An advanced pre-trip planner with personalized information on transit networks with ATIS

A. Nuzzolo, A. Comi, U. Crisalli and L. Rosati

Abstract - The paper presents the first results of a research aiming to develop a transit trip planner to support the user with personalized pre-trip information. The first part describes the user needs and the architecture of the system. The second part deals with the modeling framework implemented to provide the best path alternatives from the traveler's utility point of view according to real-time data and personal user preferences. Finally, considerations on operative aspects based on some experimental evidences are presented.

I. INTRODUCTION

Traditional Traveller Information Systems (TIS) for transit networks can be classified as: pre-trip information systems, which support user choices before departure with one or more possible alternative paths; en-route information systems, which assist user during the trip and to which user can access via mobile devices; at-stop information systems, which mainly concerns the arrival times of approaching vehicles provided via VMS (Variable Message Signs) panels at stops.

Advanced Traveler Information Systems (ATIS) include Trip Planners able to identify the best travel alternatives on multiservice transit networks (including rail, metro, tram, bus and pedestrian) by using real-time data and by providing information for each travel alternative about: travel time, monetary cost, arrival times and other service characteristics, pollution, disruptions and so on. Using an advanced trip planner, the user can easily access useful and organized information to compare the different alternatives for a rational choice of the transport mode [4].

Advanced transit trip planners, providing accurate pre-trip information, reduce the uncertainty about routes and timetable that is one of the main reasons for rejecting transit as travel mode. The added value of ITS to improve transit ridership is demonstrated in many studies [1, 10, 21, 23].

The use of real-time data to improve the accuracy of trip planner information is crucial (e.g. in case of interchanges or when the user is constrained to the arrival time at destination) and it is another of the key features of advanced trip planners. In fact, the availability of real-time information allows a continuous en-route check of path alternatives, warning users about particular problems (e.g. delays, disruptions, congestion) on the path originally (pre-trip) recommended, and including the calculation of new path alternatives from the current position to the destination. Results of providing real-time arrival information can be found in [7, 8], while [22] investigated the impact of mobile real-time information.

Moreover, trip planners have to consider the different path alternatives on the basis of the different weights of path attributes (e.g. walking time, waiting time, on board time, transfer time) defined in relation to personal user preferences, which differ among users. For this reason, advanced transit trip planners have to provide personalized (individual) real-time information by using learning process mechanisms able to track profiled travelers and to give them information according to their personal travel habits [2, 19].

The paper presents the first results of an in-progress research aiming at developing an advanced trip planner, designed for mobile applications, that gives pre-trip and en-route dynamic real-time information to support the user in the choice of the best path from his/her personal (dis)utility point of view on a multiservice transit network.

In the following, section II illustrates the logical architecture of the trip planner according to the user needs. Even if the proposed trip planner includes pre-trip and en-route components, the current state of the research concerns the pre-trip component and this paper focuses on the features of the trip planner to support the user with personalized pre-trip information. Section III presents the modeling framework to provide personalized pre-trip information, focusing on the current state of the research and presenting the proposed advances in this field. Section IV describes some experimental evidences of the proposed pre-trip choice modeling framework carried out in the metropolitan area of Rome. Finally, section V reports some conclusions and the future developments of this research.

II. USER NEEDS AND LOGICAL ARCHITECTURE

The logical architecture of the system has to be developed considering the user needs that the trip planner has to satisfy. User needs can be classified according to the required pre-trip and/or en-route support.

The pre-trip support concerns transit path alternatives according to the desired arrival time and some path attributes (departure time, walking times and distances, waiting time at stops, on-board travel time, crowding, transfer number and times, and so on) in relation to real-time transit operations.

On the other hand, the main user needs to support the en-route choice are path attributes updating on real-time data from the user current position to the destination and new paths from the current position to the destination, if the chosen path becomes not available due to unexpected events (e.g. disruptions) or if there are relevant changes in the
chosen path due to changes in transit operations (e.g. heavy delays).

According to the above user needs, Fig. 1 illustrates the logical architecture of the pre-trip module of the trip planner, which has been implemented to support transit users with personalized information.

Referring to Fig.1, the pre-trip personalized advice module is enabled by a query of the registered user \( i \), who is logged into the system. At time \( \tau \) in which user \( i \) asks for a support to travel from origin \( O \) to destination \( D \) with a desired arrival time \( \tau_{Di} \), the system identifies and ranks the path choice set of user \( i \) based on his/her preferences and the current information on the multiservice transit network (i.e. scheduled timetable and real-time data) by using the path choice set identification and ranking procedures described in section III.

The ranking of alternative paths to provide to the user \( i \) is carried out by using the Random Utility Theory [3], through which personal utility parameters \( \beta_i \) of user \( i \) are used to calculate utilities for all paths of the choice set of user \( i \) (see section III.C).

The path chosen by user \( i \) is added to the database of revealed preferences of user \( i \), which is used to update the personal (individual) parameters \( \beta_i \) by using the procedure described in section III.D.

Considering the overall logical architecture, some operative aspects have to be considered to capture personal user preferences for a new user, a registered user already profiled for the same trip purpose and a registered user already profiled for a different trip purpose.

Each new user is asked to fill the registration form, which includes data on usual trip: purpose, origin, destination and desired arrival time at destination. Then, the user is asked to answer to a Stated Preference (SP) survey aiming at capturing his/her preferences/habits in path choice in order to initialize the main parameters required to apply the path choice modeling framework of section III. Experimental evidences for the initialization of a new user are deepened on section IV. The same procedure is activated for a registered user asking to plan a trip for a different trip purpose, as different parameters can be considered for the same attribute for different trip purposes (e.g. the different weight the user can consider for the late arrival time related to business trips with respect to leisure ones).

In the case of registered users asking to plan a trip for a previous profiled trip purpose, the trip planner uses the previously estimated individual parameters for this user and activates the learning process procedure, which uses the revealed path choices to the user parameters.

The chosen path of user \( i \) enables the en-route path information module, which should be able to assist user \( i \) during the trip (Fig.1). The en-route trip planner development is one of the future perspectives of this research.

![System architecture: pre-trip personalized information](image)

### III. Transit Path Choice Modelling

Transit networks are characterized by different boarding stops and available runs for the same Origin-Destination (O/D) pair and target time (e.g. desired arrival time). So, transit trip planners have to identify travel alternatives (paths) both in space (among stops) and in time (in relation to the user desired arrival time and/or the departure time of transit vehicles at stops).

A transit path \( k \) represents the space-time sequence of transport infrastructures and services used by users travelling from an origin \( o \) at a given origin departure time \( \tau_{Di} \) to a destination \( d \) with the relative arrival time at destination \( \tau_{d} \). A
path also includes access stop with relative arrival time, run departure time (or sequence of lines and runs including the relative stop interchanges) and egress stop.

For this reason, advanced transit trip planners have to use a path choice modeling approach that explicitly takes into account the above space-time features of path alternatives, especially in the case of low-frequency services when the explicit arrival/departure time coordination of runs at interchanges makes path alternatives possible or otherwise.

A. State-of-the-art of path choice modeling

The literature review on transit path choice modeling reports two main approaches: the frequency-based and the schedule-based. The frequency-based approach [15, 20] considers services in terms of sets of runs (lines). In this case, run (vehicle) scheduled times are not considered explicitly, but we refer to the line headways or to their inverse (the service frequencies), from which the name of the approach derives. Therefore the frequency-based approach is not able to calculate explicitly space-time attributes that users consider in relation to single runs, but it can refer only to average values relative to lines and it is suitable to be used for strategic planning.

On the other hand, the schedule-based approach [18] refers to services in terms of runs using the real vehicle arrival/departure time, and hence all the values of level of service attributes, evaluated at time in which users make their choices, can be explicitly taken into account. This approach allows us to take into account the evolution in time of both supply and demand, as well as run loads and level of service attributes. For this reason, the schedule-based approach is the unique way to model transit path choice on multimodal networks in presence of ATIS and it is used in the scope of this study. The schedule-based approach requires an explicit treatment for [18]: the user desired departure or arrival times; the supply modeling, which has to take into account each single run of transit services with its departure/arrival times at stops; the path choice models, which have to consider attribute time-dependencies. For a concise state-of-the-art on schedule-based approach, the reader can refer to [16]. Schedule-based path choice models for public transport, which consider a different user behavior according to different degrees of information provided to urban transit users by ATIS are presented in [6]. The extension to consider congestion through explicit capacity constraints is presented in [17].

B. The pre-trip path choice set generation

The pre-trip path choice set can be defined through a set of rules (selective approach) that allows us to generate feasible paths according to traveler characteristics, transport system performance (e.g. travel times and costs) and traveler behavioral assumptions. It implies that only some of the topologically feasible paths are considered. The set of rules to reduce the potentially high number of path alternatives can be heuristically calibrated by matching them with a sample of observed choices (maximum coverage factor method).

C. The individual path utility model

Path choice models allow us to estimate the utility associated to each path \( k \) belonging to a set of possible alternative paths defined according to traveler characteristics, transport system performance (e.g., travel times and costs) and traveler behavioral assumptions.

Given an O/D pair \( od \) and a target time \( T_{de} \), an example of utility \( V_{od,T_{de}}^k[k] \) at time \( r \) associated by user \( i \) to the path \( k \) is given by the increment in utility if user \( i \) were to take path \( k \) and given the target time \( T_{de} \), access stop \( s \) and run \( r \) (or sequence of runs of in the case of interchanges) can be written as:

\[
V_{od,T_{de}}^k[k] = E_D + E_A + \sum m (TW_{ms} \cdot \omega_{ms} + OB_{ms}) + CFW_{ms} + \sum m (MP_{ms} \cdot \omega_{ms} + NT_{ms})
\]

where: \( E_D \) is the Early or Late arrival time (i.e. the difference between the desired and the actual arrival time at destination) using path \( k \); \( E_A \) is the sum of access and egress times on path \( k \); \( TW_{ms} \) is the waiting time spent for boarding runs of the transit service \( m \) (e.g. train, metro, bus) belonging to path \( k \); \( OB_{ms} \) is the on-board time spent on the transit service \( m \) belonging to path \( k \); \( CFW_{ms} \) is the average on-board crowding degree on runs of transit service \( m \) belonging to path \( k \); \( MP_{ms} \) is a preference attribute for transit service \( m \) on path \( k \) (e.g. expressed as a function of the travel distance on transit service \( m \) w.r.t. the total distance on path \( k \)); \( NT_{ms} \) is the number of transfers on transit service \( m \) belonging to path \( k \); \( \omega_i \) are the model parameters. They are estimated on time-dependent attribute values provided by ATIS at time \( r \).

Given path utilities, the trip planner ranks the alternative paths based on such utilities and suggests such paths to the user.

The estimation of personal model parameters \( \beta_i \) of (1) can be carried out as follows

D. The individual pre-trip path utility parameters

Calibrating user personal path choice parameters implies that the model functional form could be the same, as the same could be the values of attributes considered in path choice, but different set of parameters (different for each registered user) have to be considered and calibrated according to his/her personal travel preferences.

The estimation of individual coefficients \( \beta_i \) can be obtained by a sample of observations of a single user \( i \). The best set of \( \beta_i \) parameters is the one that allows us to provide the best perceived paths and their ranking.

It is rare to see discrete choice models estimated for single individuals [5, 9]. In fact, demand models usually simulate the average number of trips of given characteristics undertaken by homogeneous user groups. Large choice samples of single decision-maker are difficult to be obtained and it is easier to have choice samples from many decision-makers. Therefore, instead of disaggregate or individual behavioral models, user groups (homogeneous with respect to their attributes, parameters and the functional form of the models) have been used and aggregate behavioral models have been developed. But, as here detailed, the performances of aggregate models appear limited to give personal travel
advices, because of the dispersion among users and/or the variations in taste/preferences among users.

The estimation of discrete choice model parameters with repeated observations for each respondent gives a correlation of disturbances, or heterogeneity (e.g. the parameters can also vary for the same user according to travel purpose), which refers to variations in unobserved contributing factors across behavioral units. If behavioral differences are largely due to unobserved factors, and if unobserved factors are correlated with the measured explanatory variables, then estimates of model coefficients will be biased if heterogeneity and correlation are not properly treated. Besides, the collected data consist of multiple observations of the same user, the assumption of independence between choices for the same user may not be appropriate. For this reason, there has been growing interest in more advanced logit specifications, such as error component or nested logit (i.e. able to consider that the paths could be not completely independent due to possible overlaps) or mixed logit (i.e. to deal the unexplained heterogeneity in choice data, [12])

A two-step procedure was implemented to obtain the individual pre-trip path utility parameters that point out the individual preferences. The first step initializes the path utility function parameters of a new user, on the basis of Stated Preference (SP) surveys. The second step updates the initial model parameters by using the individual revealed choices acquired by Traveller Tool in past travel experiences.

As regards the initialization of the path utility function of new users, [14] proposed a method in which the updating starts from initial average parameters obtained through multi-user SP interviews. This paper explores a possible alternative approach, which requires some data of the new user, i.e. origin and destination of a typical and well-known transit trip, as well as the desired arrival time. Then, the system generates and suggests to user some path alternatives according to the system past-recorded data. The data from this SP survey are hence used for estimating the preliminary parameters of the individual path utility function.

In the second step, model parameters are updated using a user preference learning procedure based on the choice revealed when the traveler uses the system. The parameter updating procedure can use two different approaches based on: Bayesian and batch methods. The Bayesian method is characterised by the probabilities of parameter estimators, which are usually represented by a normal distribution and whose mean values represent the current best estimate of the user’s value of the parameters. At each time in which an user choice is observed, the system updates the model parameters by updating the distribution of each parameter, i.e. by updating means and standard deviations using Bayes method. The current Bayesian methods of learning continuous parameters are not feasible for incremental learning due to the long computation times needed for a learning step. For this reason, [2] proposed an advanced Bayesian method that reduces computation time by assuming a sequential processing of parameters and a systematic sampling of the parameter space. The batch method delivers a new estimation of model parameters when a given number of user choices are observed. Such estimation is carried out by a Maximum Likelihood approach [3, 13] on the basis of all the available observations.

IV. EXPERIMENTAL EVIDENCES

In order to investigate some issues of the individual path utility model, a prototype version of the pre-trip planner was applied to the transport system of the metropolitan area of Rome, served by a multiservice transit network operated by different companies, where urban bus, tram and metro regional railway lines and regional bus lines with an integrated fare policy are available. The real-time information provided by each agency were integrated within an unique trip planner to test effects of pre-trip personalized information based on real-time data. In particular, some tests were performed to focus on the individuation of the best specification of the random utility model and to the optimization of parameter initialization and updating procedures in order to reduce the time spent by user during registration phase.

The tests were performed considering a working-day journey from Frascati (a town near Rome) to the center of Rome (Piazza Sempione) with the desired arrival time at 9.30 am. The distance on the od pair by car is about 25 km (Fig. 2). Four different paths with an average total travel time of about 2 hours were available to travel on this od pair. They differ in terms of travel time, waiting and transfer times, modes to be used (train, metro or bus) and the early/late arrival time respect to desired one.

A. The individual path choice model

Different model specifications were tested starting from the simplest multinomial logit model (MNL) to the mixed-logit (MXL) and nested-logit (NSL) ones. Path choice models were estimated carrying out a SP survey with 150 sets of four alternatives on a test user (a university student) that was asked to choose the preferred one. The set of 150 scenarios with 4-alternatives was defined by path alternatives randomly extracted from the previous experimented status of the transportation system. Even if further analyses are in progress, the development of nested and error components logit models did not provide satisfactory results. Hence, in
The following only the results obtained by MNL (Multinomial Logit) and MXL (Mixed Logit with normal distribution of parameters) estimations are presented. According to [11], the random parameters were selected starting from the assumption that all parameters included in logit models could be random. Then, only the early/late arrival time was selected as random parameter, while the other parameter estimates remain quite constant between the two specifications. Calibration test using all attributes of Equation (1) were performed, but only those reported in Table I were statistically significant.

**TABLE I. MODEL ESTIMATION USING 4-ALTERNATIVE SETS**

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Multinomial Logit (MNL)</th>
<th>Mixed Logit (MXL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>t-stat</td>
</tr>
<tr>
<td>Waiting time (total)</td>
<td>-0.439</td>
<td>-4.06</td>
</tr>
<tr>
<td>On-board time (min)</td>
<td>-0.078</td>
<td>-1.30</td>
</tr>
<tr>
<td>On-board time (metro)</td>
<td>-0.336</td>
<td>-1.64</td>
</tr>
<tr>
<td>Early and late arrival time</td>
<td>-0.426</td>
<td>-3.24</td>
</tr>
<tr>
<td>Standard deviation of early and late arrival time</td>
<td>0.179</td>
<td>1.77</td>
</tr>
<tr>
<td>% of-right</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td>% of-right including the best and the 2nd best alternatives</td>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>

The estimated parameters corresponding to different components of travel time (e.g., waiting, on board and delay time) have increasing absolute values for less appreciated components. For example, the time spent on the bus is lower than on the metro and is about twice lower than the early/late arrival time. The lowest value of travel time refers to time spent on the regional train. Deepening on results of Table I, we can see that the 4-alternative model predicts at least the 84% of the chosen paths (see % of-right), but if we also consider the second paths, the % of-right grows up to the 99%. Continuing the analysis of values of Table I, it is possible to see similar performances of both MXL and MNL models. For this reason, for its well-known easy-to-apply advantages, we applications below only refer to the MNL model.

**B. The initialization parameter estimation**

The main effort in the definition of the procedure for the initialization of individual model parameters for new users is to minimize the time needed for the estimation of user-tailored model parameters able to provide path suggestions that reasonably are those expected.

Hence a particular attention should be given to the initialization phase, which should not be too long and complex to avoid the user becomes bored during this operation. The investigation of this problem regards the minimum number of scenarios to be proposed to the new user and the minimum number of attributes that can be used in the initial path utility functions, by considering that the number of quite correct estimated parameters could also depend on the number of available observations. Therefore, the possibility to limit the time the user has to spend for the initialization phase was investigated by reducing the number of alternatives to be suggested during this operation. The analysis was carried out through a further SP survey consisting of 150 scenarios with only 2-alternative paths, which were randomly selected among the same 4 path alternatives described in section IV.

As previously estimated, different behavioral models were taken into account (see Table II) and only MNL and MXL (with on-board time on the bus as random parameter) models allowed us to obtain satisfactory results. In fact, as expected, the % of-right increased w.r.t. values of Table I (e.g. from the 84% to the 93% for the MNL model), but the real comparison in model performances have to be carried out by considering the 4-alternative scenarios. Therefore we calculated the % of-right in reproducing 4-alternatives and in this case (see table II) the MNL models predicted the alternatives chosen by user with an 81% rate. Hence, with a reasonable reduction of the % of-right (about 3%), the model based on the SP scenarios with 2-alternatives instead of that with 4-alternatives can be used and acceptably applied to save time in the initialization phase.

**C. The model parameter updating**

Starting from the above results, the learning process was also simulated adding 150 SP survey data with 4-alternative sets to the previous 10 observations. The parameters of the utility function were estimated varying the number of attributes and the number of observations. The parameter updating was performed including 10 new observations at a time (batch updating). The first results reported in Table III shows that the improving of the learning process is quite slow because too many observations are necessary to obtain a statistically good model. Parameter updating and model performances.

**TABLE III. PARAMETER UPDATING AND MODEL PERFORMANCES.**

<table>
<thead>
<tr>
<th># of tot.</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of-right</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
</tr>
<tr>
<td>% of-right including the best and the 2nd best alternatives</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
</tr>
</tbody>
</table>
In fact, 60 observations are needed if we consider a model with only three parameters and the 80%-of-right, while 80 observations are required if we use four parameters, and even 120 if five parameters are taken into account. It implies that further analyses have to be carried out in order to verify if other estimation procedures (e.g. the Bayesian method) are more performing in terms of swiftness of convergence.

V. CONCLUSIONS

This paper presented some first results of a research aiming to develop a trip planner able to give pre-trip personalized information to the user about travel alternatives on a transit network by real-time data. It is based on a path choice modeling framework able to provide transit path alternatives on the basis of user personal travel preferences defined according to a learning procedure.

At first, the problem of the definition of the best model specification to provide the best paths perceived by each user was investigated. Then, the initialization and the updating of the model parameters were analysed by pointing out the choices made by some students travelling for leisure. Experimental evidences show that the model parameter initialization phase (performed by SP interviews) can use 2-alternative scenarios with a minimum number of 10 observations, through which the estimated parameters of the path utility function allow us to suggest the individual preferred paths with a good reliability. As regards the updating process of individual model parameters, we experienced that the process can be quite slow and too many observations are needed to reach statistically significant results. This conclusion gives the way for the exploration of different approaches, including those based on Bayesian methods. Further developments of this research are in progress. They concern the additional investigation of the path choice modelling and the extension of the trip planner to provide en-route personalized information. In particular, advances in path choice modelling regard the exploration of other od pairs, user preferences and model forms and the design of ad-hoc SP surveys will be studied for the initialization phase aiming at deepening on behavioural aspects related to specific user preferences.

REFERENCES