Personalized Trip Planning Considering User Preferences
and Environmental Variables with Uncertainty

Mingu KIM\textsuperscript{†+a), Student Member, Seungwoo HONG\textsuperscript{†+b), and Il Hong SUH\textsuperscript{†c), Nonmembers}

SUMMARY

Personalized trip planning is a challenging problem given that places of interest should be selected according to user preferences and sequentially arranged while satisfying various constraints. In this study, we aimed to model various uncertain aspects that should be considered during trip planning and efficiently generate personalized plans that maximize user satisfaction based on preferences and constraints. Specifically, we propose a probabilistic itinerary evaluation model based on a hybrid temporal Bayesian network that determines suitable itineraries considering preferences, constraints, and uncertain environmental variables. The model retrieves the sum of time-weighted user satisfaction, and ant colony optimization generates the trip plan that maximizes the objective function. First, the optimization algorithm generates candidate itineraries and evaluates them using the proposed model. Then, we improve candidate itineraries based on the evaluation results of previous itineraries. To validate the proposed trip planning approach, we conducted an extensive user study by asking participants to choose their preferred trip plans from options created by a human planner and our approach. The results show that our approach provides human-like trip plans, as participants selected our generated plans in 57% of the pairs. We also evaluated the efficiency of the employed ant colony optimization algorithm for trip planning by performance comparisons with other optimization methods.

key words: recommender system, personalization, trip planning, probabilistic modeling and ant colony optimization

1. Introduction

Traveling is among the most preferred leisure activities worldwide, and the number of travelers has been steadily increasing by the improving living standards\cite{1}, \cite{2}. Besides its continued growth, tourism has experienced a notable optimization algorithm for trip planning by performance comparisons with other optimization methods.

In this section, we review previous approaches related to travel recommendation systems. Section 3 formalizes the probabilistic travel planning problem. Section 4 details the proposed probabilistic itinerary evaluation model and its inference. Section 5 shows the adoption of ACO to generate the trip plan that maximizes user satisfaction. Finally, we present comprehensive experimental results and draw conclusions in Sect. 6 and 7, respectively.

2. Related Work

The interest to investigate travel recommendation approaches has been increasing with the growth of tourism. In this section, we review previous studies related to trip planning and our approach. The results show that our approach provides human-like trip plans, as participants selected our generated plans in 57% of the pairs. We also evaluated the efficiency of the employed ant colony optimization algorithm for trip planning by performance comparisons with other optimization methods.

In this paper, we propose an effective approach for trip planning that maximizes user satisfaction based on personal preferences while considering various constraints and uncertainty in environmental variables. The proposed approach uses a probabilistic graphical model to represent the satisfaction and constraints as random variables comprising user preferences and environmental variables, such as weather and traffic conditions, and allows to evaluate an itinerary in terms of the sum of time-weighted satisfaction. Then, we adopt ant colony optimization (ACO) to efficiently find the itinerary that maximizes the objective function while satisfying the constraints.

The rest of this paper is organized as follows. In Sect. 2, we review previous approaches related to travel recommendation systems. Section 3 formalizes the probabilistic travel planning problem. Section 4 details the proposed probabilistic itinerary evaluation model and its inference. Section 5 shows the adoption of ACO to generate the trip plan that maximizes user satisfaction. Finally, we present comprehensive experimental results and draw conclusions in Sects. 6 and 7, respectively.
POI recommendation system is a branch of recommendation systems that suggests a new POI to be visited by a tourist. It is closely related to travel itinerary recommendation and provides great intuition for the trip planning problem in terms of POI evaluation. It has been addressed by Feng et al. [4], who proposed an approach to select POIs using personalized ranking metric embedding based on user preferences and geographical influences. In [5], Zhao et al. present spatiotemporal latent ranking for time-aware successive POI recommendation based on ranking-based pairwise tensor factorization. Yin et al. [6] used a spatiotemporal latent Dirichlet allocation model to learn region-dependent personal interests and proposed an attribute pruning algorithm for POI recommendation. Yao et al. [7] developed a POI recommendation system based on temporal matching between users and POIs to incorporate temporal popularity of POIs and user regularities. In [8], Han et al. focused on the geographical diversification of recommended POIs and proposed a method that recommends a variety of POIs located in a user's activity district. Yang et al. [9] proposed a framework incorporating semi-supervised learning and collaborative filtering for POI recommendation to overcome data scarcity. In [10], the authors proposed a spatial-aware hierarchical collaborative deep learning model to learn spatial-aware personal preferences. They also developed methods for social regularization and spatial smoothing to overcome data sparsity, thus establishing an approach to provide POI recommendation. Zhao et al. [11] utilized hierarchical structures of both users and POIs using hierarchical geographical matrix factorization for POI recommendation. Wang et al. [12] proposed a framework for POI recommendation that incorporates users' historical check-ins and varied auxiliary information, such as geographical influences among POIs and social influences among users. Cai et al. [13] presented a feature-space separated factorization model to make POI recommendation, where the model represents multiple features of a POI in a separate latent space.

Although these works are closely related to trip planning problem, several differences exist between POI and itinerary recommendation. The most notable difference is the inability to evaluate trip plans from individual POI evaluation, because the plan can be affected by the order of the POIs in the itinerary. To address this issue, many researchers have proposed specific approaches for trip planning. In [14], De Choudhury et al. proposed an itinerary recommendation system based on photo streams from a media hosting service. They used the streams to construct a POI graph using visit counts, visit times, and travel time between POIs, and a recursive greedy algorithm was used to optimize the itinerary on the graph. Gionis et al. [15] developed tour recommendation with a specific sequence of POI categories with two different itinerary evaluation measures: additive- and coverage-based user satisfaction. Bolzoni et al. [16] proposed a cluster itinerary planning algorithm for tour recommendation with the constraint of POI category visit counts and presented two POI selection methods for exact and greedy solutions. In [17], Cheng et al. investigated information from user profiles including gender, age, and race, and adopted a Bayesian framework to represent the sequential correlation between POIs and this information. Then, Chen and Cheng [18] extended this work by considering the size of groups of people traveling together retrieved from face detection. In [19]–[21], Brilhante et al. developed a personalized itinerary planning method and formulated the trip planning problem as an instance of the generalized maximum coverage problem. The method comprises two steps for selection of sub-itineraries and combination to maximize users' personal interests. Lim et al. [22], [23] proposed an algorithm called PersTour to personalize tour recommendations considering user preferences and POI popularity, which were determined using geotagged photos from the Flickr database.

These previous works on trip recommendation investigate various factors for itinerary evaluation such as user preferences, user attributes, size of traveling group, POI popularity, sequence of POI categories, and POI category visit counts. However, they did not consider environmental uncertainties, such as opening and closing times of POIs, weather, and traffic, which are crucial to realize the recommended trip plan. To the best of our knowledge, few studies have addressed trip recommendation considering environmental uncertainty. Chen et al. [24] considered traffic conditions along with personal preferences and time budget constraints for trip planning. They collected taxi GPS footprints to obtain data for time-varying transit time modeling and constructed a dynamic POI network model on the collected GPS data, popularity and geographical location of the POIs. In terms of itinerary planning, they proposed a two-phase approach consisting of route search and route augmentation. Similarly, Gavalas et al. [25] considered time-dependent travel times among POIs as in [24].

The approaches in [24] and [25] consider time-varying travel times between two POIs but simplify modeling to travel times between POIs according to different departure times to build a POI network considering the time of user's request. In contrast, our approach directly models uncertainties using a probabilistic framework, thus enabling itinerary evaluation based on uncertainties in several aspects including travel time. Moreover, we consider variable travel times depending on the user. For example, an inexperienced driver may be delayed when arriving at a POI, and these delays will accumulate until the last POI, to which the traveler might arrive much later than scheduled. Therefore, trip planning should prepare an itinerary considering worst-case scenarios to satisfy constraints. In [26], Zhang et al. studied tour recommendation considering POI availability and uncertain traveling times represented as random variables following some probability distributions. This approach is similar to our proposed model, but our probabilistic itinerary evaluation model expands modeling capabilities by including various uncertain factors, such as weather conditions, visit duration, and costs.
3. Problem Formulation

Trip planning aims to generate an itinerary that maximizes the expected user satisfaction while satisfying various constraints. The satisfaction is affected by user preferences and various environmental variables such as weather conditions and visit duration to POIs. The environmental variables have probabilistic uncertainties, and hence it is convenient to represent both user satisfaction and constraints stochastically. More formally, we define the trip planning problem as a stochastic constrained optimization problem.

**Definition 1 (POIs):** Let $\mathbb{X} = \{x^1, x^2, \ldots, x^{|\mathbb{X}|}\}$ be a set of POIs, where $|\mathbb{X}|$ denotes the cardinality of set $\mathbb{X}$. A POI $x^i$ is defined by sextuple $(D_{x^i}, C_{x^i}, sc_{x^i}, pc_{x^i}, ot_{x^i}, ct_{x^i})$, where $D_{x^i}$ is the visit duration, $C_{x^i}$ is the cost of the visits, $sc_{x^i} \in [0, 5]$ is the POI rating provided by people who have visited it, $pc_{x^i} \in PC$ is the POI category, $ot_{x^i}$ is the opening time, and $ct_{x^i}$ is the closing time. Visit duration $D_{x^i}$ is a continuous random variable dependent on the traveler. Likewise, cost of the visits $C_{x^i}$ is also a continuous random variable to represent the varying range of prices from POIs such as restaurants, hotels, and tourist attractions.

In this paper, we denote random variables with uppercase letters and their instantiations with lowercase letters.

**Definition 2 (Trip Question):** Trip question $q$ is defined by decuple $(x_s, x_f, t_s, t_f, b, x_+, x_-, PC_+, PC_-, U)$, where $x_s$ and $x_f$ are the initial and final POIs on the itinerary, respectively, $t_s$ and $t_f$ are the initial and final times of the schedule, respectively, $b$ is the total budget for the trip, $x_+$ and $x_-$ are the sets of POIs that the itinerary should include and exclude, respectively, $PC_+$ and $PC_-$ are sets of POI categories that the itinerary should include and exclude, respectively, and $U$ is the set of user preferences about POI categories. User preference $u_i \in U$ represents the predilection for POI category $pc_{x^i}$, whose default value is 0.5 and $u_i \in [0, 1]$.

**Definition 3 (Itinerary):** Itinerary $\mathbb{I} = \langle x_0, x_1, \ldots, x_{|\mathbb{I}|} \rangle$ is defined by an ordered set of POIs, whose elements belong to POI set $\mathbb{X}$, $x_0$ and $x_{|\mathbb{I}|}$ should be equal to $x_s$ and $x_f$, respectively, and the itinerary cannot contain repeated POIs. As the set is ordered, two itineraries with the same POI set but different orders are different.

**Definition 4 (Probabilistic Trip Planning Problem):** Let $S_{x^i} \in [0, 5]$ be a continuous random variable representing the user satisfaction in POI $x_i$, where $S_{x^i}$ depends on the target POI and various other factors, such as visit duration, weather conditions, and user preferences. The probabilistic trip planning problem aims to find the best itinerary from all feasible itineraries based on trip question $q$ and is defined as the following stochastic constrained optimization problem:

$$ P(T_i \geq ot_{x^i}[j], q) > 1 - \epsilon, \quad \text{for all } 1 < i < |\mathbb{I}|, $$

subject to

$$ P(T_i \leq ct_{x^i}[j], q) > 1 - \epsilon, \quad \text{for all } 1 < i < |\mathbb{I}|, $$

$$ P(T_i \leq t_f[\mathbb{I}], q) > 1 - \epsilon, $$

$$ P\left(\sum_{i=1}^{\mathbb{I}} S_{x^i} \leq b[\mathbb{I}], q\right) > 1 - \epsilon, $$

where $\epsilon$ is a small positive real number, $T_i$ is the random variable representing the time of user arrival to POI $x_i$, $T_i$ is the random variable representing the time of user departure from POI $x_i$, and $PC_{\mathbb{I}} = \{pc_{x^i} : x_i \in \mathbb{I}\}$ is the set of POI categories in itinerary $\mathbb{I}$.

Objective function (1) maximizes the sum of expected user satisfaction $E[S_{x^i}[\mathbb{I}], q]$ weighted by expected visit duration $E[D_{x^i}[\mathbb{I}], q]$ given itinerary $\mathbb{I}$ and trip question $q$. Without the weighting term, the planner would be biased towards trip plans with POIs retrieving the shortest visit duration, because the sum of user satisfaction measures would be higher by visiting more POIs within the limited travel time. For example, a trip plan consisting of four 2-point POIs will be worth more than that consisting of one 5-point POI. Therefore, it is reasonable to weight user satisfaction by the visit.

### Table 1

| Variable | Description |
|----------|-------------|
| $\mathbb{X}$ | Set of POIs |
| $x^i$ | POI |
| $D_{x^i}$ | Visit duration of POI $x^i$ |
| $C_{x^i}$ | Cost of the visit of POI $x^i$ |
| $sc_{x^i}$ | Rating of POI $x^i$ |
| $pc_{x^i}$ | Category of POI $x^i$ |
| $ot_{x^i}$ | Opening time of POI $x^i$ |
| $ct_{x^i}$ | Closing time of POI $x^i$ |
| $q$ | Trip question |
| $x_s$ | Initial POI |
| $x_f$ | Final POI |
| $t_s$ | Initial time |
| $t_f$ | Final time |
| $b$ | Total budget |
| $x_+$ | Set of POIs should be included |
| $x_-$ | Set of POIs should be excluded |
| $PC_+$ | Set of POI categories should be included |
| $PC_-$ | Set of POI categories should be excluded |
| $U$ | Set of user preferences |
| $u_i$ | User preference for POI category $pc_{x^i}$ |
| $\mathbb{I}$ | Trip itinerary |
| $S_{x^i}$ | User satisfaction in POI $x_i$ |
| $T_i$ | Time of user arrival to POI $x_i$ |
| $T^d_i$ | Time of user departure from POI $x_i$ |
| $PC_{\mathbb{I}}$ | Set of POI categories in itinerary $\mathbb{I}$ |
| $M_i$ | Travel time between POIs $x_{i-1}$ and $x_i$ |
| $W_i$ | Weather suitability of POI $x_i$ at time $T_i$ |
| $TE_i$ | Temperature of POI $x_i$ at time $T_i$ |
| $P(T^d_i \leq ct_{x^i}[\mathbb{I}], q) > 1 - \epsilon, \quad \text{for all } 1 < i < |\mathbb{I}|, $ |
| $P(T^d_i \leq t_f[\mathbb{I}], q) > 1 - \epsilon, $ |
| $P\left(\sum_{i=1}^{\mathbb{I}} S_{x^i} \leq b[\mathbb{I}], q\right) > 1 - \epsilon, $ |
duration.

Constraints (2) and (3) imply that the arrival to and departure from a POI should be after its opening time and before its closing time, respectively. In addition, constraint (4) guarantees that the itinerary is completed before end time \( t_e \), and constraint (5) guarantees that the total cost of the itinerary activities remains below budget limit \( b \). Finally, (6) and (7) represent constraints on the itinerary and POI categories, respectively.

In (1–7), we use conditional expectations and probabilities to represent the objective function and constraints. Therefore, to solve this probabilistic trip planning problem, we need a model to infer probability distributions \( P(S_x; \beta, q) \), \( P(D_x; \beta, q) \), \( P(T_i \geq ot_x; \beta, q) \), \( P(T_i^d \leq ct_x; \beta, q) \), \( P(T_i \leq t_e; \beta, q) \), and \( P(\sum_{i=1}^{n} C_x \leq b; \beta, q) \). We adopt a hybrid temporal Bayesian network (BN) [27] that can represent both causal and temporal stochastic relations for determining the solution.

4. Probabilistic Itinerary Evaluation Model

4.1 Hybrid Temporal BN

Theoretically, if we have a joint probability distribution of all relevant random variables, then any marginal, joint, or conditional probability from a subset of those random variables can be directly inferred. In practice, however, explicitly obtaining the joint distribution is difficult for a system containing several random variables for many reasons. First, the number of probabilities to be stored increases exponentially with the number of random variables, thus being memory expensive to store this information. Moreover, it is computationally expensive to manipulate the several random variables, and the data required to obtain the probabilities will be also prohibitive.

Alternatively, a BN provides a compact representation of the joint probability distribution based on a set of conditional independence assumptions of the underlying random variables. These assumptions are represented as a directed acyclic graph, whose nodes correspond to the random variables and the edges describe the influence from origin to destination node. Therefore, a BN encodes local independence as each variable is assumed to be conditionally independent of its non-descendant nodes given its parent nodes, and it factorizes the joint distribution from the network topology and conditional probability distributions as follows:

\[
V_i \perp \text{NonDescendants}_V_i | Pa_{V_i}^G \quad \text{for all } V_i, \tag{8}
\]

\[
P(V_1, \ldots, V_n) = \prod_{i=1}^{n} P(V_i | Pa_{V_i}^G), \tag{9}
\]

where \( G \) is a BN graph over nodes (variables) \( V_1, \ldots, V_n \) and \( Pa_{V_i}^G \) denotes the parents of node \( V_i \) in \( G \).

We use a BN that admits discrete and continuous random variables, thus establishing a hybrid BN. In addition, our BN can be considered as dynamic [28], [29] because it represents sequential events. The most notable difference between our hybrid dynamic BN and a conventional dynamic BN is the explicit representation of time as a random variable in the proposed BN, which simultaneously enables the inference of both events in the continuous time domain and their time of occurrence. Therefore, we call the proposed network a hybrid temporal BN.

Figure 1 illustrates the proposed probabilistic itinerary evaluation model based on the hybrid temporal BN. Figure 1a shows a part of the model comprising the relations of variables involved in inference of time, where arrival time \( T_i \) is directly influenced by its preceding arrival time, \( T_{i-1} \), the visit duration to the previous POI, \( D_{i-1} \), and the travel time, \( M_i \), between POIs \( x_{i-1} \) and \( x_i \). In addition, given that travel time \( M_i \) usually depends on the departure time from previous POI \( x_{i-1} \) and \( x_i \), time of arrival \( T_{i-1} \), and visit duration \( D_{i-1} \) to the previous POI. Finally, visit duration \( D_i \) is affected only by POI \( x_i \). One might argue that the duration should be also influenced by the previous travel time, whose high length may induce the traveler to shorten the visit. However, considering every influence among variables would unnecessarily complicate the model, and consequently the inference and acquisition of model parameters may become intractable. Therefore, relations that we consider as negligible for model performance are assumed as conditional independence. Figure 1b describes influences among variables related to user satisfaction \( S_x \). Satisfaction is directly affected by three variables, namely, weather suitability \( W_i \), user preference \( u_x \), and average score of POI \( sc_x \). Weather suitability \( W_i \) is a binary random variable that indicates whether a POI has favorable weather conditions at time \( T_i \). In turn, weather conditions are represented by temperature \( TE_i \) and probability of precipitation \( R_i \).

4.2 Inference on Probabilistic Itinerary Evaluation Model

To represent the joint distribution and infer relevant probabilities for itinerary evaluation, each node must be associated with a corresponding conditional distribution, which can be of one from these 12 types: \( x_i, D_x, C_x, M_i, T_i, TE_i, R_i, W_i, pc_x, sc_x, u_x \), and \( S_x \). However, POI \( x_i \) is given by itinerary \( i \), and \( pc_x \), \( sc_x \), and \( u_x \) are deterministic functions of \( x_i \) given \( i \) and \( q \). Therefore, we only need the conditional probabilities of \( D_x, C_x, M_i, T_i, TE_i, R_i, W_i \), and \( S_x \) given their parent variables.

First, we gather candidate POIs and their associated data for \( pc_x, ct_x, ot_x, sc_x, D_x, \) and \( C_x \) by web crawling various travel information services. We model the conditional distributions of visit duration \( D_x \) and cost \( C_x \) as Gaussian distributions using the estimated means (\( \mu_d \) and \( \mu_c \), respectively) and variances (\( (\sigma_d^2) \) and \( (\sigma_c^2) \), respectively) from the collected data. The values of other variables are deterministically obtained from POI \( x_i \). For travel time \( M_i \), we gather real-time navigation information for all combinations of POI pairs every hour from the Daum map ser-
given by the Dirac delta function:

To infer arrival time $T_i$, we calculate the probability distribution of the first group of random variables, $M_i$ and $T_i$. Then, as $t_i$, $x_i$, and $x_j$ are given by trip question $q$ and itinerary $I$, the distributions can be computed as

$$P(M_i[I], q) = P(M_i[x_i, x_j, t_i]) = N(M_i[I][x_i, x_j, t_i]),$$

$$P(T_i[I], q) = \int P(T_i[m_i, t_i]P(m_i|x_i, x_j, t_i)dm_i = N(T_i[I][x_i, x_j, t_i]),$$

where the conditional probability density function of $T_i$ is given by the Dirac delta function:

$$P(T_i[m_i, d_{x_i}, t_{i-1}] = \delta(T_i - m_i - d_{x_i} - t_{i-1}).$$

Then, the $i$ + 1-th conditional distributions, $P(M_{i+1[I], q}$ and $P(T_{i+1[I], q}$, can be computed by using the $i$-th distributions as

$$P(M_{i+1[I], q} = \int \int P(M_{i+1[x_i, x_{i+1}, d_{x_i}, t_i]P(d_{x_i}][I], q)$$

$$P(T_{i+1[I], q} = \int \int P(T_{i+1}[m_{i+1}, x_i, x_j, t_i]P(m_{i+1}][I], q)$$

where conditional density function $P(d_{x_i}[I], q)$ is simply $P(d_{x_i}[x_i])$ according to the BN local independence assumption. Departure time $T_i$ is given by the sum of two conditionally independent variables, $T_i$ and $D_{x_i}$, and its probability density function can be calculated as the convolution of two distributions [31], $P(T_i[I], q)$ and $P(D_{x_i}[I], q)$:

$$P(T_i[I], q) = \int P(T_i[T_i]P(D_{x_i}[z[I], q])dz,$$

where $P(T_i[I], q)$ and $P(D_{x_i}[I], q)$ denote conditional distributions of arrival time $P(T_i[I], q)$ and $P(D_{x_i}[I], q)$, respectively.

For weather suitability, we obtained the temperatures and probabilities of precipitation over time using the open API of the Korea Meteorological Administration. Then, the conditional probability distribution of weather suitability is
a deterministic function of its parents, such that

\[
P(W_i|e_i, r_i, p_{c_x}) = \begin{cases} 
0 & \text{if } te_i > 35^\circ \text{C or } r_i = \text{true}, \\
p_{c_x} \notin PC_{\text{indoor}} & , \\
1 & \text{otherwise},
\end{cases}
\]

where \(PC_{\text{indoor}} \) is the set of indoor POI categories. Based on this local conditional distribution model, the marginal conditional distribution of weather suitability can be computed as

\[
P(W_i|\mathcal{J}, q) = \sum_{i} \int P(W_i|e_i, r_i, p_{c_x})P(r_i|t_i)P(te_i|t_i)P(t_i|\mathcal{J}, q) \, dt_i
\]

where \(t_i \) is the set of indoor POI categories. Based on this local conditional distribution model, the marginal conditional distribution of weather suitability can be computed as

\[
P(W_i|\mathcal{J}, q) = \sum_{i} \int P(W_i|e_i, r_i, p_{c_x})P(r_i|t_i)P(te_i|t_i)P(t_i|\mathcal{J}, q) \, dt_i
\]

Finally, the conditional density function of user satisfaction given its parents, \(P(S_x|w_i, u_x, sc_x)\), is given by Dirac delta function \(\delta(S_x - w_i \, u_x \, sc_x)\). Hence, user satisfaction is zero if POI \(x_i\) is not suitable under weather \(w_i\) at arrival time; otherwise, it is equal to overall score \(sc_x\) of the POI weighted by user preference \(u_x\).

Given these conditional distributions, we can estimate the conditional probabilities needed for itinerary evaluation:

\[
P(S_x|\mathcal{J}, q) = \sum_{i} P(S_x|w_i, u_x, sc_x)P(w_i|\mathcal{J}, q),
\]

\[
P(D_x|\mathcal{J}, q) = P(D_x|x_i) = N(D_x|\mu_{x_i}^d, \sigma_{x_i}^2),
\]

\[
P(T_i \geq ot_x|\mathcal{J}, q) = \int_{ot_x}^{\infty} P(t_i|\mathcal{J}, q) \, dt_i,
\]

\[
P(T_i^d \leq ct_x|\mathcal{J}, q) = \int_{-\infty}^{ct_x} P(t_i^d|\mathcal{J}, q) \, dt_i^d,
\]

\[
P(T_i|\mathcal{J}, q) = \int_{-\infty}^{\infty} P(t_i|\mathcal{J}, q) \, dt_i,
\]

\[
P(\sum_{i=1}^{\mathcal{J}} c_i \leq b|\mathcal{J}, q) = \int_{-\infty}^{b} \mathcal{N}(c | \sum_{i=1}^{\mathcal{J}} \mu_x^c, \sum_{i=1}^{\mathcal{J}} (\sigma_x^c)^2) \, dc,
\]

where \(c\) is the sum of all costs.

Given an itinerary and a trip question, we can evaluate the itinerary and verify whether it satisfies the given constraints with the inferred probabilities in (18)–(23). A naive way to find the optimal itinerary would be to exhaustively evaluate all possible itineraries using the probabilistic itinerary evaluation model and retrieving the itinerary with the highest objective function value. However, the number of possible itineraries exponentially increases with the trip duration and other factors, making the evaluation computationally intractable. Moreover, the orienteering problem, which resembles our trip planning problem, is a well-known NP-hard problem that has no polynomial-time algorithm retrieving its optimal solution. In fact, it took more than 3 days of computation when we evaluated all possible itineraries from 09:00 to 22:00 considering 153 POIs, although we did not use our model but only overall POI scores. Therefore, it is essential to have an efficient method to timely generate suitable candidate itineraries, as we detail in the sequel.

5. Ant Colony Optimization over Itineraries

In this section, we describe an itinerary optimization method that incorporates the probabilistic itinerary evaluation model and ACO [32]. ACO is a metaheuristic algorithm that provides near-optimal solutions using limited computation resources by iteratively simulating several artificial ants deployed on a graph representing the problem to be solved.

Figure 2 illustrates a trip planning problem, where the graph consists of nodes representing POIs and edges between all node pairs. Each artificial ant builds a candidate solution by traveling between nodes while satisfying the constraints in (2)–(7). At each ant motion between two POIs, the ant stochastically chooses the next node according to its attractiveness and the “pheromones” deposited by the other ants that previously selected that path. The amount of pheromone deposited by an ant depends on the quality of the itinerary determined from our itinerary evaluation model, and it guides subsequent ants to promising areas of the search space for improving the candidate itineraries. In this study, we used the MAX–MIN ant system [33], which is an ACO variant, as it is a relatively simple algorithm to implement.

To implement ACO, the attractiveness of the next POIs, \(\eta_{i_{x+1}, \mathcal{J}, q}\), given itinerary \(\mathcal{J}\), to the current POI should be modeled for providing a greedy local guidance for selecting next POI \(x_{i+1}\). We incorporate the proposed probabilistic itinerary evaluation model to compute the attractiveness values as follows:

\[
\eta_{i_{x+1}, \mathcal{J}, q} = \frac{E[s_{x_{i+1}}|\mathcal{J}_{i+1}, q]E[d_{x_{i+1}}|\mathcal{J}_{i+1}, q]}{E[m_{i+1}|\mathcal{J}_{i+1}, q]},
\]

where the attractiveness values are set to zero for POIs that do not satisfy the constraints in (2)–(7). In addition, we di-
vide the time-weighted user satisfaction by the travel time to inform each ant that a long travel time may adversely influence the itinerary evaluation.

Using these attractiveness values and the initial value of pheromone, \( \tau_{x_i \rightarrow x_{i+1}} = \tau_0 \), an artificial ant selects the next POI according to the selection probability:

\[
P(Sel(x_{i+1})|q) = \frac{\tau^{\alpha}_{x_i \rightarrow x_{i+1}}[\eta]_q^\beta}{\sum_{x_{i+1} \in X} \tau^{\alpha}_{x_i \rightarrow x_{i+1}}[\eta]_q^\beta},
\]

where parameters \( \alpha \) and \( \beta \) determine the influence of attractiveness \( \eta \) and pheromone \( \tau \). Equation (25) guarantees that no POI violating any constraint can be involved in the itinerary, because such POIs have zero attractiveness and therefore zero probability of being selected.

The ant repeatedly selects POIs until no more POIs satisfying the constraints remain. In most cases, the travel time budget constraint in (4) eliminates all POIs from the selectable set. When a set of ants has built candidate itineraries, pheromone trails \( \tau_{x_i \rightarrow x_{i+1}} \) are updated at each iteration as

\[
\tau_{x_i \rightarrow x_{i+1}} \leftarrow (1 - \rho) \cdot \tau_{x_i \rightarrow x_{i+1}} + \Delta \tau_{x_i \rightarrow x_{i+1}}^{\text{best}},
\]

where \( \rho \) is the pheromone evaporation coefficient, \( \tau_{\text{max}} \) and \( \tau_{\text{min}} \) are the upper and lower bounds of the pheromone trails, respectively, and \( \Delta \tau_{x_i \rightarrow x_{i+1}}^{\text{best}} \) is the amount of pheromone deposited at the iteration. The amount of newly deposited pheromone, \( \Delta \tau_{x_i \rightarrow x_{i+1}} \), is determined from the best itinerary at that iteration retrieved using objective function (1). While the attractiveness provides a greedy local guidance for selection of the next POI, the pheromone offers a global guidance in terms of completed itineraries. Therefore, as ACO iteratively generates candidate itineraries and updates pheromone trails, more pheromones are deposited on promising edges between POIs. This way, the pheromone reflects the accumulated experience of the ant colony during the stochastic exploration of the itinerary search space.

During the early iterations of ACO, it generates relatively greedy solutions. However, as the algorithm gradually improves the itineraries based on the pheromone global guidance, it converges to an itinerary with sufficiently high objective function value. Another notable merit of ACO is its nature of anytime algorithm, which preserves the ability to retrieve a valid solution even if the algorithm stops before convergence. Thus, the tradeoff between computation time and itinerary quality can be easily adjusted.

6. Experiments and Results

We verified the ability of our approach to efficiently generate personalized trip plans considering user preferences and environmental uncertainties while satisfying the given constraints. Note that it is generally difficult to compare the performance of different trip planning methods for several reasons. For instance, different recommendation algorithms have been developed focused on different factors related to the trip planning problem. In addition, there is no publicly available dataset for fair comparison among methods. Therefore, we conducted an extensive user study to validate our probabilistic itinerary evaluation model by comparing pairs of trip plans generated by our approach with those created by a human planner. Then, to demonstrate the effectiveness and efficiency of the proposed ACO-based itineraries, we compared them with those retrieved by four optimization methods: random selection (RAND), greedy nearest selection (GREEDYNEAREST), greedy highest satisfaction selection (GREEDYHIGHEST), and brute force selection with time limit (BRUTEFORCE).

6.1 Comparison with Human-Created Trip Plans

To validate our probabilistic itinerary evaluation model, we conducted a survey with 31 participants and asked them to choose their preferred trip plans from pairs of those generated by our approach and those created by a human planner. The participants were not aware of the source of the plans. For this experiment, we gathered 293 POIs from Jeju Island, one of the most famous tourist destinations in Korea, by web crawling travel information services. From the POIs, 153 were tourist attractions belonging to 36 categories, and the other POIs comprised 108 restaurants, 29 accommodations, 2 ferry terminals, and 1 airport. We also collected 30 human-created trip plans consisting of 72 daily plans and extracted trip questions from each plan. We used the same initial and final POIs, initial and final times of the schedule, and total cost if available. Then, the proposed approach generated trip plans based on the extracted questions.

Figure 3 illustrates the questionnaire adopted in this user study and shows side-by-side the trip plan created by the human planner (left panel) and that generated by our approach (right panel). For the example in the figure, our approach generated a trip plan following the trip question extracted from the human-created plan, where the initial and final POIs are the Jeju International Airport and the Sunbeach Hotel and Resort, respectively. The initial and final times are 09:30 and 20:00, respectively, and the total budget for the trip is KRW$35,000. In this example, we included an additional constraint that allows only restaurant POIs to be visited at lunch and dinner times when optimizing the itinerary.

From the 72 pairs of trip plans and 31 survey participants considered in this study, we obtained 2,232 survey results. The participants chose our travel plans 57% of the time, showing that our approach can generate human-like trip plans. Besides the survey comparing pairs of plans, we asked the participants what criteria they used to choose their preferred travel plan. Twenty-three participants (approximately 74%) stated that the most essential criterion for preferring a trip plan was the shortest travel time between POIs. Although it is difficult to conclude that this is the best criterion because there were not many participants in this study, these opinions are consistent with our consideration that the POI visitation order is crucial for trip planning.
6.2 Comparison among Itinerary Optimization Methods

We also compared itineraries generated by the proposed ACO-based approach with those generated using the other optimization methods mentioned above to verify the efficiency and itinerary quality of our approach. To generate itineraries using these methods, we considered the same set of POIs and trip questions from the previous experiment. RAND is the naivest method, as it randomly selects an itinerary comprising POIs satisfying the constraints. GREEDYNEAREST and GREEDYHIGHEST greedily select sequential POIs based on the shortest expected travel time and highest expected user satisfaction. BruteForce randomly selects an itinerary as RAND, but it repeatedly generates itineraries during a predefined period, evaluates the generated itineraries and finally selects the best itinerary among them. As mentioned above, we attempted to evaluate the method that retrieves the optimal itinerary by generating and evaluating every possible itinerary, but it was highly computationally expensive as it took more than 3 days to retrieve the best itinerary in the worst case. Hence, we used BruteForce with time limit of 10 min as an alternative to the complete exhaustive search. We also limited the execution time for the proposed ACO algorithm to 2, 3, and 4 s to verify the computation time effect on the algorithm performance.

Figure 4 shows the results from these experiments. We performed the experiments 100 times for each trip question and optimization method, 50,400 times in total. Then, the results were compared in terms of average time-weighted user satisfaction normalized by the total time to complete the itinerary, because each trip question has different total travel times. The results show that the proposed ACO algorithm generates trip plans with high objective function values. Although the value was smaller than that of BruteForce, the proposed ACO algorithm took approximately one-hundredth of the computation time required by the BruteForce.

In terms of execution time and performance, the proposed ACO algorithm generated travel plans with higher objective function value and narrower standard deviation as the execution time increased. Thus, as we mentioned in Sect. 5, the tradeoff between computation time of the proposed ACO algorithm and itinerary quality can be adjusted for prioritiz-
ing any of these aspects.

In addition, we also compared the proposed method with the method excluding weather suitability (ExWe) to show the effectiveness of considering environmental uncertainty. Figure 4 shows that the trip planner using the model excluding weather suitability has lower user satisfaction than the planner using the proposed model. Also, the method excluding the weather suitability has larger variances in user satisfaction according to the weather on the day of travel. On a good weather day, it has no significant difference from the proposed method, but the user satisfaction was greatly reduced when it rains or when the temperature is too high or too low.

7. Conclusions and Future Work

We propose an efficient approach to generate personalized trip plans considering user preferences and environmental factors with uncertainty while satisfying various constraints. Specifically, we propose the probabilistic itinerary evaluation model based on hybrid temporal BNs, which describe and allow to evaluate trip itineraries considering uncertainty from various relevant factors represented as random variables. Then, we propose an ACO algorithm for trip planning by incorporating the evaluation model. We have demonstrated the validity of our approach for trip recommendation through a comprehensive user study to compare our approach with trip plans created by a human planner. Furthermore, we compared the optimization of trip plans by using our approach and four methods aimed to retrieve the best itinerary and verified the effectiveness and efficiency of the proposed ACO algorithm. In future studies, we will incorporate more factors relevant to trip recommendation into our approach, such as time-varying user preferences and POI popularity. We will also explore real-time interactive trip planning based on user feedback to improve the recommended trip plans.

References

[1] D. Glaesser, J. Kester, H. Paulose, A. Alizadeh, and B. Valentin, “Global travel patterns: an overview;” J. travel medicine, vol.24, no.4, 2017.
[2] “UNWTO world tourism barometer and statistical annex, November 2018;” UNWTO World Tourism Barometer, vol.16, no.5, pp.1–36, 2018.
[3] J. Loo, “The future of travel: New consumer behavior and the technology giving it flight;” Think With Google, 2017, https://www.thinkwithgoogle.com/marketing-resources/new-consumer-travel-assistance, accessed Jan. 15. 2019.
[4] S. Feng, X. Li, Y. Zeng, G. Cong, Y.M. Chee, and Q. Yuan, “Personalized ranking metric embedding for next new poi recommendation,” IJCAI, vol.15, pp.2069–2075, 2015.
[5] S. Zhao, T. Zhao, H. Yang, M.R. Lyu, and I. King, “Stellar: Spatial-temporal latent ranking for successive point-of-interest recommendation;” AAAI, pp.315–322, 2016.
[6] H. Yin, X. Zhou, B. Cui, H. Wang, K. Zheng, and Q.V.H. Nguyen, “Adapting to user interest drift for poi recommendation;” IEEE Trans. Knowl. Data Eng., vol.28, no.10, pp.2566–2581, Oct. 2016.
[7] Z. Yao, Y. Fu, B. Liu, Y. Liu, and H. Xiong, “Poi recommendation: A temporal matching between poi popularity and user regularity;” 2016 IEEE 16th Int. Conf. Data Mining (ICDM), pp.549–558, 2016.
[8] J. Han and H. Yamana, “Geographical diversification in poi recommendation: Toward improved coverage on interested areas;” Proc. Eleventh ACM Conference on Recommender Systems, pp.224–228, 2017.
[9] C. Yang, L. Bai, C. Zhang, Q. Yuan, and J. Han, “Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation;” Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp.1245–1254, 2017.
[10] H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou, “Spatial-aware hierarchical collaborative deep learning for poi recommendation;” IEEE Trans. Knowl. Data Eng., vol.29, no.11, pp.2537–2551, Nov. 2017.
[11] P. Zhao, X. Xu, Y. Liu, Z. Zhou, K. Zheng, V.S. Sheng, and H. Xiong, “Exploiting hierarchical structures for poi recommendation;” 2017 IEEE Int. Conf. Data Mining (ICDM), pp.655–664, 2017.
[12] H. Wang, W. Ouyang, H. Shen, and X. Cheng, “Ule: Learning user and location embeddings for poi recommendation;” 2018 IEEE Third Int. Conf. Data Science in Cyberspace (DSC), pp.99–106, 2018.
[13] L. Cai, J. Xu, J. Liu, and T. Pei, “Integrating spatial and temporal...
contexts into a factorization model for poi recommendation,” Int. J. Geographical Information Science, vol.32, no.3, pp.524–546, 2018.

[14] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu, “Automatic construction of travel itineraries using social breadcrumbs,” Proc. 21st ACM conference on Hypertext and hypermedia, pp.35–44, 2010.

[15] A. Gionis, T. Lappas, K. Pelechrinis, and E. Terzi, “Customized tour recommendations in urban areas,” Proc. 7th ACM International Conference on Web search and data mining, pp.313–322, 2014.

[16] P. Bolzoni, S. Helmer, K. Wellenzoehn, J. Jamper, and P. Andrittos, “Efficient itinerary planning with category constraints,” Proc. 22nd ACM SIGSPATIAL Int. Conf. advances in geographic information systems, pp.203–212, 2014.

[17] A.J. Cheng, Y.Y. Chen, Y.T. Huang, W.H. Hsu, and H.Y.M. Liao, “Personalized travel recommendation by mining people attributes from community-contributed photos,” Proc. 19th ACM International Conference on Multimedia, pp.83–92, 2011.

[18] Y.Y. Chen, A.J. Cheng, and W.H. Hsu, “Travel recommendation by mining people attributes and travel group types from community-contributed photos,” IEEE Trans. Multimedia, vol.15, no.6, pp.1283–1295, Oct. 2013.

[19] I. Brilhante, J.A. Macedo, F.M. Nardini, R. Perego, and C. Renso, “Where shall we go today?: planning touristic tours with tripbuilder,” Proc. 22nd ACM International Conference on Information & Knowledge Management, pp.757–762, 2013.

[20] I. Brilhante, J.A. Macedo, F.M. Nardini, R. Perego, and C. Renso, “Tripbuilder: A tool for recommending sightseeing tours,” European Conference on Information Retrieval, pp.771–774, 2014.

[21] I.R. Brilhante, J.A. Macedo, F.M. Nardini, R. Perego, and C. Renso, “On planning sightseeing tours with tripbuilder,” Information Processing & Management, vol.51, no.2, pp.1–15, March 2015.

[22] K.H. Lim, J. Chan, C. Leckie, and S. Karunasekera, “Personalized tour recommendation based on user interests and points of interest visit durations,” IJCAI, vol.15, pp.1778–1784, Feb. 2015.

[23] K.H. Lim, J. Chan, C. Leckie, and S. Karunasekera, “Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency,” Knowledge and Information Systems, vol.54, no.2, pp.375–406, 2018.

[24] C. Chen, D. Zhang, B. Guo, X. Ma, G. Pan, and Z. Wu, “Triplanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints,” IEEE Trans. Intell. Transp. Syst., vol.16, no.3, pp.1259–1273, June 2015.

[25] D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, N. Vathis, and C. Zaroliagis, “The ecompass multimodal tourist tour planner,” Expert systems with Applications, vol.42, no.21, pp.7303–7316, Nov. 2015.

[26] C. Zhang, H. Liang, and K. Wang, “Trip recommendation meets real-world constraints: Poi availability, diversity, and traveling time uncertainty,” ACM Trans. Information Systems (TOIS), vol.35, no.1, Article No. 5, Oct. 2016.

[27] W.Y. Kwon and I.H. Suh, “A temporal bayesian network with application to design of a proactive robotic assistant,” Robotics and Automation (ICRA), 2012 IEEE Int. Conf., pp.3685–3690, 2012.

[28] P. Dagum, A. Galper, and E. Horvitz, “Dynamic network models for forecasting,” Proc. eighth Int. Conf. uncertainty in artificial intelligence, pp.41–48, 1992.

[29] P. Dagum, A. Galper, E. Horvitz, and A. Seiver, “Uncertain reasoning and forecasting,” Int. J. Forecasting, vol.11, no.1, pp.73–87, March 1995.

[30] “Daum web map service,” Jan. 2019.

[31] A. Papoulis and S.U. Pillai, Probability, Random Variables, and Stochastic Processes, Tata McGraw-Hill Education, 2002.

[32] M. Dorigo, M. Birattari, and T. Stutzle, “Ant colony optimization,” IEEE Comput. Intell. Mag., vol.1, no.4, pp.28–39, Nov. 2006.

[33] T. Stützle and H.H. Hoos, “Max–min ant system,” Future generation computer systems, vol.16, no.8, pp.889–914, June 2000.