A DECISION SUPPORT SYSTEM FOR TOURIST TRIP RECOMMENDATIONS INTEGRATING SOCIAL MEDIA NETWORK

PATRIZIA BERALDI[1], ANNARITA DE MAIO[1], FILOMENA OLIVITO[2], GIUSEPPE POTRINO, IMMACOLATA STRAFACE[3], ANTONIO VIOLI[4,]
[1], Department of Mechanical, Energy and Management Engineering, University of Calabria, Italy,
[2]Eurora sas Rende, Italy,
[3]Valuetech s.r.l., Rende, Italy,
[4]University of Sannio, Benevento, Italy

ABSTRACT

The rapid growth in the use of recommendation systems in the tourism sector is mainly related to the possibility to access updated data deriving from social networks, thus providing more appropriate and personalized suggestions. The paper presents a tourist trip recommendation system that suggests personalized itineraries defined as sequence of Point of Interest to visit. The system core integrates two software modules: a neural network and an optimization engine. For every pair user-PoI typology, the neural network provides, on the basis of the analysis of the social media data, a score between 0 and 1. These latter values are then used as input parameters for a routing optimization problem that suggests the itinerary by considering additional restriction, as, for example, time windows, budget and time limitations, specified by the end-user. Being a computational demanding problem, the model solution is carried out by applying a heuristic approach that is proven to provide high quality solution in a limited amount of time.

Keywords: Social media, neural network, routing problem

1 INTRODUCTION

Nowadays tourism and technology are becoming more and more interconnected. According to Google, each traveller visits, on average, 22 websites before booking a vacation, and the percentage of those who use mobile technologies to register a flight or hotel is approaching 70%. The same rapid changes that other sectors of the economy are facing have started to deeply influence the behaviour of tourists and require an evolution of the structures and capabilities of the offer. The term, borrowed by the Industry sector, to denote this new trend is Tourism 4.0. The novelty of the paradigm relies on the processing of large amount of data collected by social media networks to create personalized travel experiences. The suggested travel plans are more efficient, safer, ecological and less problematic by optimizing travel times and minimizing costs for travellers.

The aim of the paper is to present a tour planning system developed within an Italian funded R&D project to boost the tourism offer in the Southern Italy. The system exploits user's preferences learned from the social networks analysis to suggest personalized and optimized itineraries for visiting a set of Points of Interest (PoI, for short).

With millions of active users, social media platforms, such as Facebook, Twitter, Instagram and Flickr, have become potential big data sources of individual behaviour, thus creating a tremendous opportunity to gather digital traces. Analysing millions of user

footprints, it is possible to extract travel behaviour at a scale unimaginable before (Hendrik and Perdana [1]).

A valuable approach to analyse a so huge amount of data is based on the use of data-mining techniques. The basic idea is to derive for a given user a list of possible PoIs together with personalized scores without explicitly asking users about their specific tastes. The system allows overcoming the so-called "cold start" effect typical of the content-based recommendation systems (see, for example, Lee et al. [2]), and can be casted into collaborative filtering systems with recommendation process based on ratings of other users who have similar preferences (Claypool et al. [3]).

The use of data-mining techniques is not new in the tourism sector. Among the others, we cite the recent paper by Hasnat and Hasan [4] (see, also, the references therein), where a tailored framework is proposed for understanding out tourists' travel behaviours from social media data. Clustering methods are then used to determine destination choice patterns of tourists. It is worthwhile mentioning that, besides machine learning approaches, on line, social network data can be analysed and integrated into tourist recommendation system also by using non-machine learning techniques. Here we mention the very recent contribution by Persia et al. [5] who propose the use of a social sensing approach to obtain personalized PoI scores to use in the definition of personalized routing plans.

The decision support system proposed in this paper integrates neural network for the user's characterization in terms of selection of PoI categories he/she can prefer, with advanced optimization tools to generate efficient and reliable routing plans.

As for this second element is concerned, we note that the optimization engine provides the definition and solution of a routing problem with a special structure. The problem, known in the scientific literature as orienteering problem (see, for example, Golden et al. [6]) includes time windows regarding the time interval a given PoI can be visited and budget constraints that the decision maker is willing to account for. Since the identification and suggestion of the recommended plan is required to be returned in a very short time, the system is empowered with a specialized heuristic approach.

In summary, the proposed system is based on the design and integration of the following elements:

- a neural network approach for the selection of PoI categories the user can be more interested to;
- a clustering approach used to cast the set of PoIs that can be potentially interesting for the user into subsets of PoIs which are close according to some criteria;
- a tailored algorithm for generating high quality tourist routes in a limited amount of time.

The system has been developed and tested within the "ASSD – Acceleratori semantici social driven per la generazione di itinerari turistici" project (ASSD, for short), funded by Calabria region, in the South of Italy.

The rest of the paper is organized as follows. Section 2 describes the main functionalities and building blocks of the decision support system, while Section 3 is devoted to the introduction of the methodological kernel. Section 4 presents and analyses the results in terms of effectiveness and applicability on a real-life context. Conclusions and future research directions are highlighted in Section 5.

2 THE ASSD DECISION SUPPORT SYSTEM

The final output of the research project ASSD is the design and implementation of a decision support system for the definition of itineraries visiting sets of PoI that can match the tourist implicit and explicit preferences. The system, accessible via both web and mobile applications, has a high-level architecture composed with three main parts.

The client side is devoted to the interaction with the end-user, while the persistence side contains the modules providing the connection to the data warehouse for data persistence. The main methodological modules are included in the server side, which has been articulated in several components, as depicted in Fig. 1.

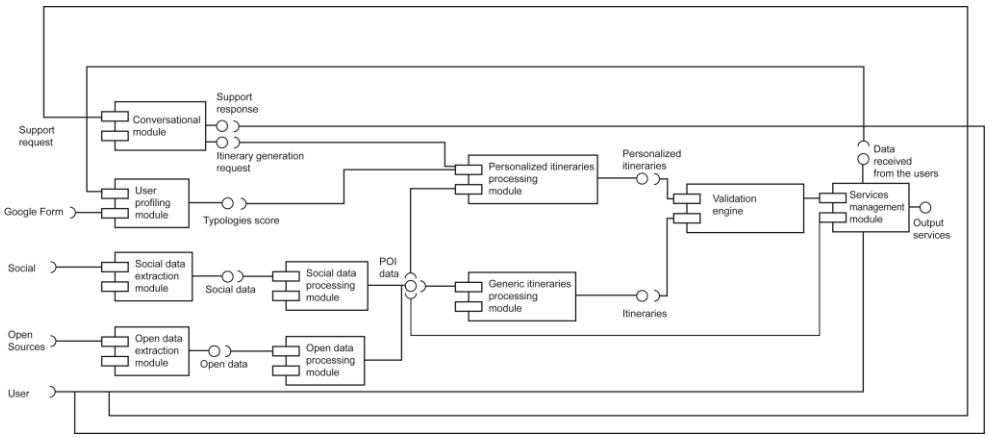

Figure 1: [Server side functional components]

Here, there are some modules for input definition, for both user and PoI characterization. First, a user-profiling module collects individual preferences and the specific parameters for a single request. Moreover, there are modules for data extraction from both open data sources and social media and for the processing of information in order to update PoI features, in terms of ranking, costs, time window and so on. These modules provide input for the "decision" modules where different tourist itineraries are defined. In particular, in addition to classic "personalized" itineraries, defined on the basis of the user request, "generic" routes, that are based on average preferences of all the users in the last month, are generated as well. Finally, a validation engine processes the solutions and a service management module transforms data into responses and coordinates the communication with the client for the output visualization.

The decision problem at the kernel of the system consists in the definition of tourist routes, i.e. sequences of PoIs to visit, that match better with the user preferences. In order to effectively model such a problem some assumptions have been made. Each PoI is assumed to belong to a specific category, while each category refers to an area of interest. For example, an art gallery belongs to category "Museum" of area of interest "Art".

The decision support process that is implemented on the system can be articulated into the following steps:

- user preferences profiling;
- PoI ranking update;
- routes definition.

As regards the first step is concerned, when the user accesses the system and sets/updates some configuration parameters and the areas of interest (e.g. Art, Cultural Heritage, Music, etc.), as shown in Fig. 2, his/her preferences are updated by using a machine learning procedure.

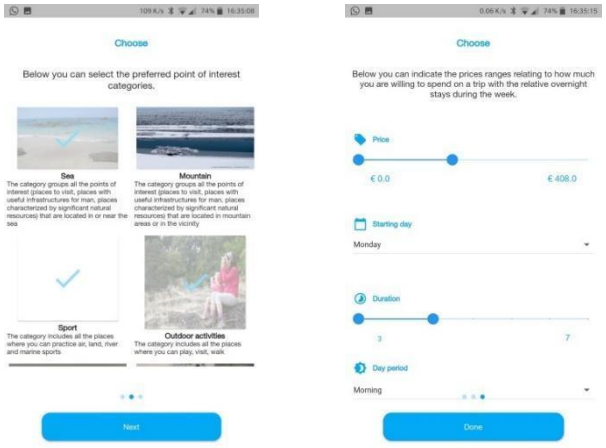

Figure 2: [User interfaces for preferences set-up]

In particular, a score in the range [0, 1] for each pair {User-PoI category} is assigned, so to help the selection for the tourist itinerary of PoIs of categories belonging to user's areas of interest. The procedure is based on a *feed-forward* neural network, whose training is periodically performed based on a database of surveys, by means of a Resilient Backpropagation training algorithm. The neural network has an input layer, which consists of the following attributes:

- residence area
- age
- gender
- educational qualification
- current job
- travel team
- journey time
- daily distance.

Then a certain number of hidden layers, which is updated each time the neural network training is performed, and one output layer with a Sigmoidal activation function.

The PoI ranking step is performed periodically by means of the analysis of the presence of the PoI on the main social media (Instagram, Facebook, Twitter and Google Plus) and on other specific web sites like TripAdvisor. A specific procedure collects and analyses the posts that refer to the PoI and considers for the previous week the following data:

- the number of posts;
- the average "sentiment" (positive, negative, neutral) expressed on the PoI;
- the level of engagement, that is the number of interactions of social media users about the posts.

By the analysis of a cumulative function of these data for all the PoI of the same category, the score of each PoI, and then the ranking within the category, is updated.

PoIs scores are used in the definition of the routes proposed to the tourist. Such a phase is further articulated into 3 steps:

1. PoIs selection, performed according to the length of the visit imposed by the user; indeed, a number *n* of PoIs with the higher score within the categories the user prefers are selected;
2. PoIs clustering where the *n* PoI are clustered into subsets that can be visited in one day according to the visit time and budget of each PoI;
3. Routes definition where, for each PoI cluster, the route that maximizes a certain score function and satisfies further constraints, like time windows for example, is defined, if any.

From a computational standpoint, this latter step is the more critical since it has to be performed within a limited amount of time. For this reason, specific heuristics procedures have been designed and implemented for both the clustering and the routing steps.

In the next Section 3 the methodological approaches adopted for these procedures are described in detail.

3 TOURIST TOUR DEFINITION

The definition of the suggested tourist tours relies on the formulation and solution of a vehicle routing problem, known in the scientific literature, as Orienteering Problem (OP) (Golden et al. [6]). The general model is used in many application domains where the selection of points to be visited represents a critical issue. In the tourist sector, the problem is also known as Tourist Trip Design Problem (TTDP).

The model is defined starting from the set of nodes that represent the PoIs, each one characterized by a score, a geographical position, an average time required for the visit and a time window, denoting the opening time interval the point can be visited.

Because of the side constraints related, for example, to limitations of time and budget, not all the PoIs could be visited and the aim is to perform a PoI selection so to maximize the total score. The solution of the mathematical problem returns an itinerary that starts from the initial PoI, visits a subset of PoIs and finally arrives at the ending PoI. The suggested tour satisfies the time windows, length, time and budget constraints. Both TTDP and OP cannot be solved in polynomial time, for this reason all existing web and mobile applications are based on the use of efficient heuristic algorithms (Gendreau et al. [7]).

In this case, we describe the algorithm considering the following features. Let *G = (V, A)* denote a complete directed graph where *V* is the set of PoIs and *A* is the set of arcs. We assume that *G* is a Euclidean graph, so the triangular inequality holds. In the set *V* index 1 corresponds to the starting point of tour and index *n* corresponds to the ending point of the tour. At each arc $(i, j) \in A$ three non-negative parameters are associated: t_{ij} denotes the traversing time, l_{ij} is the traversing distance and c_{ij} denotes the corresponding cost. Different parameters are associated with each node: R_i is the score assigned to the POI, t_i is time necessary to visit the PoI, c_i is price-ticket to visit the PoI, the time window specified by S_i, i.e. the starting time in which the POI can be visited and E_i is the ending time in which the PoI can be visited. We also consider *L* as maximum distance of the route, *B* as the maximum budget for the route and T as the maximum time for the route. The score of each PoI is pre-computed through the neural network, as described before. In the following, we refer to the formulation described in Ciancio et al. [8] and we concentrate directly to the algorithm description. The computational effort required to solve the model is affected by the number of PoIs and, as a consequence, it can be huge. In order to provide the system with a fast tool to support the users, we design a heuristic algorithm based on a clustering

engine aimed at reducing the number of potential PoIs to be evaluated. So the algorithm is divided in two main phases:

- **Phase 1 - Clustering**: in the first phase of the heuristic a clustering procedure is used to select the more convenient PoIs from the total set provided by the platform. They are chosen according to feasibility requirements related to the maximum length and budget decided by the user for his tour while maximizing the score. The clustering procedure is described in Section 3.1.
- **Phase 2 - Optimization**: in the second phase, the set composed by the POIs selected in Phase 1 becomes the input set for the model described in Ciancio et al. [8], in order to find the good quality solution and building the tour, in a short computational time.

In the routing field, the approach to *cluster before-route after* is largely used and it could be considered very effective for the solution of routing problems with different operative constraints (e.g. time windows, budget, duration, capacity, etc.). Similar approaches were used in different industrial routing problems, with application to real case study. For example, in Beraldi et al. [9] a similar approach is applied to solve the problem of a company specialised in van-sharing for heavy freight, so large instances of a pick-up and delivery problem with time windows were approached with success. In Bertazzi et al. [10] a matheuristic algorithm based on a clustering procedure is used for solving an inventory-routing problem in order to deliver freight to 942 different shops situated in the same urban area. Finally different authors described the clustered vehicle routing problem, based on the concept that the customers are aggregated on the base of capacity in clusters that have to be served in a precise order (Baratta et al. [11], Pop et al. [12]).

The rest of the paper will be available after the publication.