Context Awareness in the Travel Companion of the Shift2Rail Initiative*

Alireza Javadian Sabet‡, Matteo Rossi⋆,
Fabio A. Schreiber‡, and Letizia Tanca‡

‡Dipartimento di Elettronica, Informazione e Bioingegneria
⋆Dipartimento di Meccanica
Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
{firstname.lastname@polimi.it}

Discussion Paper

Abstract. Providing personalized offers, and services in general, for the users of a system requires perceiving the context in which the users' preferences are rooted. In this work, we introduce the use of an already known model and methodology – based on the so-called Context Dimension Tree – along with a conceptual architecture to build a recommender system that offers personalized services for travelers. The research is performed in the frame of the Shift2Rail initiative as part of the Innovation Programme 4 of EU Horizon 2020.

Keywords: Context Dimension Tree · Preferences · Journey Planning · Data Tailoring · Recommender Systems.

1 Introduction

The demand for systems that provide *personalized* services increases the need to extract knowledge from different sources and appropriately reshape it. Besides, services cannot be properly adapted just by considering the static information obtained from the users' profiles: using instead a combination of such profiles with the context in which the user is going to be served is definitely more realistic. Generally speaking, *Context* can be recognized as a set of features (*i.e.* values for variables) contributing to the decisions of a user in a system [4].

This work explores and presents the essential elements vital to design a user-centered *Recommender System* (RS) for the *Travel Companion* (TC) module, currently being developed within the Shift2Rail (S2R) initiative as part of the Innovation Programme 4 (IP4). TC acts as an interface between users (typically travelers) and the other modules of the S2R IP4 ecosystem, supporting the

*This work was supported by Shift2Rail and the EU Horizon 2020 research and innovation programme under grant agreement No: 881825 (RIDE2RAIL)

users in all steps of their travel. Precisely, since supporting context-dependent data and service tailoring is paramount to ensure personalized services, we aim at extending the work carried out in [1] on the *Traveler Context-aware User Preferences*, by designing the Traveler Context Dimension Tree (TCDT) and the conceptual system architecture that identifies the essential components dealing with the creation and management of travelers' preferences. The rest of the paper is organized as follows. Section 2 discusses background; Section 3 explains the proposed methodology; Section 4 concludes the work.

2 Background and Related Work

Recommender Systems: RS are designed to recommend items to users considering their needs. The *user profiling* approaches, that emerged in the literature with the aim to determine the users' requirements and behavioral patterns [10], fall into one of the following categories: *Explicit* approaches, often referred to as *static user profiling*, predict the user preferences and activities through data mostly obtained from filling forms; *Implicit* approaches, instead, mostly disregard the users' static information and rely on the information obtained from observing their behaviors; *Hybrid* approaches are a combination of the two [12].
Context-awareness: Various models for designing context-aware systems have been described in many surveys [2][8]. The work [3] introduced the Context Dimension Tree (CDT) model – and an associated methodology – aimed at representing, and later exploiting, the information usage of *contexts*, in order to capture different situations in which the user can act, and formalize them hierarchically as a rooted labeled tree. Among the various advantages of the CDT model, its flexibility to capture the context both in the *conceptual* and *detail* level pursued us to utilize it for the sake of this work. An example of a CDT is depicted in Figure 1; the root of the tree represents the most general context, N is the set of nodes, which are either black *dimension nodes* N_D or white *concept nodes* N_C *a.k.a* dimension's values; N_D and N_C must alternate along the branches. White squares are *parameters*, shorthands to represent the nodes that have many possible values. Dashed lines specify if the N_C children of N_D are mutually exclusive. The children of r define the main analysis dimensions and are known as *top dimensions*. Each N_D should have at least one N_C.

3 Traveler Context Dimension Tree

This section explores the travelers' contexts and preferences through the CDT methodology, and introduces a conceptual system architecture that includes the main components for the ranking of the trips according to the travelers' context.

To enable context-aware recommendations for travel purposes, we identified the aspects characterizing contexts which correspond to the TC users' choice criteria that are potentially useful to score the available trips (see Figure 1).

Note that, in the application design phase, designing a CDT is performed independently of, yet in parallel with, the other routine activities involved in this phase [5]. The modeling mechanism of the TCDT intends neither to model all the available data and their structure nor how they are acquired and where they are stored; rather, it models the information that constitutes the various contexts in which the travelers may find themselves during their *reservation and travel experiences*; this information is potentially useful for supporting the system in understanding and seconding the users' preferences. The user variable *Name* is an example: the TCDT does not include it since it does not vary with the user's context; unless one might decide to use it to estimate the user gender. Note that it is common practice to employ *feature engineering* techniques to transform the dataset's feature space, improving the performance of the predictive models [11]. The ultimate purpose of using the CDT model is thus to support the analysis of such domain knowledge.

Practically, the CDT design is done iteratively, and depends on the final requirements of the application. The TCDT we propose defines the fundamental dimension and concept nodes in such a way that any further interesting features can be added by increasing or decreasing the level of granularity of the model.

3.1 Main Dimensions and Concepts

In the TCDT of Figure 1, the `Personal Data` N_D captures the socio-economic characteristics of the users through the `Profile` N_C. Different groups can be extracted according to the values of the `socio-economic factor` concept node, such as geographical origin, profession, and so on. Each group carries a set of preferences that are rather stable, thus can be associated with the notion of *user profile*. The main motivations for this dimension are to enable a warm start for the system and also to provide the chance of detecting and possibly supporting some group behaviors. As an example, consider two regions `X` and `Y` in the category of geographical groups, and suppose that, for some reason, users from `X` tend to choose eco-friendly travels much more frequently than users from `Y`. Investigating the reasons behind this tendency enables the authorities and the system to take the required actions (if applicable) for increasing the popularity of eco-friendly travels for the users living in region `Y`.

Moreover, a person can be member of some communities. The `Memberships` N_C captures the memberships of the user along with their payment methods. The `Loyalty Cards` N_D captures membership of the user in a community that potentially provides specific discounts. Moreover, we introduced N_D—`Other Memberships`—as a placeholder to capture other, less structured, communities that may follow different patterns compared to those based on `Loyalty Cards`. Tailoring the `Memberships` N_C with more levels of granularity through a combination of domain experts' knowledge and machine learning approaches like *clustering* is one of our future works.

We put a particular emphasis on supporting the needs of people with disabilities and health-related issues (HI), dedicating the `Health Issues` N_D to this purpose. The `HI_Category` N_D distinguishes the case that the issue belongs to the `PRM` category, which stands for *Person with Reduced Mobility*. The descendants of `HI_Type` N_D list some of the most critical issues. Moreover the TCDT

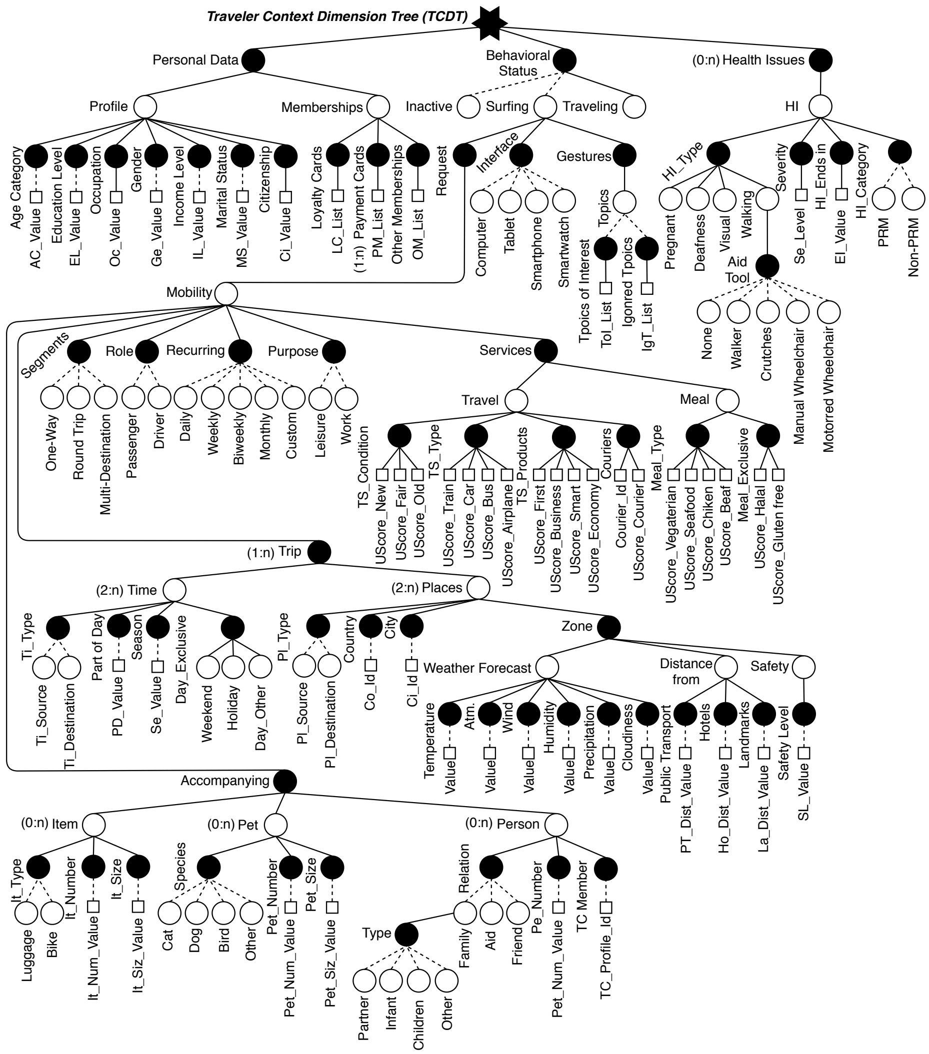

Fig. 1. Proposed Traveler Context Dimension Tree. Black circles represent dimension nodes and white circles and squares represent concept nodes. Dashed lines indicate the mutually exclusive concepts.

increases the level of granularity for the `Walking` N_C through the `Aid Tool` N_D to exemplify the concepts which should be considered during the future expansion. Knowing this concept is essential because of the particular space each of the `Aid Tools` require while recommending trips. The same importance is also applied to the *Severity* N_D, which can potentially limit the travel choices. Lastly, `HI_Ends in` N_D, enables the TC to determine if the issue is permanent or temporary.

Among the other top dimensions, `Behavioral Status` captures the current situation of the user as follows: the `Traveling` concept captures the state in which the user is traveling, or has purchased a *travel offer* and is waiting for the upcoming trip. Obviously, its two sibling N_C, drawn with dashed lines, are mutually exclusive: the `Inactive` N_C is true if the user is not interacting with the TC, while the `Surfing` N_C incorporates both implicit and explicit momentary user behaviors while interacting with the TC. More precisely, the `Interface` and `Gestures` N_D capture implicit behaviours: while users are interacting with the TC through a `Computer`, since there is extra visual space, and presumably the user has no urgent travel request, the TC proposes information regarding “eco-friendly” offers, which have lower CO_2 emissions. The users may decide to click, scroll, or ignore this information, which in turn can provide useful insights about their preferences regarding eco-friendly offers.

Explicit behaviours, instead, are captured via the `Request` N_D. Eco-friendly traveling behaviors can be promoted through so-called ride-sharing. For this reason, we foresee that when the user requests a travel offer through the TC and driving a car is a possibility, they can specify whether their `Role` is that of *Driver* or of *Passenger*, as well as, their `Purpose` and preferred `Segments`.

The user provides the locations that they are going to visit (at least source and destination). In the TCDT this value is transformed into appropriate concepts such as `Country`, `City` and `Zone`. For example, the `Zone` N_D enables the TCDT to capture the factors contributing to the user's decision through its children nodes *i.e.* `Distance from Public Transportation` services, `Hotels` and `Landmarks` (`PT`). Moreover, the `Weather Forecast` N_C is used to capture weather information according to the `Time` when the user will be in that `Zone`. The same strategy is applied to transform the actual value of the requested departure and arrival times to the `Time` N_C and its descendant nodes.

It may happen that the user has some `Accompanying Items` (e.g., a bike) and `Pets`, whose characteristics—such as their *Type*, *Species*, *Size* and *Number* —should be considered when recommending trips. Also, accompanying `Persons` not only affect travel choices from the logistic aspect but, if the `Person` is also a user of the TC, their preferences should be considered for recommending trips.

The `Services` N_D encompasses the variety of `Travel`- and `Meal`-related preferences according to the *optional* user-provided scores. Precisely, the TC, through *scores*, enables its users to provide scores between 0 and 5 representing excluded and requested respectively for the available services. Beside the preferences obtained by analyzing the history of the user's choices, the services scored as 0 allow the system to filter out such *travel offers*.

3.2 System Architecture

Figure 2 shows the conceptual architecture envisaging three main actors, namely *End Users*, *Travel Companion*, and *Third Parties*, along with their main elements to learn the users' preferences and recommending the best travel options accordingly. It provides an overall view, hiding the details of all the TC's blocks and functions; also, it does not specify the modules' physical locations (cloud, *etc.*).

Vitally, since the system deals with sensitive data, best practices concerning issues related to security and individual privacy through technical solutions [7] such as security protocols and algorithms should be taken into account.

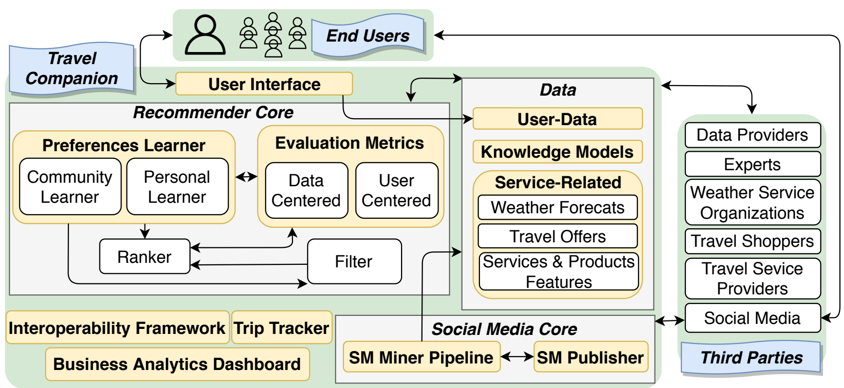

Fig. 2. Conceptual system architecture showing the main elements.

Naturally, the **Travel Companion** is, for our purposes, the most important actor, for which we provide a further breakdown into modules. The *Knowledge Models* block is useful to have a warm start for the newly-registered users, and for whose behavioral activities and specific preferences the system does not have any prior data. It also provides the opportunity to investigate the behavioral drifts that happen for the person, groups, and communities, and to acquire initial information for logistic purposes. The *Social Media (SM) Core* is composed of the two main blocks *SM Miner Pipeline* and *SM Publisher*. The *SM Miner Pipeline* employs Natural Language Processing techniques to harvest knowledge from SM platforms, seeking explicit travel-related mentions and keywords. The *SM Publisher* enables the TC to publish tailored news, promotions, and responses to specific users on SM, and allows users to share their trip information and other socializing functionalities. *Recommender Core* elements take as input user data, knowledge models, and service-related information and accordingly provide a ranked list of the trips for the user. The TC uses the S2R *Interoperability Framework* of the Sprint Project [13] and its rich domain ontology to facilitate the exchange of information between TC and other modules through the automatic concept mapping, both semantically and technologically [6,9]. After the trip plan is finalized, the *Trip Tracker* provides proper notifications about the trip (*e.g.*, disruptions), which include information explicitly required by the user or info deemed useful according to the implicitly learned preferences. Finally, the *Business Analytics Dashboard* tracks the system performance according to various KPIs and provides a platform to observe the trends and behavioral drifts.

The main **Third parties**, playing different roles about the provision of information and services to the TC, are the following: (i) *Data Providers* serve various information about the user, service, *etc.*, e.g. data about the safety of zones; (ii) *Experts* from different domains like sociology, transportation, *etc.* provide and modify the knowledge models of the TC; (iii) *Weather Service Organizations* provide weather forecasts associated with the places; (iv) *Travel Shoppers* (TS) are the organizations and services that arrange the journeys; (v) *SM* can play two primary roles: they can collect data regarding the users' attitudes and preferences, and the TC can use them to publish tailored news and promotions.

3.3 Trip Recommendations

The *travel offers* received by the TC from TS contain variables describing their characteristics (*e.g.* duration, CO_2 emissions, *etc.*). As a first step in the recommendation of *travel offers* to the user, the *Filter* block (see Figure 2) hides some of them according to the knowledge provided by the values associated with specific TCDT dimensions that are stronger preferences and act as a kind of personal constraints—*e.g.* offers that include TSPs with `Services` which users already excluded by scoring them 0. Then, the *Ranker* receives the remaining *travel offers*, plus a vector of preferences containing the *weights* capturing the importance of the TCDT values to the user and to divers communities and groups. For each received *travel offer*, the *Ranker* computes a numerical score in the interval $[0, 1]$ according to suitable evaluation metrics and uses this score to rank offers.

For example, consider a pregnant traveler, traveling with her partner for leisure; for the trip, she has excluded a specific type of meal by giving 0 score. Besides, through knowledge models, TC knows that leisure travels, have higher score for economy products (assuming that neither she has provided the optional scores for any of the products nor the TC has any history regarding her preference in leisure travels). Among the *travel offers* provided by the TS, TC filters out those that include the excluded meal. Considering her current context, the weight of her pregnancy condition potentially might be higher than that of her possible preference for economical offers, thus a *travel offer* that includes a direct flight will take over indirect flights (even though more economical), and, due to the accompanying partner, two seats next to each other might be more scored than seats with more space but in different places (even though more comfort).

4 Conclusion and Future Work

The design of an advanced learning system for travelers' preferences should not only provide the best possible rankings (and, consequently, suggestions) for travel offers, but should also adapt to changes in the behaviors and preferences of users.

The latter requirement is of great importance because preferences are highly dynamic and prone to changes according to different contexts.

In this work, we proposed a methodology to describe, at the conceptual level, the different contexts in which travelers can find themselves, with the advantage of being able to specify, for each traveler, how their preferences are affected by context changes. The methodology consists, on one side, in representing the characteristics of users, services and specific circumstances by means of a TCDT, and, on the other side, in designing a system architecture that identifies the potential sources of data and the interactions among the various system elements.

We are currently working on enriching the proposed TCDT by increasing the dimensions' level of granularity to explore the other contexts whose characteristics can contribute to the users' preferences and to their traveling decisions.

References

1. D5.2 – travel companion specifications. Technical report, IT2Rail Consortium, 2017. available from http://www.it2rail.eu.
2. Cristiana Bolchini, Carlo A. Curino, Elisa Quintarelli, Fabio A. Schreiber, and Letizia Tanca. A data-oriented survey of context models. *ACM Sigmod Record*, 36(4):19–26, 2007.
3. Cristiana Bolchini, Carlo A. Curino, Elisa Quintarelli, Fabio A. Schreiber, and Letizia Tanca. Context information for knowledge reshaping. *International Journal of Web Engineering and Technology*, 5(1):88–103, 2009.
4. Cristiana Bolchini, Carlo Aldo Curino, Giorgio Orsi, Elisa Quintarelli, Rosalba Rossato, Fabio A Schreiber, and Letizia Tanca. And what can context do for data? *Communications of the ACM*, 52(11):136–140, 2009.
5. Cristiana Bolchini, Elisa Quintarelli, and Letizia Tanca. Carve: Context-aware automatic view definition over relational databases. *IS*, 38(1):45–67, 2013.
6. Alessio Carenini, Ugo Dell'Arciprete, Stefanos Gogos, Mohammad Mehdi Pourhashem Kallehbasti, Matteo Rossi, and Riccardo Santoro. ST4RT – semantic transformations for rail transportation. In *TRA 2018*, pages 1–10, 2018.
7. George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. Privacy and data protection by design-from policy to engineering. *preprint arXiv:1501.03726*, 2015.
8. Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim. Context-aware systems: A literature review and classification. *Expert Systems with applications*, 36(4):8509–8522, 2009.
9. Marjan Hosseini, Safia Kalwar, Matteo Rossi, and Mersedeh Sadeghi. Automated mapping for semantic-based conversion of transportation data formats. In *Proceedings of the International Workshop On Semantics For Transport (Sem4TRA)*, volume 2447 of *CEUR-WS*, pages 1–6, 2019.
10. Sumitkumar Kanoje, Sheetal Girase, and Debajyoti Mukhopadhyay. User profiling trends, techniques and applications. *arXiv preprint arXiv:1503.07474*, 2015.
11. F. Nargesian, H. Samulowitz, U. Khurana, Elias B Khalil, and Deepak S Turaga. Learning feature engineering for classification. In *IJCAI*, pages 2529–2535, 2017.
12. Danny Poo, Brian Chng, and Jie-Mein Goh. A hybrid approach for user profiling. In *36th Annual HICSS, 2003. Proceedings of the*, pages 9–pp. IEEE, 2003.
13. M Sadeghi, Petr Buchníček, A Carenini, O Corcho, S Gogos, M Rossi, R Santoro, et al. Sprint: Semantics for performant and scalable interoperability of multimodal transport. In *8th Transport Research Arena TRA 2020*, pages 1–10, 2020.