A Future Internet Oriented User Centric Extended Intelligent Transportation System*

S. Canale, A. Di Giorgio, F. Lisi, M. Panfili, L. Ricciardi Celsi, V. Suraci, and F. Delli Priscoli

Abstract—Intelligent Transportation Systems (ITS) are changing the way people plan a journey and travel around the world. Advanced mobility information systems, as well as intelligent multimodal mobility services, may take considerable advantage of consolidated technologies from emerging ICT frameworks. In this paper we propose an Extended Intelligent Transportation System (ExITS) consisting of a basic ITS equipped with a User Centric Control System (UCCS). The proposed ExITS relies on service personalization methodologies and is conceived as a Future Internet (FI) oriented, closed-loop, user centric architecture integrating and controlling ITS services. The proposed UCCS considers the trip planning service and takes into account both explicit and implicit user preferences in selecting travel solutions satisfying a given user request. The aim of the UCCS is to drive the trip planning service in proposing to the user travel typologies tailored to preferences in selecting travel solutions satisfying a given user request. The aim of the UCCS is to drive the trip planning service in proposing to the user travel typologies tailored to preferences in selecting travel solutions satisfying a given user request. Implicit preferences are automatically inducted by similarity based unsupervised machine learning techniques and verified by a closed-loop control mechanism triggered by explicit user feedback.

I. INTRODUCTION

Smart data aggregation, transmission and analysis are key features in Intelligent Transportation System (ITS). Since the beginning, these mobility support systems allowed both urban, local trip planning and control solutions integrating private and public transportation means, and door-to-door long distance multimodal journey planning with the aim of optimizing specific travel aspects (e.g., cultural visit, low emission, etc.). Due to the wide deployment and large range of applications, ITSs can take advantage of the most advanced ICT architectures and technologies. Among them, Future Internet (FI) offers one of the most promising frameworks for efficient, large-scale solutions.

In this work, we present a FI oriented system for a closed-loop, control-based approach driving existing ITSs in personalizing basic services for end users. The proposed approach is user centric, in the sense that it is sufficiently general to allow the personalization of a family of services, ranging from trip planning and control services to tariff design. In particular, we consider the personalization of a multimodal trip planning service. In Section II methods and solutions concerning ITSs and FI are described. Section III describes the reference scenario fostering the introduction of the control-based system proposed in Section IV. Section V describes models and algorithmic aspects of the proposed system, while Section VI summaries future works and conclusions.

II. STATE OF THE ART

A. Intelligent Transportation Systems

Nowadays ITSs represent an important topic from the technological and business point of view. They mainly rely on models and algorithms from Transportation Engineering that are deeply changing our travel habits. The main goal is that of making current transportation systems more and more safe, secure and efficient, as well as providing intelligent multimodal trip plans and reducing risks, traffic congestion, and CO2 emissions. From the ICT point of view, ITSs take advantage of real-time information and communication for supporting end user decision-making [1]. In this respect, an ITS can be considered as intelligent since it combines several disciplines including, but not limited to, effective and efficient optimization and control algorithms, as well as computer science innovative services and advanced architectures.

According to [2], from the functional point of view, an ITS can be divided into six fundamental components (see Fig. 1): (i) Advanced Traveller Information System (ATIS), (ii) Advanced Transportation Management System (ATMS), (iii) Advanced Vehicle Management (AVM), (iv) Business Vehicle Management (BVM), (v) Advanced Urban Transportation System (AUTS), and (vi) Advanced Public Transportation System (APTS). Among them, ATIS represents one of the most important components in ITS, especially because it provides travellers with suitable real-time detailed information about journey, traffic status, public transport information, including time table and current availability of each transport means. From the conceptual point of view, ATIS consists of two main modules [3]: (a) Pre-Trip Traveller Information System (PTTIS) and (b) En-Route (also known as On-Trip) Traveller Information System (OTTIS). In applications of practical interest, ATIS can include the Multimodal Trip Planner (MTP) module (see [4]-[6]). In order to provide suitable real-time information about travel, usually data coming from several, heterogeneous sensors are gathered and subsequently processed into the Advanced Transportation Management System (ATMS) [1]. A real challenge in ITS is integrating different, heterogeneous and dynamic data sources and providing information for each end user in a multimodal transport model that could cover long distance journeys [7].

*Research supported by European Commission in the framework of the BONVOYAGE project (From Bilbao to Oslo, intermodal mobility solutions and interfaces for people and goods, supported by an innovative communication network) funded by EU under Grant Agreement n° 635867.

S. Canale, A. Di Giorgio, F. Lisi, M. Panfili, L. Ricciardi Celsi, V. Suraci, and F. Delli Priscoli are with the Department of Computer, Control and Management Engineering Antonio Ruberti, University of Rome La Sapienza, via Ariosto 25, 00185 Rome, Italy (email:{canale, digiorgio, lisi, panfili, ricciardicelsi, suraci, dellipriscoli}@diag.uniroma1.it).

V. Suraci is also with the Department of Engineering at eCampus University, Novedrate (CO), Italy.
B. Future Internet and ITS

FI concepts include the presence of technology-independent cooperative controllers (also referred to as Generic Enablers in the FIWARE/FICORE FP7 projects, [8] and [9]) able to efficiently control the available network resources in a context-aware closed-loop fashion (e.g., [10]-20) and, at the same time, able to guarantee the user satisfaction in terms of a personalized Quality of Experience [21]. In this respect, an interesting approach is the so-called cognitive approach of FI, as introduced in [22]-[25], and personalized driving parameters are introduced in [26] with respect to telecommunication networks. This concept is exploited in [27] in combination with web services [28], to improve the operation of power systems taking advantage of telecommunications while integrating electric vehicles. Preliminary research contributions to ITS from FI perspective are given in [29] and [30], where a FI based multimodal travel platform is simulated. The proposed approach has been developed within the Instant Mobility FP7 project [31].

C. Personalization in mobility services

A highly challenging feature in ITSs is the personalization of real-time information, aimed at providing end users with suitable, on-line support taking into account user (i) requirements, (ii) preferences, and (iii) behavioural profile. Several approaches have been considered, ranging from Recommendation Systems to advanced Expert Systems. In [4], an advanced trip planner is presented aimed at providing personalized user information and travel alternative path. A personalized information retrieval system for filtering travel results, such as to satisfy user specific needs, is conceived in [32]. In [33], a multi-criteria decision making approach is proposed and validated in order to help, in a personalized way, the end user in the journey selection among several different travel solutions. In [34], a framework integrating Case Based Reasoning and Multi Criteria Decision Making is proposed in order to improve the user satisfactions for itinerary search in urban area. Three distinct methods aimed at predicting journey times and ranking typologies of Points of Interest (i.e., stations, museums, etc.) according to user interests are proposed in [35].

Sophisticated machine learning methodologies have been proposed for user profiling. A Markov Decision Process and Reinforcement Learning based approach is proposed in [36] to learn how to support users in decision-making process through human-computer interaction. Bayesian methods have been adopted in [37] in order to learn user travel preferences by considering user past travel choices.

This paper is in line with FI based frameworks for ITSs proposed in the literature. We propose an Extended Intelligent Transportation System (EXITs) whose key difference, with respect to previous control and/or learning based approaches for ITS, consists in jointly taking into account the user request submitted to the Multimodal Trip Planning (MTP) module and the actual choice carried out by the user as one or more Candidate Travel Solutions are returned. As a matter of fact, these choices are elaborated by a dedicated User Centric Control System (UCCS) in order to refine the user behavioural profiling and eventually producing the so-called Personalized Optimality Criteria which drive the basic ITS to provide personalized travel solutions representing the Personalized Control Decisions as explained in Sections IV. The UCCS extends traditional ITS functionalities by automatically learning user preferences from the user behaviour, even in the case the user preferences are in contrast with the ones explicitly declared.

III. Reference Scenario

The transport network is modelled as a multigraph $G(V,E)$, namely a graph with possible multiple edges between the same vertices. A multigraph can be described by a function $f: E \rightarrow V \times V$ indicating the vertices $(v_i,v_j)$ connected by a given oriented edge $e_i$, i.e. $f(e) = (v_i,v_j)$. In our model, there exists an oriented edge $e_i$ connecting the vertices $(v_i,v_j)$ if there exists a transport mean $m$ directly linking the node $v_i$ with the node $v_j$ starting at time $t_i$ from the node $v_i$ and arriving at time $t_f$ at the node $v_j$. We assume that there are $M$ distinct transport means. A transit node in $V$ is a node $v$ such that $v$ is the source node of a link $e_i$ served by a transport means $m_i$ and the destination node of a link $e_j$ served by a transport means $m_i$ for some $i \neq j$ and $i,j \in \{1,...,M\}$.

By referring to Fig. 2, the ITS includes the so-called Advanced Travel Information System (ATIS), whose main task is to keep dynamically updated the structure of the above-mentioned multigraph $G(V,E)$ on the basis of the real-time mobility information provided by Transport Operators. As a matter of fact, $G(V,E)$ is a basic input for the Multimodal Trip Planner (MTP) module. In this paper, we assume that $U$ users are registered in the ITS platform and have subscribed the trip planning service. This service is used by user $u$ whenever $u$ needs to plan a door-to-door travel, i.e. a complete trip from a specific source location $a$ to a target...
destination location \( b \). The source and destination locations of the travels are two particular vertices of \( G(V,E) \). We indicate by \( d_\text{s}(a,b) \) the Euclidean distance between two dimensional vectors of GPS coordinates of physical locations \( a \) and \( b \). User \( u \) can indicate the time \( t_a \) for travel to begin. \( t_a \) can be a specific time in a day, a generic daytime or a range of time in a day.

Additionally, user \( u \) can provide a set of User Constraints, indicated by \( UC(u,a,b,t_a) \), for the travel from \( a \) to \( b \) starting at time \( t_a \). Typical constraints concern the number of passengers travelling with \( u \), special needs indicated by \( u \), allowed transport means among the possible \( M \) transport means, etc. We consider \( S \) possible special needs that user \( u \) can declare. Finally, user \( u \) can indicate an explicit set \( UP(u,a,b,t_a) \) of User Preference Criteria about the desired travel. The criteria \( UP(u,a,b,t_a) \) are provided in order to explicitly guide the MTP in proposing the most suitable travel solutions. In this work, we consider a basic family of criteria consisting of: (1) minimizing the overall travel time; (2) minimizing the overall travel cost; (3) maximizing comfort level during the journey; (4) maximizing the class category of each transport modality; (5) minimizing \( CO_2 \) emissions during the trip; (6) minimizing the number of transit nodes; (7) minimizing the walking distance; and (8) minimizing the number of distinct modality means. These criteria represent a subset of a more general set of Optimality Criteria, here indicated as \( OCR = \{oc_1,\ldots,oc_P\} \), consisting of all optimality criteria considerred by MTP. In general, the cardinality \( P \) of OCR depends on the specific MTP and can be a very high number, often including particular combinations of criteria (e.g., convex combinations).

We refer to as User Query, indicated as \( Q(u) = (u,a,b,t_a,UC,UP) \), the \( i \)-th query submitted by the user \( u \) who wants to plan the travel from the source location \( a \) to the destination location \( b \) starting at time \( t_a \) with the User Constraints \( UC(u,a,b,t_a) \) and the User Preference Criteria \( UP(u,a,b,t_a) \). For the sake of readability, we will refer to a generic \( i \)-th User Query \( Q(u) \) as \( Q(u) \). A Travel Sequence related to \( Q(u) \), indicated as \( s(Q(u)) \), consists of a list of \( L \) distinct and consecutive edges \( e_l \) (\( l = 1,2,\ldots,L \)) linking the source \( a \) with the destination \( b \) in such a way that \( f(e_l) = (v_{l,v_l}), f(e_{l+1}) = (v_{l+1,v_{l+1}}), t_{l} < t_{l+1} \), \( f(e_1) = (a,v_1) \) and \( f(e_L) = (v_b,b) \).

A Travel Sequence \( s(Q(u)) \) is referred to as Feasible if \( s(Q(u)) \) respects all constraints \( UC(u,a,b,t_a) \). A Feasible Travel Sequence \( s(Q(u)) \) is Optimal if \( s(Q(u)) \) optimizes some Optimality Criterion in \( OCR \). Note that, for processing time constraints, MTP cannot consider all the \( P \) possible Optimality Criteria; conversely, MTP will consider a suitably selected subset of \( OCR \) including, at the minimum, \( UP(u,a,b,t_a) \). This subset will be hereinafter referred to as the set \( SOC \subseteq OCR \) of Selected Optimality Criteria. A Feasible Travel Sequence \( s(Q(u)) \) being relevant to a given \( Q(u) \) and optimal according to some optimality criterion in \( SOC \) identifies a Candidate Travel Solution for \( Q(u) \). We assume that MTP is able to find the entire set of Candidate Travel Solutions for \( Q(u) \) on the basis of the set \( SOC \) and of the information related to the present structure of \( G(V,E) \). Let \( TS(Q(u)) = \{s_1(Q(u)), \ldots, s_m(Q(u))\} \) be the set of Candidate Travel Solutions for \( Q(u) \) returned by MTP. Typically, the cardinality \( n \) of \( TS(Q(u)) \) increases as the number \( M \) of transport modalities gets larger and/or the distance \( d_\text{s}(a,b) \) increases.

We define Selected Travel Solution and indicate by \( s^*(Q(u)) \) the Candidate Travel Solution in \( TS(Q(u)) \) which is actually selected by the user \( u \). We assume that the user \( u \) can be assigned the User Profile grouping all users being more similar to \( u \) with respect to three families of parameters: the distance \( d_\text{s}(a,b) \), the constraints \( UC(u,a,b,t_a) \) and the criteria \( UP(u,a,b,t_a) \). The identification of \( K \) possible User Profiles is a task demanded to a specific functionality of the UCCS as shown in Section IV. Now, we introduce three important functions:

- \( OCR \): \( s \rightarrow OCR \) is the function indicating the optimality criterion \( oc(s) \) in \( OCR \) considered by MTP to provide the Candidate Travel Solution \( s(Q(u)) \) in \( TS(Q(u)) \);
- \( prl \): \( u \rightarrow prl_1, prl_2, \ldots, prl_K \) is the function assigning a user \( u \) to her/his User Profile \( prl(u) \) properly selected in a set of \( K \) possible User Profiles;
- \( prs \): \( prl_k \rightarrow \{0,1\}^P \) is the function identifying the subset of \( OCR \) being suitable for the User Profile \( prl_k \); in other words, for each possible \( prl_k \) for \( k = 1,2,\ldots,K \) the function \( prs \) allows the identification of the most relevant optimality criteria in the User Profile \( prl_k \).

IV. EXTENDED INTELLIGENT TRANSPORTATION SYSTEM

In this section, the Extended Intelligent Transportation System (ExITS) is outlined with respect to two working modes: the Basic Mode and the Cognitive Mode. In the following, we will refer to Fig. 2 in order to illustrate these working modes and to clarify the main differences between them. Both working modes are based on a basic ITS consisting of the main components ATIS and MTP as explained in Section III.

Both working modes allow each user \( u \) to submit a User Query \( Q(u) \) in the pre-trip phase to the MTP. In both working modes, MTP is in charge of selecting the set \( TS(Q(u)) \) of \( n \) Candidate Travel Solutions which is returned to the user \( u \) (arrow “F” in Fig. 2). The MTP performs such selection on the basis of \( Q(u) \) (arrow “A”) and \( G(V,E) \) (arrow “C”). Multigraph information is dynamically elaborated by the ATIS on the basis of the real-time Mobility Information received from Transport Providers (arrow “B”). Note that the dynamic information included in \( G(V,E) \) allows the MTP to react, in real-time, with respect to unforeseen events affecting the mobility dynamics. The Basic Mode, differently from the Cognitive Mode, does not make use of the User Centric Control System (UCCS). Therefore, with reference to Figure 2, the information flow of the Basic Mode just foressees the sequence “A”, “B”, “C”, and “F”.

Conversely, the Cognitive Mode makes use of the UCCS including two fundamental functionalities being in charge for the actual personalization of mobility services: (i) the off-line User Profile Identification functionality and (ii) the on-line User Profiling functionality. In the Cognitive Mode, \( Q(u) \) submitted by the user \( u \) (arrow “A”) is forwarded to the UCCS (arrow “D”). As the UCCS receives \( Q(u) \), the User Profiling functionality identifies, in real-time, the User Profile \( prl(u) \) the user \( u \) belongs to in the set of \( K \) possible User Profiles. The identification of this set is demanded, off-
line, to the User Profile Identification functionality. Then, the User Profiling functionality identifies, via the \( prs(prl) \) function when \( prl = prl(u) \), the Personalized Optimality Criteria \( POC(Q(u)) \leq OCR \), associated to the User Profile \( prl(u) \), i.e., optimality criteria being more likely preferred by the user \( u \) when travelling from \( a \) to \( b \) starting at time \( t_a \) as indicated in \( Q(u) \). In general, the cardinality of \( POC(Q(u)) \) will be much lower than the cardinality of \( OCR \).

The set \( POC(Q(u)) \) of Personalized Optimality Criteria are provided by the UCCS to MTP (arrow “E”). Accordingly, MTP can calculate the set \( TS(Q(u)) \) of Candidate Travel Solutions with respect not only to the explicit User Preference Criteria \( UP(u, a, b, t_a) \) (as in the Basic Scenario, arrow “A”), but also to the implicit Personalized Optimality Criteria \( POC(Q(u)) \) (arrow “E”). In other words, in the ExITS, the set \( SOC \) includes both the set of User Preference Criteria and the set of Personalized Optimality Criteria, i.e. \( SOC = UP(u, a, b, t_a) \cup POC(Q(u)) \). Therefore, the MTP considers both preferences explicitly indicated by the user \( u \) in \( Q(u) \) and the Personalized Optimality Criteria selected by the UCCS as user profile based preferences implicitly inducted by the machine learning approach described in Section V.B. In the Cognitive Mode, the set \( TS(Q(u)) \) of Candidate Travel Solutions, computed by the MTP according to \( SOC \), is returned to user \( u \) (arrow “E”) once ranked according to the following rule. The first Candidate Travel Solutions appearing in the ranked list are the ones satisfying the explicit User Preference Criteria \( UP(u, a, b, t_a) \). Instead, the following Candidate Travel Solutions satisfy the implicit Personalized Optimality Criteria \( POC(Q(u)) \). Once the ranked set \( TS(Q(u)) \) of Candidate Travel Solutions is returned to user \( u \), such a user makes her/his choice by selecting one of Candidate Travel Solutions in \( TS(Q(u)) \), i.e., the Selected Travel Solution \( s^*(Q(u)) \). It is fundamental to note that, differently from the Basic Mode, the Selected Travel Solution \( s^*(Q(u)) \) is provided to the UCCS (arrow “G”), in order to be used as a feedback of paramount importance in the machine learning algorithms embedded in the User Profiling Identification module. Therefore, with reference to Fig. 2, the information flow of the Cognitive Mode foresees the complete sequence from “A” to “G”.

It is important to remark that the Cognitive Mode is fully in line with the Future Internet concept of considering a service/technology independent Controller (here represented by the MTP). In fact, the Controller takes control decisions (here represented by the set \( TS(Q(u)) \) of Candidate Travel Solutions) on the basis of the feedbacks directly coming from monitoring of the present system status (here represented by the User Query \( Q(u) \) and the Mobility Information), according to a control law depending on some user centric, personalized Driving Parameters (here represented by the Personalized Optimality Criteria \( POC(Q(u)) \)).

Within the UCCS, four main functional modules are in charge of dealing with User Query \( Q(u) \) at the aim of generating the Personalized Optimality Criteria \( POC(Q(u)) \).

A. Travel Knowledge Base

The Travel Knowledge Base has the role of storing all additional data considered in the Cognitive Mode but not in the Basic Mode. These interactions are represented by a set HP of records \( R(u) \) of the form

\[
R(u) = \{ Q(u), POC(Q(u)), OCR(s^*(Q(u))) \}^T
\]

Each \( R(u) \) includes the \( i \)-th User Query \( Q(u) \) submitted by user \( u \), the Personalized Optimality Criteria \( POC(Q(u)) \) returned by UCCS with respect to \( Q(u) \), and the optimality criterion \( OCR(s^*(Q(u))) \) considered by the MTP when \( s^*(Q(u)) \) is the Selected Travel Solution \( s^*(Q(u)) \). The Travel Knowledge Base stores records having the structure (1). A new record of this structure is created whenever a Candidate Travel Solution in \( TS(Q(u)) \) is selected by \( u \). The set HP of historical records are provided to the User Profile Identification module so that the set of \( K \) User Profiles can be calculated off-line. In addition, when a new \( Q(u) \) triggers the UCCS (arrow “D”), the User Profiling module detects on-line the User Profile \( prl(u) \) associated to user \( u \) on the basis of the information stored in the Travel Knowledge Base. Accordingly, Personalized Optimality Criteria \( POC(Q(u)) \) are provided to the MTP via the \( prs \) function (arrow “E”). Once the Selected Travel Solution \( s^*(Q(u)) \) is chosen in \( TS(Q(u)) \) by the user \( u \) (arrow “F”), a new record \( R(u) \) is inserted in the Travel Knowledge Base.

B. Meta Data Handling Tool

The Meta Data Handling Tool offers basic procedures to gather, store and manage data records of the form (1) in the Travel Knowledge Base. This module is in charge for extracting the features of a given \( Q(u) \), namely the distance \( d(a, b) \), the vector \( UC(u, a, b, t_a) \) and the vector \( UP(u, a, b, t_a) \), as made available by the ITS (arrow “D”). Moreover, this module is in charge of transmitting the set \( POC(Q(u)) \), identified by the User Profiling, to MTP. Finally, the Meta Data Handling Tool implements the function \( OCR : s \rightarrow OCR \) indicating the optimality criterion \( OCR(s) \) considered by the MTP for Candidate Travel Solution \( s \).

The proposed ExITS is sufficiently general in the sense that UCCS can adapt to any MTP by means of the Metadata Handling Tool. In fact, every MTP accepts User Query \( Q(u) \) containing at least the source \( a \) and the destination \( b \). In this elementary case, \( UC(u, a, b, t_a) \) and \( UP(u, a, b, t_a) \) are empty vectors and the only data that the Meta Data Handling Tool can extract from \( Q(u) \) is the distance \( d(a, b) \). Accordingly, other components in UCCS can work seamlessly.

C. Off-line User Profile Identification

The User Profile Identification is the core module of UCCS. In fact, this module is able to analyse a huge amount of historical records HP of the form (1) stored in the Travel Knowledge Base and to identify the set of \( K \) User Profiles that will constitute the basic family of user profiles driving the User Profiling module. The automatic identification of the User Profiles is made on the basis of the similarity between historical records \( R(u) \) in HP. The main idea is that similar User Queries \( Q(u) \) and \( Q(w) \) should be characterized by similar sets \( TS(Q(u)) \) and \( TS(Q(w)) \) of Candidate Travel Solutions, that means similar Personalized Optimality
Criteria \( POC(Q(u)) \) and \( POC(Q(w)) \). Note that the Personalized Optimality Criteria \( POC(Q(u)) \) are proposed by the ExITS on the basis of historical patterns of User Query being similar to \( Q(u) \). On the contrary, the Selected Travel Solution \( s^*(Q(u)) \) is explicitly chosen by the user \( u \) and, therefore, the optimality criterion \( ocr(s^*(Q(u))) \) represents the actual, implicit preference of user \( u \) with respect to \( Q(u) \).

The User Profile Identification algorithm for determining the Personalized Optimality Criteria \( POC(Q(u)) \) is described in Section V.C.

### D. On-line User Profiling

The User Profiling module implements two distinct functions. The former is the function \( prl: u \rightarrow \{prl_1, prl_2, ..., prl_K\} \) assigning a user \( u \) to the User Profile \( prl(u) \) in the set of \( K \) possible User Profiles already identified by the User Profile Identification module. The latter is the function \( prs: \{prl_1, prl_2, ..., prl_K\} \rightarrow \{0,1\}^p \) identifying the subset of OCR being suitable for each User Profile \( prl_k \) for \( k = 1, ..., K \).

### V. DATA DRIVEN USER PROFILING

In this section, functionalities of the User Centric Control System (UCCS) presented in Section IV (see Fig. 2) are described from the methodological and algorithmic point of view. These functionalities rely on similarity based behavioural profiling techniques typically used to learn user preferences by analysing a huge amount of data records and inferring similar behaviours. Data describing both user preferences and behaviours with respect to the considered service, in this case the MTP, are a key aspect of the analysis.

#### A. Structure of data record

The complete set of entries of record \( R(u) \) is summarized in Table I. It is important to remark that all entries are numerical values (real, integer or Boolean). Of course, according to the specific features of the considered MTP, some sub-vectors in the User Query \( Q(u) \) can be available or not. In this case, the Meta Data Handling Tool will not manage these data and the record stored in the Travel Knowledge Base will not include the resulting missing data with no consequence for the machine learning algorithms described in Sections V.B and V.C.

#### B. Machine learning approach for user feedback analysis

Analysing the feedback provided by users with respect to a given service is fundamental for user profiling and service personalization. In this work, we consider the mobility service offered by MTP integrated in the ITS via ATIS, as explained in the reference scenario presented in Section III.

In this case, the explicit feedback provided by the user \( u \) is the Selected Travel Solution \( s^*(Q(u)) \) actually chosen by the user \( u \) in the set \( TS(Q(u)) \) of Candidate Travel Solutions returned by the MTP. We are interested in defining a finite number \( K \) of User Profiles describing general, but recurrent patterns of user behaviours when users choose \( s^*(Q(u)) \) in \( TS(Q(u)) \). In order to identify the set of possible User Profiles, we consider a pattern of the form

\[
I(Q(u)) := [d_{i,a,b}, UC(u,a,b,t_a), UP(u,a,b,t_a), ocr(s^*(Q(u)))]^T
\]

We consider that for each user request, i.e., for each interaction between an user \( u \) and the ExITS producing a User Query \( Q(u) \), a record \( R(u) \) (1) is stored in the Travel Knowledge Base and a pattern \( I(Q(u)) \) can be used for user feedback analysis by the User Profile Identification. The user feedback analysis consists of a partitional clustering procedure that, given a set \( HP \) of historical patterns \( I(Q(u)) \), yields a partition \( \Pi(HP) = \{W_1, ..., W_K\} \) of \( K \) non empty subsets \( W_i \subset HP \) such that \( \cup_{i=1,K} W_i = HP \) and \( W_i \cap W_j = \emptyset \) for each \( i, j = 1, ..., K \) and \( i \neq j \). The \( K \) components of partition \( \Pi(HP) \), also called clusters, represent groups of similar patterns with respect to an inter cluster separation criterion or, alternatively, an intra cluster homogeneity criterion ([38]). Both approaches rely on a priori selected distances (typically \( l_1 \)-norm metrics, e.g. [39]) measuring the inter cluster similarity or the intra cluster dissimilarity, respectively. According to the criterion to optimize, a number of partitional clustering algorithms have been proposed in the literature. Some algorithms are indicated in case of numerical attributes describing the patterns to be clustered, others are more suitable when dealing with mixed attributes. The most of clustering procedures consider the number \( K \) of clusters as input parameter to the procedure (e.g., \( k \)-means, see [40] and [41]). There exist approaches not requiring the number of clusters as input parameter (e.g., Clique Partitioning Problem, see [42]). The selection of the
suitable partitional clustering algorithm depends on the characteristics of the user profiling problem.

In this work, we consider patterns \( I(Q(u)) \) including all numeric attributes (see Table 1), and adopt the \( k\)-means algorithms for selecting the optimal partition in \( k \) clusters by minimizing the so-called Sum of Squares Error (SSE), i.e. the sum of the squared Euclidean distances between each pattern \( I(Q(u)) \) and the cluster centroid. This choice is particularly suitable, since the Euclidean distance allows capturing the differences between different patterns, once their entries are normalized (in our case, we consider a normalization in [0,1]). The selection of parameter \( k \) of the clustering algorithm is a key feature of our approach. In fact, even if the clustering procedure is completely unsupervised, we deal with the problem of selecting the suitable number \( K \) of clusters by analysing the results of the clustering obtained with different values of the input parameter \( k \) and by selecting the value \( K \) corresponding to the minimum number of distinct optimality criteria \( ocr(s^*(Q(u))) \) represented in each cluster. In such a way, similar users belong to same cluster \( W_i \) and share the same user profile given by the centroid of \( W_i \). The user profile is characterized by the most representative optimality criteria appearing in the cluster \( W_i \) that are in a restricted number of meaningful user profile based optimality criteria. The representative optimality criteria in cluster \( W_i \) are the ones returned as the set \( POC(Q(u)) \) of Personalized Optimality Criteria by the on-line User Profiling module for all \( Q(u) \) such that \( I(Q(u)) \) is similar to the centroid of the cluster \( W_i \). In this sense, we say that Personalized Optimality Criteria returned by the User Profiling module are personalized.

The output of the off-line User Profile Identification module is the set \( \Pi(HP) = \{W_1, ..., W_K\} \) of clusters of homogeneous patterns \( I(Q(u)) \). For each cluster \( W_i \), the centroid of the cluster is given by the mean values of entries of patterns \( I(Q(u)) \) belonging to \( W_i \). In the following, for each \( k = 1, ..., K \), we indicate by \( prl_k \) the sub vector

\[
prl_k = [d_0(a,b), UC(u,a,b,t_a), UP(u,a,b,t_a)]^T
\]  

including the mean values of entries \( d_0(a,b), UC(u,a,b,t_a) \) and \( UP(u,a,b,t_a) \) of all patterns \( I(Q(u)) \) belonging to \( W_i \). It is important to remark that a single user can belong to different clusters. In fact, if two User Queries are assigned to distinct clusters, then the user who submitted the queries behaved differently and, accordingly, his behaviours have been assigned to distinct groups in \( \Pi(HP) \).

C. User profiling procedure

Given the output of the off-line User Profile Identification, namely the set \( \Pi(HP) = \{W_1, ..., W_K\} \) and the related sub vectors \( \{prl_1, ..., prl_K\} \), we denote by \( J(Q(u)) \) the sub vector of the pattern \( I(Q(u)) \) given by

\[
J(Q(u)) := [d_0(a,b), UC(u,a,b,t_a), UP(u,a,b,t_a)]^T
\]  

Once a new User Query \( Q(u) \) is submitted by a user \( u \) to the MTP, the on-line User Profiling procedure evaluates the cluster \( prl(u) \) whom \( Q(u) \) belongs to. In order to do that, the Euclidean distance between the sub vector \( J(Q(u)) \) and each sub vector \( prl_k \) of the \( K \) centroids is calculated. \( Q(u) \) is assigned to the cluster whose centroid is the closest to \( J(Q(u)) \). In this sense, the User Profiling module implements the function \( prl(u) \rightarrow \{prl_1,prl_2, ..., prl_K\} \) by assigning to the user \( u \) the cluster \( prl_u \) obtained as follows

\[
k^* := \arg \min_{k=1,...,K} d_{l_2}(J(Q(u)), prl_k)
\]  

In order to provide the Personalized Optimality Criteria \( POC(Q(u)) \subseteq OCR \), the User Profiling module implements the function \( prs: \{prl_1,prl_2, ..., prl_K\} \rightarrow \{0,1\}^T \) that identifies the subset \( POC(Q(u)) \) of the \( P \) possible Optimality Criteria being suitable for a given User Profile \( prl_k \) for \( k = 1, ..., K \). In order to do that, given the User Profile \( prl(u) \) assigned to User Query \( Q(u) \) by (4), the function \( prs(prl(u)) \) returns the subset of \( OCR \) being mostly represented in cluster \( prl(u) \) in terms of high percentage of patterns \( I(Q(u)) \) sharing the same selected optimality criterion \( ocr(s^*(Q(u))) \).

VI. CONCLUSION

In this work, we presented a Future Internet (FI) based framework for a closed-loop, User Centric Control System able to identify user preferences and behavioural profiles in terms of pre-trip selection of multimodal travel solutions in a generic Intelligent Transportation System (ITS). The automatic identification process takes into account feedback that current ITSs get from the user but typically do not take into account. The proposed approach is aimed at combining advances ICT technologies, Machine Learning and Expert Systems in order to control multimodal mobility services and to make them completely user centric by closing the loop between user request and user feedback with respect to specific services. In this sense, we propose an unsupervised machine learning based approach to extend ITSs by exploiting and then improving the potentiality of the above mentioned closed-loop FI architecture. Future work will consider an in-depth investigation of the proposed approach and its integration in multimodal trip planning algorithms considering new trends and challenging aspects in the field of multimodal transportation such as car sharing, pooling and on-trip electric vehicle charging (e.g., [43], [44]).

ACKNOWLEDGMENT

This paper is based on the work performed by the authors in the framework of the BONVOYAGE project (From Bilbao to Oslo, intermodal mobility solutions and interfaces for people and goods, supported by an innovative communication network), funded by European Commission under the Grant Agreement n° 635867. The authors wish to thank Prof. Nicola Blefari Melazzi and all the partners of the BONVOYAGE project for the constant and constructive cooperation and helpful suggestions.
