Towards Learning Travelers' Preferences in a Context-Aware Fashion

A. Javadian Sabet[‡], M. Rossi[⋆], F. A. Schreiber[‡], and L. 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}

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. Accordingly, context modeling is becoming a relevant issue and an expanding research field. Moreover, the frequent changes of context may induce a change in the current preferences; thus, appropriate learning methods should be employed for the system to adapt automatically. In this work, we introduce a methodology based on the so-called Context Dimension Tree—a model for representing the possible contexts in the very first stages of Application Design—as well as an appropriate conceptual architecture to build a recommender system for travelers.

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

1 Introduction

The demand for the 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 (*a.k.a.* variables) contributing to the decision of a user in a system [5].

One of the novel and challenging characteristics of Intelligent Transportation Systems (ITSs) is the real-time personalization of the information considering the user *(i)* requirements, *(ii)* preferences, and *(iii)* behavioural profile [7].

This work investigates and presents the essential elements required to design a user-centered recommender system 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 the users (typically travelers) and the other modules of the S2R IP4 ecosystem, supporting the users in all steps of their travel. More 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 *Traveler Context-aware*

User Preferences, by designing the Traveler Context Dimension Tree (CDT) and the conceptual system architecture that identifies the essential components dealing with the creation and management of travelers' preferences. Notice that the CDT discussed in Section 3.1 is a proof of concept and is not meant to be a complete real-life system. The rest of the paper is organized as follows. Section 2 discusses some related work and necessary background; Section 3 explains the proposed methodology, and Section 4 contains some discussions and future works.

2 Background and Related Work

This section discusses different trends in the fields of recommender systems and context-aware models.

Recommender Systems Recommender Systems are designed to help users fulfill their needs by recommending them appropriate items. Different *user profiling* approaches emerged in the literature, with the aim to determine the users' requirements and behavioral patterns. Each approach falls into one of the so-called *Explicit*, *Implicit* or *Hybrid* categories. *Explicit* approaches, often referred to as *static user profiling*, predict the user preferences and activities through data mostly obtained from filling forms, and do not consider the context. *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 other two [11].

Context-aware Models The demands and models for designing context-aware systems have been described in many works [2][3] [14] . Bolchini *et al.* [4] introduced the Context Dimension Tree (CDT) model (and associated methodology), aimed at representing and later exploiting the information usage *contexts* to capture different situations in which the user can act, and formalize them hierarchically as a rooted labeled tree $\mathcal{T} = \langle N, E, r \rangle$. An example of a CDT is depicted in Figure 1, thoroughly explained in the next sections: r is the root of the tree, which represents the most general context, and N is the set of nodes, which are either dimension nodes N_D (black circles) or concept nodes N_C *a.k.a* dimension's values (white circles). For further analysis of the dimension and concept nodes, it is possible to add one or more parameter nodes (white squares) that characterize their parent node. The children of r should be dimension nodes, which are known as *top dimensions*. They define the main analysis dimensions. Each dimension node should have at least one concept or a parameter node.

Dimension nodes should not directly generate dimensions; that is, they cannot have immediate descendants that are dimension nodes themselves. Similarly, concept nodes cannot directly generate other concept nodes. If they are not followed by any parameter node, they represent a Boolean value, and in case of continuous values or a large number of values, they are followed by suitable parameter node(s).

3 Methodology

In this section we explore the travelers' preferences through the CDT methodology, and design a conceptual system architecture that includes the main components for the ranking of the proposed trips according to the travelers' context.

3.1 Traveler Context Dimension Tree (TCDT)

To enable context-aware recommendation for the travel purposes, we identified the aspects characterizing contexts which correspond to the TC users' choice criteria that are potentially useful to score the available trip choices. Figure 1 presents the proposed Traveler CDT (TCDT).

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 [6]. The modeling mechanism of the TCDT neither intends 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, information potentially useful for supporting the system in understanding and seconding the users' preferences. Consider the user variable *Name* as an example; the TCDT does not include it because it does not vary with the user's context; however, it might be useful within the user profile because one might decide to use it to estimate the user gender.

To improve the performance of the predictive models, it is a common practice to employ different *feature engineering* techniques to transform the dataset by transforming its feature space [12]. When designing the TCDT of Figure 1, we have employed the expansion and transformation of some raw features into different context dimensions, so that the same feature plays different roles in the TCDT to provide a better understanding of the context. An example is the `Places` dimension, explained in Section 3.2. In practice, the design of a CDT requires an iterative approach, and it is dependent on the final requirements of the application. In the proposed TCDT, we designed the top dimensions in such a way that any further modification can be applied by increasing or decreasing the level of granularity to tailor the TCDT. Moreover, the TCDT encompasses some of the essential primary dimension and concept nodes to pave the way for further investigation and tailoring.

3.2 Main Dimensions and Concepts by TCDT

In the TCDT of Figure 1, the `Profile` dimension captures the socio-economic characteristics of the users, along with their payment methods. 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

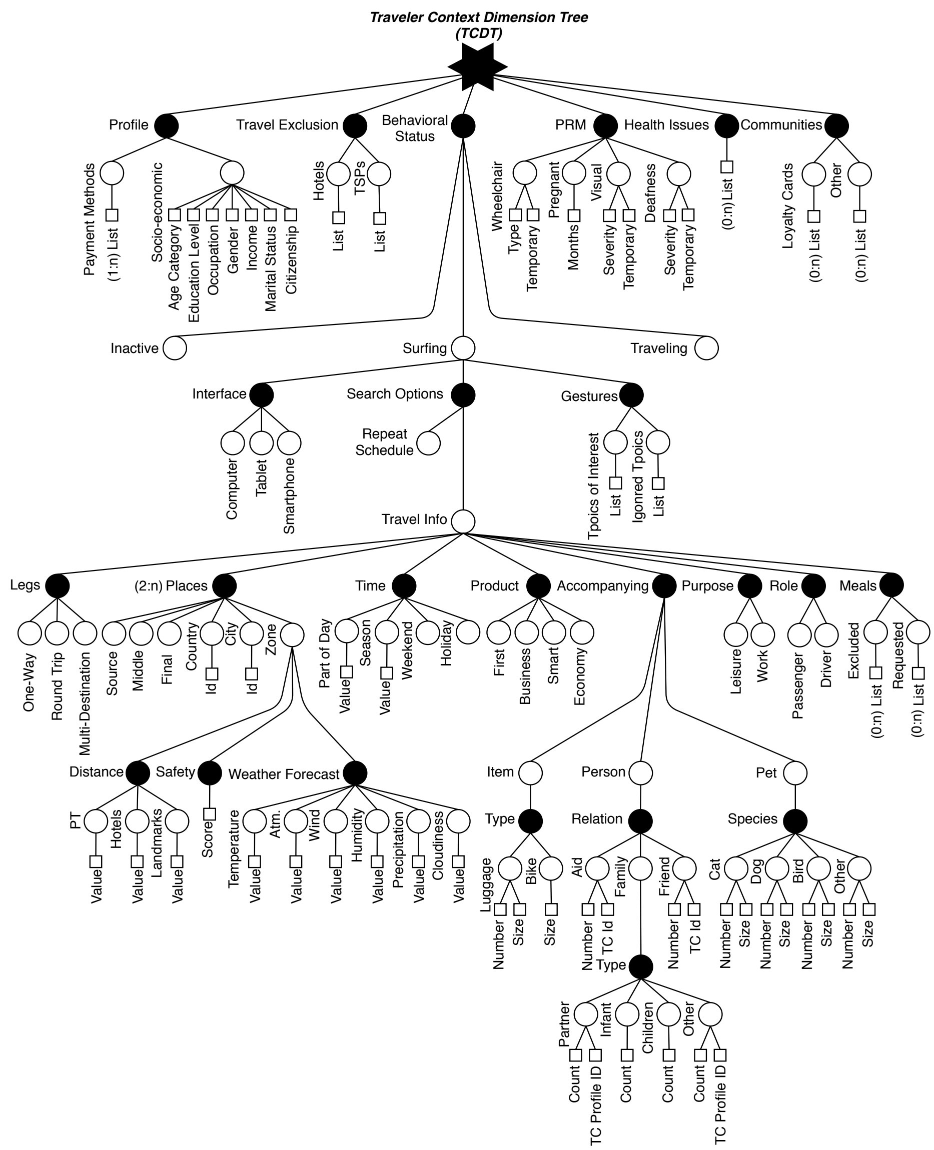

Fig. 1. Proposed Traveler Context Dimension Tree. Black circles, white circles, and white squares represent, respectively, dimensions, concepts and parameters.

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.

Beside the preferences obtained by analyzing the history of the user's choices, the `Travel Exclusion` dimension allows the system to filter out *travel offers* that include features, such as specific hotels and transport service providers (TSPs), that have already been explicitly excluded by the user.

In order to support the needs of people with disabilities and health-related issues, we put a particular emphasis on dedicating a group of dimensions—namely `Health Issues` and `PRM`, which stands for Person with Reduced Mobility–to this problem. The suggested concept nodes for the `PRM` dimension identify some of the most critical mobility issues and their related parameters. For example, by nature, pregnancy is a temporary concept, while the others include a parameter that represents the fact that the reduced mobility situation of the user is temporary or permanent. Another example of the parameters to be taken into account during the expansion stage of the TCDT is the *Type* parameter of the `Wheelchair` concept; possible values for this parameter are *manual* or *motored*. Knowing this concept is essential because of the particular space each of these types requires while recommending trips. The same importance is also applied to the *Severity* parameters, which can potentially limit the travel choices.

A person can belong to a community if they have joined that community. In the TCDT, the `Communities` dimension captures the memberships of the user. The `Loyalty Cards` concept captures membership of the user in a community that potentially provides specific discounts. Moreover, we introduced another concept node—`Other`—as a placeholder to capture other, less structured, communities that may follow different patterns compared to those based on `Loyalty Cards`. Tailoring the `Communities` dimension with more levels of granularity through a combination of domain experts' knowledge and machine learning approaches like *clustering* is one of our future works.

The `Behavioral Status` dimension describes the current situation of the user through three sub-dimensions. The `Inactive` concept is true if the user is not interacting with the TC. The `Traveling` concept, instead, captures the state in which the user is traveling, or has purchased a *travel offer* and is waiting for the upcoming trip. The `Surfing` concept encompasses both implicit and explicit momentary user behaviors. More precisely, the `Interface` and `Gestures` dimensions capture implicit behaviours. As an example, consider a context in which the user is interacting with the TC through a computer; since this interface provides more space for showing information, and potentially may suggest that the user has no urgent travel request, the TC promotes 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 through the `Search Options` dimension.

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*. Through the TC, they may also specify the `Purpose`, `Legs` (One-way, Round Trip, Multi-Destination) and preferred `Product` (first-class trip, second-class trip, etc.) for the trip.

Naturally, the user provides the locations that they are going to visit (at least source and destination). As far as the TCDT is concerned, this value is transformed into appropriate concepts such as `Country`, `City` and `Zone`. For example, consider the `Zone` concept as a representation of the location; it enables the TCDT to capture the factors contributing to the user's decision through its subdimensions, *i.e.*, `Distance` from public transportation (`PT`), `Hotels` and `Landmarks`. In addition, the `Weather Forecast` dimension 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` dimension and its descendant concepts.

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 taken into account when recommending trips. Also, accompanying `Persons` not only affect travel choices from the logistic point of view, but, if the `Person` is also a user of the TC, their preferences should be taken into account.

3.3 Dynamic vs. Static Dimensions

We categorize dimensions as *static* and *dynamic* ones. Among the top dimensions, `Profile`, `Travel Exclusion` and `Communities` are static, since their values are rarely modified (though they can indeed change over time). On the other hand, `Behavioral Status` is a dynamic dimension capturing features that usually refer to a specific moment in time and are not necessarily valid for the future interactions of the user with the TC. Moreover, `PRM` and `Health Issues` are dimensions that can be categorized both as static and dynamic, depending on whether they are permanent or temporary.

To clarify, we illustrate how the interpretation of a situation can change through the following example. For the sake of reporting to the Business Analytics Dashboard (discussed in Section 3.4) or of adapting the preferences learner, the TC needs to query the list of TSPs that are un-favored by users. Indeed, according to the TCDT, this information can be obtained from the `Travel Exclusion` dimension's child—`TSPs` concept node—or from the `Gestures` dimension's child—`Ignored Topics` node. Suppose a TSP with eco-friendly, but comparatively expensive *travel offers* has appeared in the list of `Ignored Topics`. As the latter is a dynamic concept, the TSP should not be considered one that users in general do not favor; indeed, the TSP might have been made less popular by the short-term circumstances (say, "Holidays in a touristic region", in which users might tend to opt for the more economical offers). Consequently, any further action, like updates in the preferences learner, should be temporary. If, on the other hand, the TSP appeared in the list associated with the `TSPs` concept node (which is static), it should be considered as one that is excluded by a group of users, and any decision for it might not be temporary.

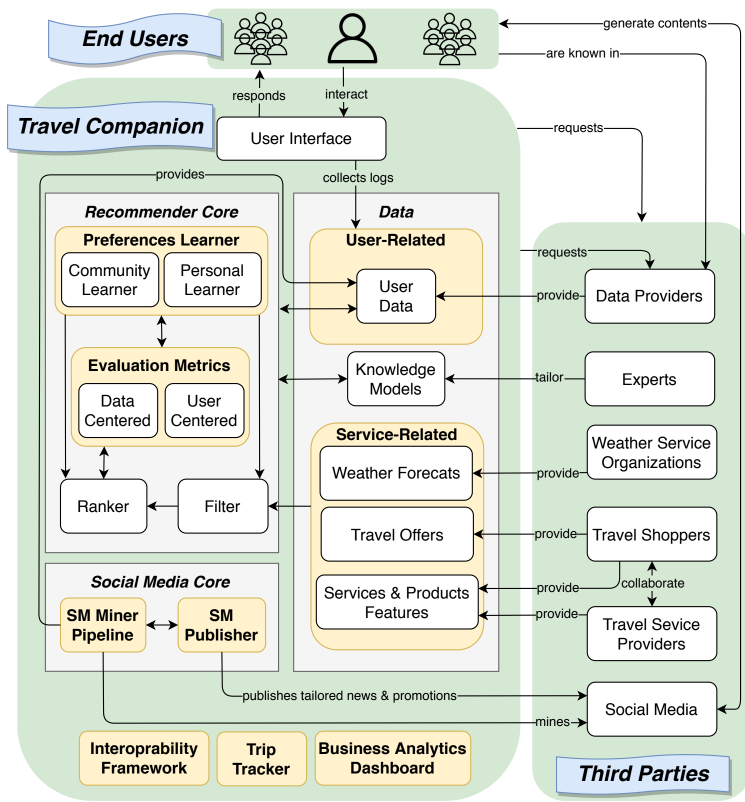

Fig. 2. Conceptual system architecture showing the main elements engaged in learning the travelers' preferences.

3.4 System Architecture

Figure 2 shows the conceptual architecture defining the main blocks and elements required to learn users' preferences and to recommend the best travel options accordingly. Notice that Figure 2 provides only a partial representation of the TC system, and it does not include all TC's blocks and functions; also, it does not specify where modules are deployed (in the cloud, on the client app, *etc.*).

The architecture depicts three main actors, namely *End Users*, *Travel Companion*, and *Third Parties*.

Naturally, **Travel Companion** is, for our purposes, the most important actor, for which in the following we provide a further breakdown into modules. The *Knowledge Models* block is useful to have a warm start for users who just registered to the system, and for which the system does not have any prior data regarding their behavioral activities and specific preferences. Moreover, the *Knowledge Models* block provides the opportunity to study and understand the behavioral drifts that happen for the person, groups, and communities. Another advantage includes acquiring initial information for logistic purposes. *Social Media (SM) Core* is composed of two main blocks, namely *SM Miner Pipeline* and *SM Publisher*. The *SM Miner Pipeline* queries different SM platforms seeking explicit mentions, keywords, and hashtags relevant to TC and travel-related contents; it employs Natural Language Processing techniques to harvest knowledge from those platforms. The *SM Publisher* enables the TC to publish tailored news, promotions, and responses to specific users on SM. In addition, it allows users to share their trip information and provides 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* [13] to facilitate the exchange of information between TC and other modules through the automatic mapping between concepts, both semantically and technologically [8,10]. After the trip planning is finalized, *Trip Tracker* provides appropriate notifications about the trip (*e.g.*, disruptions), which include both information that the user explicitly decided to receive, and information that is deemed useful according to the preferences that are implicitly learned. Finally, the *Business Analytics Dashboard* keeps track of the system performance according to different KPIs and provides a platform for observing the trends and behavioral drifts that are happening.

The main **Third parties**, playing different roles concerning the provision of information and services to the TC, are the following. *Data Providers* provide many different types of information related to the user, service, *etc.* For example, they could provide data regarding the safety of zones. *Experts* from different domains like sociology, transportation, *etc.* provide and modify the knowledge models of the TC. *Weather Service Organizations* provide weather forecasts associated with places that the user will visit. *Travel Shoppers* are the organizations and services which are in charge of planning the journeys. Different *Social Media* platforms can play two primary roles. On one side, they can be employed to collect data regarding the users' attitudes and preferences. On the other hand, TC makes use of them to publish tailored news and promotions.

3.5 Trip Recommendations

Each *travel offer* received by the TC from Travel Shoppers contains a set of variables that describe its characteristics, such as duration, type of vehicle, CO2 emissions, type of seat, and many others.

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 listed in the `Travel Exclusion` dimension.

Then, the *Ranker* block receives the list of remaining *travel offers*, plus a vector of preferences containing the *weights* capturing the importance of the TCDT values to the user and to different communities and groups. For each received *travel offer*, the *Ranker* computes a numerical score in the interval $[0, 1]$ according to some suitable evaluation metrics and uses this score to rank offers.

Note that, among the user's preferences, some explicit choices should be treated differently from the others. For instance, once the user has chosen first class (which is one of the offer categories), the context becomes more precise because now the system knows that the first-class-related offers are to be considered as the most probably chosen. Therefore, this specification acts as another filter that filters out the offers not pertaining to this travel category, or assigns to them a lower weight while scoring the offers.

We illustrate the ranking step through an example. Consider a traveler who is pregnant and who is traveling for leisure accompanied by her husband; for the trip, she has excluded a specific type of meal. Moreover, through preference learning, TC knows that she favors eco-friendly offers. Among the *travel offers* received by the Travel Shopper, TC filters out those that include the type of meal to be avoided. Considering her current context, the weight of her pregnancy condition is higher than that of her preference for eco-friendly solutions. As a result, a *travel offer* that includes a direct flight providing two aisle seats next to each other will have a higher score compared to another with the same characteristics, but one window and one aisle seat separated by a corridor, which is less favorable considering the presence of the accompanying husband.

4 Conclusion and Future Work

The design of an advanced learning system for travelers' preferences should be such that it not only provides the best possible rankings (and, consequently, suggestions) for travel offers, but it should also be capable of adapting to changes in the behaviors and preferences of users. The latter requirement is of great importance because preferences are highly dynamic, and they are prone to changes from time to time 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.

Last, but not least, the recommender system should be able to provide the appropriate exploration-exploitation tradeoff [9], which stems from the observation that due to the lack of information about the existence of offers, users may take actions that might mislead the learning system.

Acknowledgements Work supported by Shift2Rail and the EU H2020 research and innovation programme under grant agreement No: 881825 (RIDE2RAIL)

References

1. D5.2 – travel companion specifications. Technical report, IT2Rail Consortium, 2017. available from http://www.it2rail.eu.
2. Unai Alegre, Juan Carlos Augusto, and Tony Clark. Engineering context-aware systems and applications: A survey. *J SYST SOFTWARE*, 117:55–83, 2016.
3. 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.
4. 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.
5. 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.
6. Cristiana Bolchini, Elisa Quintarelli, and Letizia Tanca. Carve: Context-aware automatic view definition over relational databases. *Inf. Syst.*, 38(1):45–67, 2013.
7. Silvia Canale, Alessandro Di Giorgio, F Lisi, Martina Panfili, L Ricciardi Celsi, Vincenzo Suraci, and F Delli Priscoli. A future internet oriented user centric extended intelligent transportation system. In *2016 24th Mediterranean Conference on Control and Automation (MED)*, pages 1133–1139. IEEE, 2016.
8. Alessio Carenini, Ugo Dell'Arciprete, Stefanos Gogos, Mohammad Mehdi Pourhashem Kallehbasti, Matteo Rossi, and Riccardo Santoro. ST4RT – semantic transformations for rail transportation. In *Transport Research Arena (TRA 2018)*, pages 1–10, 2018.
9. Madalina M Drugan. Reinforcement learning versus evolutionary computation: A survey on hybrid algorithms. *SWARM EVOL COMPUT*, 44:228–246, 2019.
10. 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.
11. Sumitkumar Kanoje, Sheetal Girase, and Debajyoti Mukhopadhyay. User profiling trends, techniques and applications. *arXiv preprint arXiv:1503.07474*, 2015.
12. Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. Data lifecycle challenges in production machine learning: a survey. *ACM SIGMOD Record*, 47(2):17–28, 2018.
13. Mersedeh Sadeghi, Petr Buchníček, Alessio Carenini, Oscar Corcho, Stefanos Gogos, Matteo Rossi, and Riccardo Santoro. SPRINT: Semantics for PerfoRmant and scalable INteroperability of multimodal Transport. In *Transport Research Arena (TRA 2020)*, pages 1–10, 2020. to appear.
14. Gregory Vert, S Sitharama Iyengar, and Vir V Phoha. *Introduction to Contextual Processing: Theory and Applications*. CRC press, 2016.