Managing travel demand: Location recommendation for system efficiency based on mobile phone data

Yan Leng
MIT Media Lab
Cambridge, MA
yleng@mit.edu

Larry Rudolph
Two Sigma LP & MIT
New York, NY
larry.rudolph@twosigma.com

Alex 'Sandy' Pentland
MIT Media Lab
Cambridge, MA
pentland@mit.edu

Jinhua Zhao
MIT
Cambridge, MA
jinhua@mit.edu

Haris N. Koutsopolous
Northeastern University
Boston, MA
haris@mit.edu

ABSTRACT

Growth in leisure travel has become increasingly significat economically, socially, and environmentally. However, flexible but uncoordinated travel behaviors exacerbate traffic congestion. Mobile phone records not only reveal human mobility patterns, but also enable us to manage travel demand for system efficiency. In this paper, we propose a location recommendation system that infers personal preferences while accounting for constraints imposed by road capacity in order to manage travel demand. We first infer unobserved preferences using a machine learning technique from phone records. We then formulate an optimization method to improve system efficiency. Coupling mobile phone data with traffic counts and road network infrastructures collected in Andorra, this study shows that uncoordinated travel behaviors lead to longer average travel delay, implying the opportunities in managing travel demand by collective decisions. The interplay between congestion relief and overall satisfied location preferences observed in extensive simulations indicate that moderate sacrifices of individual utility lead to significant travel time savings. Specifically, the results show that under full compliance rate, travel delay fell by 52% at a cost of 31% less satisfcation. Under 60% compliance rate, 41% travel delay is saved with a 17% reduction in satisfaction. This paper highlights the effectiveness of the synergy among collective behaviors in increasing system efficiency.

Keywords

Congestion alleviation; location recommendation; data mining; optimization; mobile phone data

1. INTRODUCTION

Careful coordination of the travel behavior of tourists may reduce congestion, improve their travel experience, improve the quality of the environment, improve the quality of life for the local population, and avoid the many problems with tourists [25, 13, 12]. Coordination has grown to become even more important as leisure travel has continued to increase, both internationally and domestically, contributing to 10% of global GDP and 6% exports in 2015 [1]. International tourism increases around 4.2% to 6.6% since 2010 and
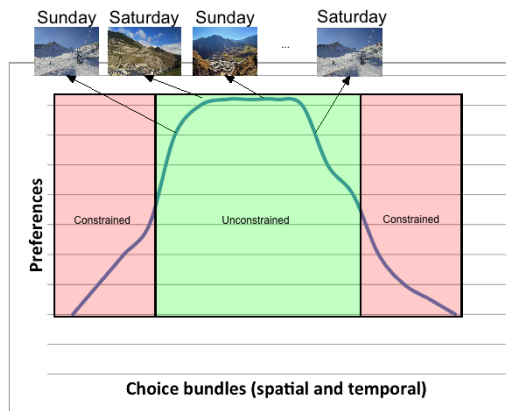

Figure 1: Choice flexibilities.

reached a record of 1.2 billion traveler-arrivals in the same year [1]. This paper proposes a personalized event recommendation system that mitigates the adverse effects of mass tourism, while respecting the needs of all stakeholders.

Our solution can be considered a type of *travel demand management*, i.e., a paradigm to reduce or shift travel demand in space or time to reduce the negative impacts[23]. However, little research has focused on managing tourism travel demand. Unlike commuting trips, where the travel time and desinations are fixed, leisure travelers are more flexible. Figure 1 illustrates how individuals' preferences of location and time bundles are relatively flat above a certain threshold, indicating the flexibility of tourists' travel [13].

Recommendation systems have been successful tools in eCommerce. Current location recommendation algorithms, mostly adopted from movie or book recommendations, simply make recommendations based on inferred personal preferences. Applied to travel, however, this method may lead to even more severe congestion and longer waiting times if the most popular locations are recommended at the same time. We argue that *system efficiency*, the interplay between the location preferences at a user level and the traffic congestion at a system level, should be balanced when using location recom-

mendations, which can serve as strategies in travel demand management.

The availability of large-scale geolocation data from mobile devices, such as the Call Detail Records (CDR) used in this study, offers an unprecedented opportunity for location-based service providers, transportation agencies, tourism departments, and governments to understand human mobility pattern, provide personalized information and improve system-wide performance [11, 22, 38]. Making the comprehensive picture of population-wide behaviors enables decision-makers to intervene at a system level. Call Detail Records, used to understand travel behaviors, have rarely been used to understand and manage travel demand or recommend locations [13, 26, 4]. Therefore, a recommendation system can be built based upon not only satisfying personal preferences at the demand side but also making the best use of the system capacity at the supply side.

In this paper, we use location recommendation to manage travel demand to achieve system efficiency. We show that uncoordinated travel behaviors lead to severe traffic congestion. At the same time, we propose a method as a solution for transportation practitioners or authorities to optimize and trade-off satisfied preferences and road congestion. We use matrix factorization to mine travelers' implicit preferences, taking advantages of underlying similarities among locations and travelers. We then formulate an optimization problem to maximize satisfied location preferences at the user level under pre-defined road congestion constraints. The method reveals the interplay between system congestions and user preferences. With an implementation with the CDR data in Andorra under various compliance rates, we show the effectiveness of the method. For example, under a 100% compliance rate, a 52% reduction in travel delay (from 11.73 minutes per hour to 5.61 minutes per hour) only sacrifices 31% satisfaction regarding the recommendations. Similarly, under a 60% compliance rate, a 41% reduction in travel delay (to 6.98 minutes per hour) only sacrifices 17% in satisfaction regarding the recommendations.

This paper is organized as follows: Section 2 summarizes current literature regarding travel demand management and location recommendations. The data required to perform the study is described in section 3. Section 4 demonstrates the framework of the method and details each step, including preference inference, and collective satisfaction maximization. Section 5 describes a case study in Andorra. The performance of the proposed method is compared with a baseline model where location is recommended based simply on travelers' preferences. The impact of the method is analyzed under various compliance rates. Section 6 concludes the paper and discuss future work.

2. RELATED WORKS

2.1 Travel demand management
Travel Demand Management (TDM) encompasses strategies that alter demand patterns to increase transportation system efficiency, instead of adding more capacity to the system [23]. TDM has broad applications, including energy savings, air quality improvements, peak period congestion alleviation, etc [17].

Different categories of TDM strategies include, economic policies, physical change measures, legal policies, and information or education measures [20, 23]. Economic policies, the most popular TDM strategies, include taxing vehicles, congestion pricing, lowering transit costs, etc [37, 23]. Physical change measures, such as walking/cycling improvements, park-and-ride schemas, represents another category in TDM [31]. Legal policies, prohibit traffic in some areas or parking controls.

Many of the above strategies are not applicable in the context of tourism. Relatively little TDM research has targeted tourism demand, which is more flexible than commuting-related travels. Therefore, the following characteristics of this work differentiate it from prior travel demand management researches, namely:

- We focus on flexible travel demand, which can be manipulated at the destination and time-of-day levels.

- We propose to use Call Detail Records, a large-scale and opportunistic data source, to understand travel patterns.

2.2 Location recommendations
In the early 2010s, several studies introduced traditional recommender engines to personalized location recommendation. Ye (2011) [41] introduced user-based and item-based Collaborative Filtering (CF) to location recommendations using user check-in data, based on the assumption that similar users have similar tastes and users are interested in similar Points of interests (POIs). Berjani (2011) [10] employed the more effective and efficient matrix factorization in POI recommendations on check-in history. Regularized matrix factorization is used to learn the latent feature matrix, which has better performance than item-based CF. Recent research focuses on utilizing additional information. Geospatial factors, social networks, and temporal influences are three main examples. Other researchers argue that users prefer nearby locations rather than distant ones, which is defined as the geographical clustering phenomenon [43, 42]. Some researchers make the assumption that friends share more common interests than non-friends to utilize social influence in recommendation [34, 19]. Finally, to make use of temporal influence on activities, some researchers make seperate location recommendations for different temporal states [8, 42, 27]. Quercia (2010) [35] is the only work that makes recommendations using mobile phone data. However, his paper is based exclusively on item-based CF, which is computationally inefficient and hard to scale [41, 39]. Application of existing methods, ignoring service capacity constraints, may result in traffic congestion and long waiting times, no matter how sophisticated these methods are in inferring preferences.

Therefore, this paper has the following improvements in location recommendation:

- We argue that capacity constraints are an important characteristic of location recommendation, which is currently ignored by the literature. We integrate capacity constraints in the location recommendation method.

- We develop a framework to recommend locations for system efficiency based on Call Detail Records. It can be used in other cities when call records, traffic counts and road network GIS files are available. It can also be applied on other longitudinal data sources, such as WiFi, GPS, AFC, etc.

2.3 Next-location prediction

Next-location prediction has been an increasingly popular topic in pervasive computing based on GPS, bluetooth, check-in histories, etc. Accurate next-location predictions, given previous footprints from these data sources, is a significant building block benefiting many areas, including mobile advertising, public transit planning, and urban infrastructure management [5, 21, 28, 33]. Different data sources vary at spatial and temporal scales, and depending on the availability of contextual information [14]. Most researchers build Markov models and predict longitude and latitude as continuous variables based on previous travel trajectories [6, 7, 18]. Mathew [29] predicts next-location using a Hidden Markov Model with contextual information, such as activities and purposes. Domenico [14] and Alhasoun [5] uses mobility correlations, either measured by social interactions or mutual information, to improve forecasting accuracy. Though extensive researches have acceptable accuracy in predicting next locations, the performance is poor with when Call Detail Records are sparse [5].

In this paper, we develop a new perspective in viewing mobility behaviors based on Call Detail Records as sentences. We then introduce the use of Recurrent Neural Network, a successful tool in language modeling, into mobility prediction.

3. DATA

In addition to the Call Detail Records, we make use of the topology of the Andorra road network, its capacity, and periodically recorded traffic counts. This information is combined to understand travel demand patterns and transportation system performance.

3.1 Call Detail Records

The Call Detail Records were originally used for billing purposes. A record is stored when a mobile phone user connects to the network of mobile carriers. Each Call Detail Record entry contains an encrypted user ID, start and end times of the phone call, the IDs of the connected cell tower ID, and the origin nationality and model type of the phone, see Figure 2. The cell tower ID is easily converted to geographic location of the cell tower in the provider's network.

The anonymized CDR data used in this paper is collected from Andorra, a small country situated between France and Spain. As a case study, we target travelers visiting Andorra during a week in May, 2015 to provide personalized recommendations for system efficiency. These 47743 tourists include 20311 tourists from France and 27432 from Spain. The spatial distribution of the cell towers are shown in Figure 3, with different colors representing different cities in which the tower is located.

user_id	datetime	antenna_id	direction	interaction	Country	Phone type
dsdasdfasdfqwerwoiouiruf9w90283u4oijofjlksdkfansdkgklsjoqi4029830498203	2015.08.30 08:17:47	90	out	text	21465	986745342
f2436657210815553388ed1adacec6e38c1fwewer20ed3d560a4a6bb8e55571	2015.08.30 08:16:06	461	in	call	21464	986745342
efdkssdfeedwerwoiouiruf9w90283u4oijofjlksdkfansdkgklsjoqi4029830sdfasddf	2015.08.30 08:04:32	223	in	call	24134	986745342
7a27adb89209cb27eec47bc79aaasdfwewef612f77760e608fd2ea53c3a74f0c95	2015.08.30 08:40:02	454	out	call	23413	986745342

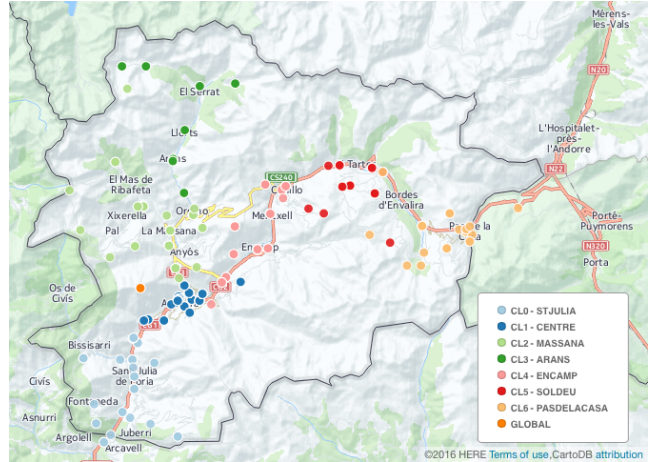**Figure 2: Snapshot of Call Detail Records.**

Figure 3: Cell tower distributions in Andorra. Each circle represents a cell tower. Cell towers in the same city are colored the same, with the legend an abbreviation of a city name.

3.2 Traffic flow derived from CDR and traffic counts

We assume that the number of travellers using their cell phones while they on the highway is a constant fraction of the number of vehicles on the highway. Under this assumption, traffic flow derived from the Call Detail Records can be used to approximate actual traffic flow.

The Andorra road network is limited with a single major route between each city. Traffic counts have been collected at key locations using cameras by local authorities to monitor internal mobility. This data is publicly available [16]. The GIS shapefiles of road networks were acquired from the Andorra Transportation Department. Important attributes used in the simulation include connecting cities, number of lanes, capacities and free flow travel time, which enable the computation of travel times based on traffic flow. Free-flow travel time were obtained from the Google Map API. Road capacities were obtained from a NCHRP report [30].

From the phone record time and cell tower information, we can generate an Origin-Destination matrix, the number of trips between cell towers, i.e. Combining it with the road network information gives traffic flow over time, by group, and between two locations.

4. METHOD

The goal is to send recommendations for places to visit and when to visit them. This is done in three steps, which are outlined in Figure 4. The first step infers travel demand in terms of vehicle trips along road links, from mobile phone records. The second step infers personal location preferences based on location traces with no explicit ratings. With matrix factorization, we infer these implicit location preferences
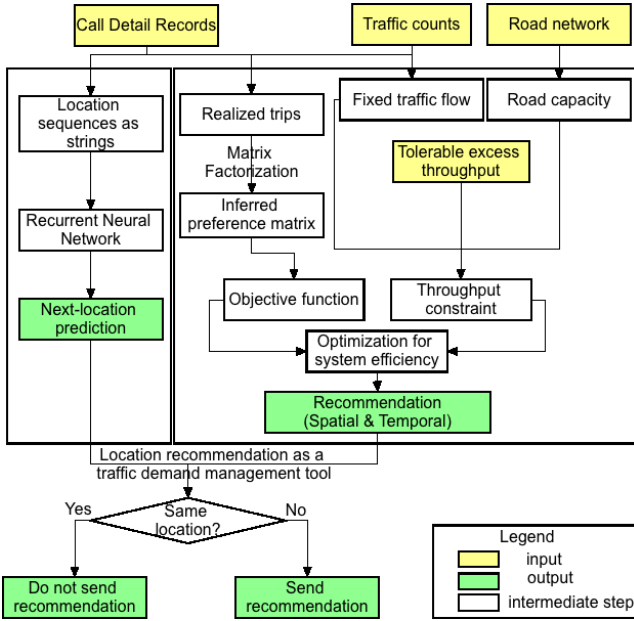
Figure 4: Methodological framework.

regarding all the locations with hidden factors, which correlate with the characteristics of both travelers and locations. With the inferred preferences, an optimization problem is formulated with the objective of maximizing satisfactions regarding the recommendations subject to tolerable congestion levels, which can be determined by the decision-makers. Finally, we predict next locations based on historical traces and compare them with the recommended locations to determine whether a recommendation will be sent.

4.1 Terminology

The following terms are used throughout the method as defined below.

User profile. User profiles contain the longitudes, latitudes, timestamps, and characteristics of the user. A user profile $l_{u,g}$ is generated for each user based on individual mobility traces $(l_{u,t_1,g}, l_{u,t_2,g}, ...)$. g is the user group of user u based on his/her characteristics directly obtained or inferred from CDR. t_i is the number of presences at location i.

Realized trips. Realized trips p_{ij} are calculated by the number of times individual i traveled to location j. Realized trips are used as a proxy for location preferences.

Idealized trips. Idealized trips $\hat{p}_{ij}$ measure the preferece of traveler i regarding the location j with no observed presences. Similarly, it is a measure of proxy travelers' preferences regarding the locations with no observations.

Tolerable excess throughput. Tolerable excess throughput is determined by the tolerable congestions by the decision-makers.

4.2 Traffic flow inference

Mobile phone data provides an imperfect estimate of the travel demand and traffic flow along road links. In order to understand travel delays, CDR data need to be processed

to derive individual movements and vehicle trips from the actual flow which is derived from the CDR-based travel demand and the actual travel demand.

The first step is to extract the tower-to-tower Origin-Destination (O-D) matrix, aggregated by the individual movements from one cell tower to another cell tower. Peak hour traffic flow can be learned as traffic flow varies heterogenously by links and hours of the day. The O-D pairs are assigned to road links given the road network. The last step is to scale the aggregated movements to actual vehicle trips using traffic counts as the ground truth, which is calculated in Equation 2.

$$R_{it} = \sum_{jkt} O_{jt} D_{kt} \quad (1)$$

$$TC_{it} = R_{it} \times \beta_{it} \quad (2)$$

where i is the index of road link, t is the index for hour of day, $O_j D_k$ represents OD pair with origin j and destination k, R_i is the vehicle trips along road link i, TC is the actual traffic counts and β is the scaling factor.
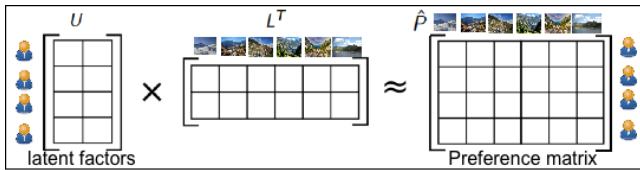
4.3 Preference inference

As no explicit review or rating regarding locations is available in CDR, we propose to use "visiting frequencies" as a proxy for location preferences. The next problem is to infer preferences regarding locations with no observed visits.

Matrix factorization, one type of latent factor model, is used to infer travelers' preferences regarding new locations. This model characterizes both the locations and users by vector of factors inferred from location visiting patterns, mapping both travelers and locations to a joint latent factor space of dimensionality k. The latent factor space determines why and how traverlers like each location based on hidden characteristics, which can be interpreted as personal interests or land use categories. High correspondence between location and user factors, based on the characterization of both the locations and the travelers, leads to a recommendation [27].

4.4 Optimization for system efficiency - individual preferences vs. congestion

The key idea of the proposed method is to optimize travelers' location preferences with the constraint of acceptable congestion. The authorities will then have the freedom to tradeoff between these two factors. An optimization model was built to maximize preferences regarding location recommendations subject to road capacity constraints. This model can be easily extended to other cases where capacity constraints exist by modifying the constraint accordingly.

This paper formalizes the traffic problem by modeling destination and time choice as follows: each traveler i makes a choice of location j and the day for travel day t. The choice is made based on personal utility p_{ijt}, which is assumed to be the preferences regarding the location j inferred from the call records. Since travelers make selfish choices independent

Figure 5: Illustration of the preference inference methodology via matrix factorization to infer the preferences of travelers regarding locations with no observations. Matrix U captures user's characteristics or interests (hidden factors). Location matrix L characterizes the associations of the locations with the latent factors (point of interest categories). The multiplication of the two predicts travelers' preferences across the location space.

of any other larger constants, the system may settle into a suboptimal state. In a suboptimal state, the travel time delay of the whole system as well as the congestion are higher than they should be. The set of destination choices that occur when every traveler maximizes their satisfied preferences is referred to as the user equilibrium flows [13, 2], which is similar to Wardrop's principles in route choice [40].

The objective function of the formulated optimization model aims to maximize the overall satisfied preferences regarding the location recommendations. The constraint function is determined by the acceptable congestion of the decision-makes.

4.5 Next-location prediction with Recurrent Neural Network

In order to distribute recommendations more efficiently, the method sends recommendations only when the predicted next-locations are not in line with the ones to be recommended. In this section, we predict individual next-locations based on historical location traces.

Recurrent neural network is an adaptation of the traditional feed forward neural network, which can process variable-length sequences of inputs. It has been successfully used in language models, learning word embeddings, financial time series predictions [32, 36]. In this project, we apply sophisticated recurrent neural network into mobility prediction by mapping between mobility models to language models. Cell tower traces for each individual are modeled as a sentence and each cell tower as a word. We use a simple RNN architecture, with a input layer, a hidden layer, a long short-term memory layer and output layer. The cell tower with the maximum probability is predicted to be the next location. The predicted next-location can be compared with the recommended location to determine the action to be taken.

RNN is advantageous in predicting next locations in two ways.
Location sequences: Travelers visit locations in a sequential way and RNN reads in data sequentially.
Variable number of visited locations: The heterogeneity in the size of location traces and frequency of mobile phone usage makes traditional machine learning techniques inap-

plicable. The ability to handle variable input lengths makes RNN appropriate in this situation [15].

5. APPLICATION

We performed extensive simulations to gain insights into the interplay between satisfaction regarding the recommendations and the travel delay caused by congestion. We vary the compliance rate across the population, i.e. the probability that travelers will follow the recommendation, in order to evaluate the potential traffic improvement of the recommendation system. The simulation assumes that the individuals who do not comply will follow their preference with no behavioral change.

5.1 Location recommendation for system efficiency

Using the method described in section 4, we simulate the average travel delay per hour and overall idealized trips to assess the system-wide impacts and effectiveness of the method. The notion of idealized trips is a measure of satisfaction regarding the recommendations. Average travel delay measures the congestion externalities of the recommendations. In transportation, practitioners and planners model the relationships between traffic flow and travel time based on the characteristics of the road infrastructure. One of the most widely used methods is Bureau of Public Road function (BPR) [3], which models travel time as a function of the ratio between actual traffic volume and road capacity, volume-over-capacity (VOC) [9], as shown in Equation (3).

$$t_{\text{simulated}} = t_{\text{free-flow}} \times (1 + \alpha(V/C)^{\beta}) \tag{3}$$

Average travel delay:

$$\Delta t = t_{\text{simulated}} - t_{\text{free flow}} \tag{4}$$

where $t_{\text{free-flow}}$ is the free flow travel time on the road segment, $t_{\text{simulated}}$ is the simulated travel time, Δt is the delay caused by congestion, α and β are parameters that are used to characterize the non-linear relationship between V/C and $t_{simulated}$. The default parameter for the BPR equation are $\alpha = 0.15$ and $\beta = 4$.

We compare our method with a baseline model, which is referred to as *preference only* method. This method makes recommendations simply according to personal preferences with no system performances taken into account. Table (1) summarizes average travel delay per hour and idealized trips of: 1) the preference only method; 2) the proposed method under various compliance rates. Under full compliance, the average travel delay is 5.61 minutes per hour with 44930 idealized trips, which indicates that with a 52% reduction in congestion time, only 31% of idealized trips are sacrificed. When the compliance rate is 80%, a 47% reduction in congestion time is achieved with only 23% of idealized trips being sacrificed. The lower the compliance rate, the larger the idealized trips and the longer travel delay. Figure 6 shows the relationship between idealized trips and tolerable excess throughput, which indicates individual perceived benefits under various level of congestion controls. Higher compliance rates satisfy larger individual benefits, which generate

Scenario	Δt (min/h)	Idealized trips
Preference only	11.73	64925
100% comp.	5.61	44930
80% comp.	6.17	49997
60% comp.	6.98	53680
40% comp.	8.37	57442
20% comp.	10.40	61219

Table 1: Results comparisons. Δt is the average travel delay during the peak hour. Idealized trips measures travelers' satisfaction regarding the recommended locations. Comp. is short for compliance rate. Preference only is the baseline model where recommendations are based on predicted preferences.
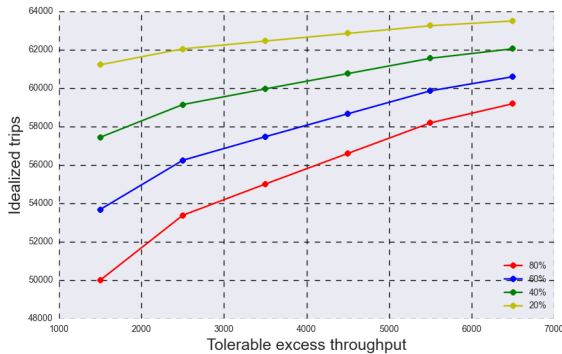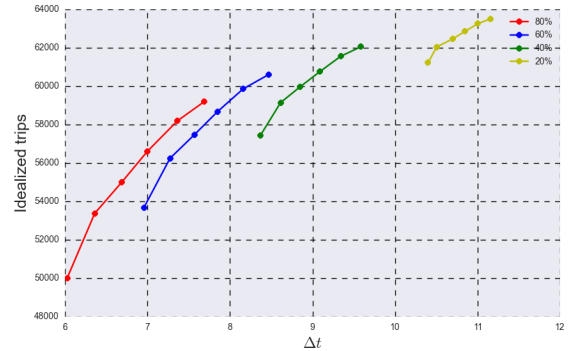

Figure 6: Relationships between idealized trips and tolerable excess throughput.

Figure 7: Interplay between idealized trips and average travel delay.

Table 2: Cell-tower level		
	Accuracy	Improvement
Naive model	50%	NA
Markov model	54%	8%
RNN	67%	34%

Table 3: Merged-cell-tower level		
	Accuracy	Improvement
Naive model	63%	NA
Markov model	57%	-16%
RNN	78%	41%

more traffic. The concave relationship reveals that preferences are satisfied more quickly in the beginning and slow down afterwards. The tolerable excess throughput enables decision-makers to manage congestion at an acceptable level. Figure 7 reveals the interplay between idealized trips and average travel delay. A horizontal reference line shows that for the same level of idealized trips, synergized behaviors generate less congestion. On the other hand, a vertical reference line shows that for the same travel delay, coordinated behaviors are more effective. This indicates the importance of effective schemas to incentivize behavioral change to achieve synergetic effects.

5.2 Next-location prediction

As a case study, we evaluate different methods using CDR data collected over three weeks during May, 2015 in Andorra. We apply the method specifically on tourists, which can be filtered on the country code from CDR. We do not exclude travelers with too few observations as long as their travel call includes more than one cell tower, suggesting the generalarization of the method. We use two settings with different spatial resolutions, at the cell-tower level and the merged -cell-tower level.

We introduce two baseline models: "Most Frequent" model and Markov model. "Most Frequent" model, referred to as the naive model, predicts the next location using the most frequent location. The Markov model is built on the contex-

tual co-occurences between sequences of locations [24, 18, 28]. Various parameters have been tuned to optimize the performance of RNN, including activation function, dimension of embedding layer, drop out rates, sample size, and batch size.

Tables 2 and 3 show that the naive model has 50% and 63% accuracies in predicting next location at the cell-tower and merged-cell-tower level. The Markov model across the population has 54% accuracy at the cell tower level, with 8% improvement. However, at the merged cell tower level, the population-wide markov model has lower accuracy than the naive model. Comparatively, RNN improves the accuracy of location prediction significantly, with an accuracy of 67% and 78% on the two settings. It improves 34% and 41% in accuracy compare with the two settings, indicating that RNN significantly improves the accuracy of the next-location prediction.

6. CONCLUSIONS AND FUTURE WORK

We have shown that individual travel decisions, without accounting for system efficiency, lead to traffic congestion. Existing location recommendation algorithms exacerbate congestion by recommending popular locations. To address this issue, we develop a location recommendation method to manage travel demand. Call Detail Records are the raw input and are processed to infer personal location preferences, traffic congestion, and other information about both tourists and native travellers. This study, as far as we know, is the

first one making location recommendations based on Call Detail Records. Most importantly, we factor in the special characteristics, capacity constraints, of location recommendation, which distinguish our system from existing recommendation methods. The simultaneous trade-off between congestion relief and overall satisfied location preferences learned from the simulation results indicate that moderate sacrifices for individual utilities lead to significant collective travel time saving.

The simulation results from our Andorra case study reveal a noticeable impact in reducing traffic congestion with moderate sacrifices on individual preferences. For example, under 100% compliance, there is a 52% reduction in travel delay (from 11.73 minutes per hour to 5.61 minutes per hour) with 31% dissatisfaction rate regarding the recommendations. Even with a much smaller compliance rate, under 60%, there is a 41% reduction in travel delay (to 6.98 minutes per hour) with only a 17% dissatisfaction rate. We use a recurrent Neural Network approach to make sense of the input data. With the implementation of RNN on the large-scale CDR collected in Andorra, this paper demonstrates the applicability of the method with accuracies of 67% and 78% at cell-tower and merged-cell-tower levels, representing an improvement of greater than 30% compared to two base-line models.

The method developed in this paper specifically targets leisure travel, where travelers are relatively more flexible in the decision-making processes at spatial and temporal levels. We aim to divert travelers to their "secondary" choices at the location and the time to visit, while gaining more travel-time savings both for the individuals and for the society as a whole.

This research opens up multiple directions to advance in the topic of location recommendation for system efficiency using Call Detail Records. In this study, we only use travel delay as the systematic efficiency measure. However, other externality measures, such as air pollution or energy consumption, should also be factored in for comprehensive evaluations from the system side. In addition, except for time and destination, other interesting factors could be included in the choice bundle, such as travel routes, budgets, etc.

The proof-of-concept experiments in this study demonstrate the effectiveness of our approach. The natural next step is to investigate how to incentivize users to sacrifice perceived benefits for better system performances. A comprehensive framework of the application in real situations, detailing the distribution channel, frequencies, target markets, needs to be studied further from a marketing perspective.

An interesting extension of the paper is the information configurations for travelers, specifically, how to strategically present information accounting for travelers' willingness to accept the information. An web-interface-based experiment could be developed to help understand travel behaviors and decision-making processes, and how the system dynamics perform by providing various recommendation configurations.

Call Detail Records constitute one longitudinal data source in understanding travel behaviors. Integrating with other data, such as WiFi and bluetooth data, could supplement it and enhance the application - providing better spatial and temporal resolutions, social media data - capturing momentary feelings.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] W. T. O. (2016). Unwto annual report 2015. Technical report.

[2] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Informational braess' paradox: The effect of information on traffic congestion. *arXiv preprint arXiv:1601.02039*, 2016.

[3] R. Akcelik. Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. *Australian Road Research*, 21(3), 1991.

[4] L. Alexander, S. Jiang, M. Murga, and M. C. González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 2015.

[5] F. Alhasoun, M. Alhazzani, and M. C. González. City scale next place prediction from sparse data through similar strangers. 2016.

[6] A. Asahara, K. Maruyama, A. Sato, and K. Seto. Pedestrian-movement prediction based on mixed markov-chain model. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 25–33. ACM, 2011.

[7] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[8] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel. Recommendations in location-based social networks: a survey. *GeoInformatica*, 19(3):525–565, 2015.

[9] M. Ben-Akiva and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, pages 5–33. Springer, 1999.

[10] B. Berjani and T. Strufe. A recommendation system for spots in location-based online social networks. In *Proceedings of the 4th Workshop on Social Network Systems*, page 4. ACM, 2011.

[11] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[12] S. Çolak, A. Lima, and M. C. González. Understanding congested travel in urban areas. *Nature communications*, 7, 2016.

[13] M. De Domenico, A. Lima, M. C. González, and A. Arenas. Personalized routing for multitudes in smart cities. *EPJ Data Science*, 4(1):1–11, 2015.

[14] M. De Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility

and social interactions. *Pervasive and Mobile Computing*, 9(6):798–807, 2013.

[15] W. De Mulder, S. Bethard, and M.-F. Moens. A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech & Language*, 30(1):61–98, 2015.

[16] D. D'Estadistica. Internal mobility. 2016.

[17] FHWA. Travel demand management. 2014.

[18] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[19] H. Gao, J. Tang, and H. Liu. Personalized location recommendation on location-based social networks. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 399–400, New York, NY, USA, 2014. ACM.

[20] T. Gärling and G. Schuitema. Travel demand management targeting reduced private car use: effectiveness, public acceptability and political feasibility. *Journal of Social Issues*, 63(1):139–153, 2007.

[21] J. B. Gomes, C. Phua, and S. Krishnaswamy. Where will you go? mobile data mining for next place prediction. In *Data Warehousing and Knowledge Discovery*, pages 146–158. Springer, 2013.

[22] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[23] A. A. F. Halvorsen. *Improving transit demand management with Smart Card data: general framework and applications*. PhD thesis, Massachusetts Institute of Technology, 2015.

[24] H.-P. Hsieh, C.-T. Li, and X. Gao. T-gram: A time-aware language model to predict human mobility. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[25] M. Ivanovic, P. KHUNON, N. Reynish, R. Pawson, L. Tseane, and N. Wassung. Fresh perspectives: tourism development. *Cape Town: Pearson*, 2009.

[26] S. Jiang, J. Ferreira Jr, and M. C. González. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. In *Int. Workshop on Urban Computing*, 2015.

[27] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 831–840. ACM, 2014.

[28] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson. Approaching the limit of predictability in human mobility. *Scientific reports*, 3, 2013.

[29] W. Mathew, R. Raposo, and B. Martins. Predicting future locations with hidden markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 911–918. ACM, 2012.

[30] T. R. B. of the National Academies. Default values for highway capacity and level of service analyses. 2008.

[31] V. D. of Transportation. Statewide park & ride program best practices guide. Technical report, 2013.

[32] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*, 2013.

[33] J. Petzold, F. Bagci, W. Trumler, and T. Ungerer. Comparison of different methods for next location prediction. In *Euro-Par 2006 Parallel Processing*, pages 909–918. Springer, 2006.

[34] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke. A clustering approach for collaborative filtering recommendation using social network analysis. *J. UCS*, 2011.

[35] D. Quercia, N. Lathia, F. Calabrese, G. D. 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, Dec 2010.

[36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Technical report, DTIC Document, 1985.

[37] G. Santos. Urban congestion charging: a comparison between london and singapore. *Transport Reviews*, 25(5):511–534, 2005.

[38] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[39] H. Wang, M. Terrovitis, and N. Mamoulis. Location recommendation in location-based social networks using user check-in data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 374–383. ACM, 2013.

[40] J. G. Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.

[41] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 325–334. ACM, 2011.

[42] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 363–372. ACM, 2013.

[43] J.-D. Zhang and C.-Y. Chow. igslr: Personalized geo-social location recommendation: A kernel density estimation approach. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, pages 334–343, New York, NY, USA, 2013. ACM.