner.** A journey planning system is used to compute travel times between relevant pairs of locations, such as a candidate house and a work place. The input required to the planning system includes knowledge about the public transport network, called a *network snapshot*, and a *journey plan request*. These are described in the next section.
- **Optimisation Component.** In order to recommend the optimum homes based on the criteria of travel times and rental price, a depth-first heuristic search algorithm is used to approximate the set of Pareto optimal solutions taking into account the multi-objective constraints.

2.2 Computing Journey Travel Times

Though services exist such as Google Maps to provide estimated travel times using public transportation, it is important to note these services typically provide point-to-point travel times, working for one origin-destination (OD) pair at a time. For the purposes of evaluating many OD pairs, to support a real-time service taking into account user preferences, we employ a customized journey computation solution. This generalisation allows to greatly improve the computational speed. For instance, it allows to compute multiple journey plans, from one target location to many candidate houses, with one Dijkstra search.

The public transport data is represented in the GTFS format.¹ It includes knowledge about stops, routes and scheduled vehicle trips that follow a timetable. Stop details include their location (as latitude and longitude coordinates), their name and a unique id. Each trip belongs to a given route and represents the actual trip of a vehicle (e.g., bus) through an ordered sequence of stops. For each stop along the trip, the arrival and the departure time are provided as part of the GTFS data.

A standard journey plan request specifies the origin of a journey, the destination, and the start or the arrival time. We generalize this to be able to specify multiple OD pairs. We take into account so-called quotas, which are maximum acceptable values for parameters such as the total walking time and the number of hops (legs) in the journey. Additional constraints specify what transport modes can be considered in a given journey.

¹<https://developers.google.com/transit/gtfs/>

Journey plans are computed by searching in a state space whose states and transitions are defined as follows. A state $s = (l, t, q)$ is a vector containing a *relevant location* l , a time t when the traveller arrives at that location, and the quotas q available in that state. Relevant locations include all stops, all candidate houses, and the target locations indicated by a user, such as a work place or a school. As defined here, the set of relevant locations is much smaller than the set of all intersections in a city, which leads to a corresponding reduction in the computational effort needed to obtain journey plans. Transitions between states include walking actions and public-transport riding actions. Quotas q get updated after each application of an action. Every such an action updates the number of hops by 1. Walking actions update the walking time quota by an amount equivalent to the duration of the walking hop at hand.

The number of relevant locations on a map is reasonably small. For example, in Dublin, our GTFS data² contains 4,739 bus stops. However, the state space can be much larger, due to constraints such as bus schedules and quotas. For each bus stop, there can be many states in the search space, each having a different combination of quotas (e.g., max walking time and max number of hops allowed in a journey from that state to the destination). Such combinatorial blow-ups make table-based approaches including the Floyd-Warshall algorithm infeasible in terms of both time complexity and memory requirements.

In calculating travel times, our approach relies on the fact that there are many more houses than user specified target locations (hundreds of houses versus a few target locations). One invocation of the Dijkstra's algorithm computes optimal travel times from one user specified location to all houses, for one departure time. Given one user-specified target location, several runs of Dijkstra's algorithm are performed with different departure times. The travel times are averaged for each pair of a user specified location and a house. This process is repeated for every user-specified location.

Our implementation of Dijkstra's algorithm uses a priority queue to select the best state to examine next. It dequeues the best state s from the priority queue, generates successors of s , saves s in a hash table called the closed list, and enqueues these successors in the priority queue. These procedures are repeated until the algorithm either reaches all houses or proves that it cannot reach some houses without exceeding the available quotas specified by the user.

2.3 Pareto Optimal Solution for Multi-Criteria Evaluation

In many real-world situations, a decision maker has more than one objective. In the case of housing, examples of such objectives include the rent and the travel times to a few user-specified target locations. Each of these criteria has a different scale and therefore mapping these objectives to a single utility scale can be problematic, since the decision maker may be unwilling or unable to provide precise trade-offs between objectives [1].

As a result, it is natural to consider multi-attribute or multi-criteria utility functions to cope with multiple and non-commensurate utility scales on which the decision maker's preferences are expressed.

Given p attributes, a *multi-attribute utility value* is characterized by a vector $\vec{u} = (u_1, \dots, u_p) \in \mathbb{R}^p$, where u_i represents the utility with respect to attribute $i \in \{1, \dots, p\}$. Given $\vec{u}, \vec{v} \in \mathbb{R}^p$, we say that \vec{u} *Pareto dominates* \vec{v} (denoted $\vec{u} \succcurlyeq \vec{v}$) iff $\forall i \in \{1, \dots, p\} u_i \geq v_i$. As usual, the symbol \succ refers to the asymmetric part of \succcurlyeq , namely $\vec{u} \succ \vec{v}$ iff $\vec{u} \succcurlyeq \vec{v}$ and it is not the case that $\vec{v} \succcurlyeq \vec{u}$. Given a finite set of utility vectors $\mathcal{U} \subseteq \mathbb{R}^p$, we define

²Downloaded from <http://dublinked.ie>.

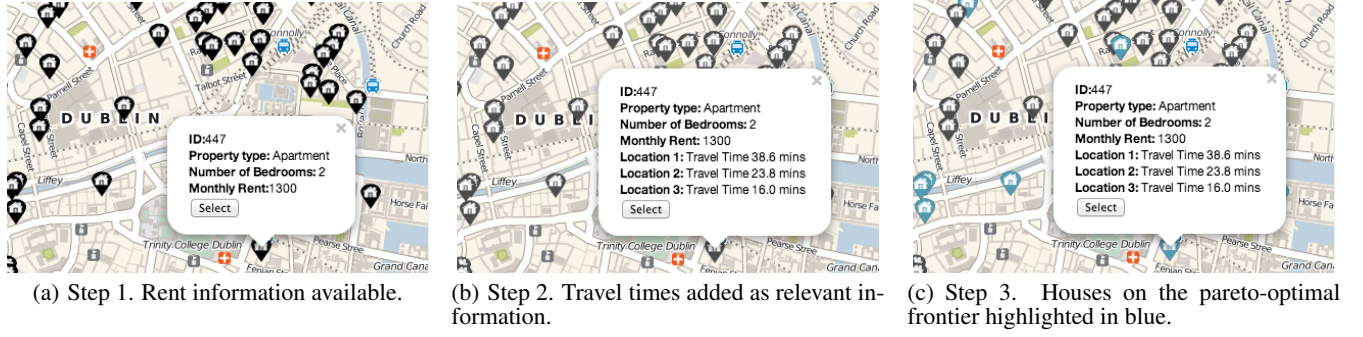


Figure 2: Main steps of the user study

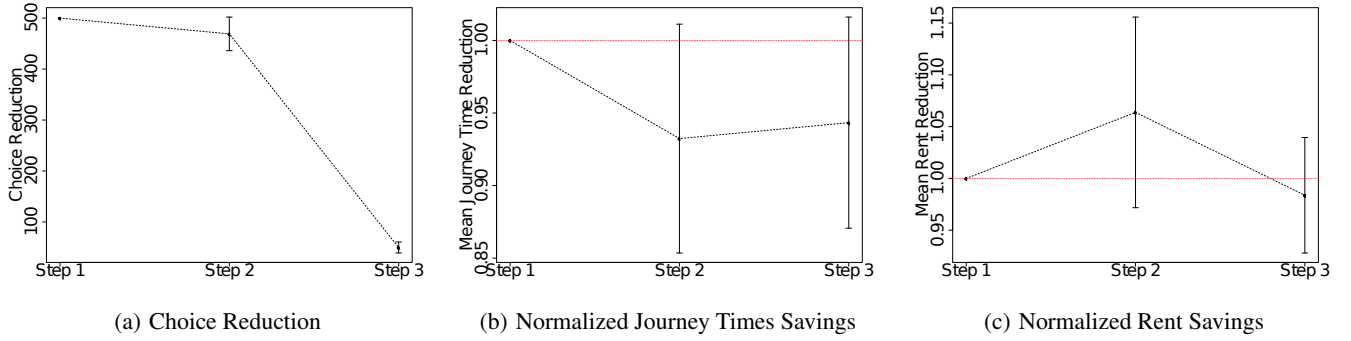


Figure 3: User Choice Optimization at each step. Journey times and rent are normalized by the user selection at step 1

the *maximal set*, denoted by $\max_{\succsim}(\mathcal{U})$, to be the set consisting of the undominated elements in \mathcal{U} , i.e., $\max_{\succsim}(\mathcal{U}) = \{\vec{v} \in \mathcal{U} \mid \nexists \vec{u} \in \mathcal{U}, \vec{u} \succ \vec{v}\}$. Computing $\max_{\succsim}(\mathcal{U})$ is quadratic in the size of \mathcal{U} .

In general, the Pareto ordering may lead to relatively large maximal sets and yet be useful to make an informed decision. Therefore, the decision maker may be happy to allow some additional trade-offs between attributes. Based on previous work by Marinescu et al. [6], we consider a set Θ of vector pairs of the form (\vec{u}, \vec{v}) , where $\vec{u}, \vec{v} \in \mathbb{R}^p$ and $\vec{u} \succ \vec{v}$. Such trade-offs may be elicited using some structured method, or in a more *ad hoc* way. The input preferences Θ (if consistent) give rise to a relation \succsim_{Θ} which may induce a much stronger ordering than the weak Pareto ordering. However, computing the maximal set with respect to \succsim_{Θ} is more involved and requires solving a linear program representing Θ [6].

In JOURA, each candidate housing location is associated with a four-attribute value representing the rent of that house and the average travel times from the house to each of the three target locations. The solution recommended by the system is the set of undominated housing locations. If additional trade-offs are elicited (e.g., one prefers to travel 15 more minutes to the first target location to decrease the rent by 100 euros) then the subset of candidate houses can be reduced even further, to the set of optimal (i.e., non-dominated) housing locations to account for the trade-offs.

3. EVALUATION - USER STUDY

To test the effectiveness of the JOURA system we created a web application to allow users to view a selection of available homes throughout Dublin. The purpose of this study was to better understand the effect of a) showing travel times to users and b) providing

a recommended subset of the homes that reflect an optimal trade-off across multiple criteria. For the purpose of this study we limited our criteria to rent and travel times, all homes were displayed as 2 bedroom apartments to control for other preferences. When the journey times are calculated, the maximum walking time is set to 15 minutes, and the maximum number of hops is set to 5. The Dublin public transportation data that we used contains 4,739 bus stops and 120 bus routes. Each route is served by multiple trips, at different times of the day, for a total of 16,777 trips per day.

We recruited 20 participants, all working and living in Dublin. Users were asked to select three target locations of interest and in real time the journey times between each candidate home and the targets were computed. The users were then guided through a three step process where they could select from 500 homes located in various parts of the city. Each step provided more information to the users, with the benefits of a more informed decision and, in the case of the third step, a significantly simpler decision process for the user, as a result of filtering away dominated candidate houses. Clicking on a home allows the user to see the details of that home available at a given step (e.g., rent at Step 1).

- **Step 1:** In order to understand where a user would select a house based purely on their own knowledge of the city we first ask users to select a home they felt provided the best trade-off between rent and accessibility to their chosen locations. A user was given only basic information about the home and the rent, as shown in Figure 2 a). The user could select a home where the details remain displayed to the right of the map, in order to support the user comparing different options before making their final selection.

- **Step 2:** Users are provided with additional information where the journey times to each of their chosen locations is shown to the user, as seen in Figure 2 b).
- **Step 3:** Finally, as shown in Figure 2 c), non-dominated homes, which best reflect a trade-off between journey times and rent, are highlighted in blue. As all homes remain visible, users were still given the option of choosing a non-optimal home if desired.

3.1 Choice Reduction

Navigating through the complete set of available homes is a time consuming task for the user, particularly when processing several criteria at once. Algorithmically however, it is very efficient to compute homes that reflect a trade-off and therefore narrowing down the selection process for the user.

Figure 3 a) shows the reduction in the number of homes a user needs to evaluate at each step in the process. There is a small reduction from step 1 to step 2 where some homes are inaccessible through public transportation links due to the constraints of the maximum walking time and maximum number of hops. The reduction in step 3 includes only the recommended options, even though all options were available to the user. At the final stage an average of 50 homes were included in the recommended subset. This 10-fold reduction allows the users to focus on a much smaller set of candidate homes, under the guarantee that each house filtered out is dominated on all four criteria by at least one house in the preserved subset.

3.2 Effectiveness of Recommendations

Overall, 90% of the participants had chosen a house that was included in the recommended subset by the end of the experiment. However, 45% of the users did not change their selection between step 1 and step 3. When examining whether their original selection was included in the recommended subset, in 77.77% of the instances the house was included in the final recommendations meaning the user had already selected a home that represented a valid trade-off between their goals of travel times and rent.

3.3 Journey Times Savings

Figure 3 b) shows the normalized travel time reduction for users as they progress through the user study. The confidence intervals are quite large, this is a result of 45% of the users selecting the same home for each step, resulting in zero savings. For the remaining users however we do see a reduction in overall travel time between step 1 and step 2. Interestingly there is a slight increase in overall travel time between step 2 and step 3 as users are shown the recommended options. This is due to a reduction in rent at the final step where the users make a compromise of a slight increase in travel time for a lower rent.

3.4 Rent Savings

Figure 3 c) shows the normalized rent reductions for users at each step. As can be seen at step 2, when users are shown information about journey times they optimize for travel times and have trouble optimizing their selection over multiple different criteria. At step 3 however, we see that when they are presented with the optimal recommendations, the users select a better trade-off across the goals of reduced rent and journey times.

4. CONCLUSION

In real life, travel times to relevant destinations, such as a workplace, play an important role in the overall evaluation of a house

to purchase or rent. For instance, houses located in the proximity of main transportation routes, such as a metro line, might be seen as more suitable to one's mobility needs. In this work we have focused on incorporating public transportation links and user journey requirements into a housing recommender system.

Clearly, selecting a home is affected by more factors besides the set considered in this work (e.g., rent and travel times to a few selected locations). In future work, we plan to integrate additional environmental factors, such as security, amenities, and convenience facilities.

5. REFERENCES

- [1] T. Brink, R. Keeney, and H. Raiffa. Decisions with multiple objectives, preferences and value tradeoffs. *Behavioral Science*, 39(2):169–170, 1994.
- [2] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: Recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 201–210, New York, NY, USA, 2009. ACM.
- [3] A. Garcia, P. Vansteenwegen, O. Arbelaitz, W. Souffriau, and M. T. Linaza. Integrating public transportation in personalised electronic tourist guides. *Computers & Operations Research*, 40(3):758–774, 2013.
- [4] J. A. Konstan and J. Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.
- [5] N. Lathia and L. Capra. Mining mobility data to minimise travellers' spending on public transport. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1181–1189. ACM, 2011.
- [6] R. Marinescu, A. Razak, and N. Wilson. Multi-objective influence diagrams. In *Proceedings of the Twenty-Eighth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-12)*, pages 574–583, Corvallis, Oregon, 2012. AUAI Press.
- [7] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft. Recommending social events from mobile phone location data. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 971–976, 2010.
- [8] F. Ricci and F. del Missier. Designing personalized user experiences in ecommerce. chapter Supporting Travel Decision Making Through Personalized Recommendation, pages 231–251. Kluwer Academic Publishers, Norwell, MA, USA, 2004.
- [9] J. B. Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. In *Applications of Data Mining to Electronic Commerce*, pages 115–153. Springer US, 2001.
- [10] G. Tumas and F. Ricci. Personalized mobile city transport advisory system. In *ENTER*, pages 173–183. Springer, 2009.
- [11] L. Wang, X. Hu, J. Wei, and X. Cui. A collaborative filtering based personalized top-k recommender system for housing. In *Proceedings of the 2012 International Conference of Modern Computer Science and Applications*, pages 461–466. Springer, 2013.
- [12] X. Yuan, J.-H. Lee, S.-J. Kim, and Y.-H. Kim. Toward a user-oriented recommendation system for real estate websites. *Information Systems*, 38(2):231 – 243, 2013.