PREFACE

Special issue on knowledge discovery and user modeling for smart cities

Marcelo G. Armentano[1] · Frank Hopfgartner[2] · Ioanna Lykourentzou[3] · Antonela Tommasel[1]

1 Introduction

User modeling and personalization are commonly used in multiple systems, in which users are characterized based on explicit information about their prior knowledge, behavior, social relations, or preferences, with the aim to adapt a generic system behavior to the particularities of each user. In parallel, the ubiquitous use of social networking sites and mobile and smart devices generates massive data that open opportunities for enhancing and changing the personalization paradigm.

The analysis of the data obtained through the aforementioned sources offers new research opportunities across a wide range of disciplines, including media and communication studies, linguistics, sociology, health, psychology, information and computer sciences, or education. This allows researchers and practitioners to mine and analyze user behavior aiming at discovering knowledge that would allow them to better understand users, and thereby create more accurate models and personalization strategies. At the same time, this influx of new types of data gives rise to the need for further development of innovative methods and approaches to mine this data, and this has important implications in the context of inclusive eGovernment and Smart Cities. In this context, applications could leverage on the mined user models to design and tailor services according to the characteristics and needs of each particular citizen.

The aim of this special issue is to explore recent advances in mining and understanding data generated by citizens, as well as proposing new approaches in regard to how to tackle the new challenges that arise. Such challenges for example include the process of knowledge discovery, the long-term availability of data, the interpretation of user-generated information, ethical and legal considerations, the heterogeneous nature of information, the high volume of available data, and the creation of long-term user models that adapt to the dynamics of life.

The papers received are a well-balanced combination of original contributions in the form of theoretical foundations, experimental and methodological developments, comparative analyses, experiments, and case studies in the field.

In this preface, we first summarize in Section 2 the manuscripts that have been accepted as part of this special issue, and then discuss, in Section 3, the next steps and challenges that lie ahead.

2 Accepted articles

We accepted a total of eight out of the 23 submitted manuscripts as part of this special issue. In the remainder of this section, we summarize the main research challenges addressed in these articles.

Amritha and Sandeep [1] tackle the problem of city crowd and traffic management focusing on the scheduling of dynamic plans for tourists in India, based on the real-time characteristics of traffic and information provided by travel agencies. To that end, the authors propose an architecture of a centralized travel scheduling system for efficiently managing the tourist crowd by distributing and scheduling their travel plan for different sites. Their results showed improvements regarding other state-of-the-art routing protocols.

Casella et al. [2] present an approach to recognize human activities from mobility traces acquired through wearable devices, such as GPS loggers and smartphones. Their approach relies on grammatical inference to construct syntactic models in the form of finite state automata. A similarity measure is also proposed to consider the intrinsic hierarchical nature of such models. This measure enables the identification of

✉ Marcelo G. Armentano
marcelo.armentano@isistan.unicen.edu.ar

[1] ISISTAN (CONICET-UNICEN), Campus Universitario, Paraje Arroyo Seco, Tandil, Buenos Aires 7000, Argentina

[2] University of Sheffield, Sheffield, United Kingdom

[3] Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, Utrecht 3584 CC, The Netherlands

common traits in the paths induced by different activities at various granularity levels. Experiments were conducted in a large metropolitan area to support the proposed approach.

Fernandez et al. [3] study the privacy implications of smart city infrastructures such as transportation and energy networks, which collect and leverage citizens' data in order to adapt services to citizen' needs. According to the authors, current systems try to comply with privacy regulations via anonymization or by using very rigid, hard-coded workflows that have been agreed with a data protection authority. These alternatives can affect the data quality and richness, while diminishing their functionality and potential. In this context, the authors propose an extension of the domain-agnostic SPECIAL policy language, which they apply to different case studies provided by Vienna's largest utility provider. Their extension aims at reducing the semantic gaps between use cases such as the above and the policy language definitions.

Ibrahim et al. [4] present a review of the indoor base station placement problem. The article discusses the parameters that affect the topology of heterogeneous networks and compares existing solutions to the problem of base station layout planning, aiming at improving the coverage of densely populated areas. The authors highlight directions for future work in the area, which includes addressing the problem of interference due to massive Femtocell deployment and the different materials of walls and floors, and minimizing power consumption both in the mobile phones and in the Femtocell base stations.

Jiang et al. [5] address the topic of indoor map construction using crowdsourcing techniques. The paper presents a map construction system able to generate a grammar map based on the integration of the crowdsourcing traces, which include semantic information from user activities, and build a semantic map exploiting Conditional Random Field prediction. Experiments conducted on a floor of a shopping mall of around 10,000 m^2 showed a semantic prediction accuracy varying from 61 to 80% depending on the kind of locations, which in turn depends on the continuity of traces.

Sivanantham and Gopalakrishnan [6] explore energy consumption patterns in smart grids, with the aim to reduce the peak load and alleviate the deviation between the demanded and supplied energy. In this context, the authors propose an optimization-based energy consumption scheme for customers in a smart grid, based on a Stackelberg game. The results showed that the proposed model can reduce peak load and the mismatch between actual load and planned supply, while avoiding the grid overload. According to the authors, the proposed scheduling approach represents an optimal strategy to implement the relationship between the consumer and the service provider in the smart grid environment.

Varona et al. [7] explore how smart technologies can be employed to improve urban infrastructure. In their study, they propose to exploit accelerometer data that have been recorded using mobile phone devices to estimate road surface conditions. Their proposed solution is a convolutional neural network where the input layer is a 3D tensor of the accelerometer data, and the output is a classification of the road surface. In their study, they use real-world data from Tandil, Buenos Aires, Argentina, to categorize road surfaces as either concrete panels, cobblestones, asphalt, or dirt roads. In order to evaluate the performance of their method, they compare it with several suitable baseline algorithms.

Finally, Zinman and Lerner [8] present an analysis of social function of urban areas using digital traces collected in a district of Tel Aviv. The urban area in this study was divided in a grid-like manner, and each cell was labeled with a leveled hierarchy of semantic categories depending on the use (i.e., residential neighborhoods, commercial areas, industrial areas), including different levels of detail resolution. After extracting 158 features from call detail records collected for 62 days, a random forest algorithm was used to classify each cell in the grid. Experiments showed better performance than existing approaches that only consider cellular communications as their data source, although different cities were analyzed in each of these works.

3 Steps ahead

With sensor platforms and data-rich mobile apps becoming mainstream and affordable within our cities, more and more data is generated, which in turn opens many opportunities to exploit this data for the common good. This special issue contains several articles that introduce different case studies to analyze such data. The authors of these articles hail from different continents (i.e., North and South America, Asia, Africa, and Europe), and their case studies reflect a truly global interest in knowledge discovery and user modeling in a smart city context.

The presented case studies can be roughly categorized into two main groups. In the first group, three papers explore how the infrastructure of a smart city can be improved based on the analysis of sensor data. In Ibrahim et al. [4], a case study is presented on how to determine the optimal location for indoor mobile networking base stations. In Sivanantham and Gopalakrishnan [6], the authors aim to identify peak energy usage in a smart grid to optimize energy demand and supply. An approach to identify road conditions based on the analysis of mobile phone accelerometer data recordings is presented in Varona et al. [7]. The second group of papers focuses on how the mobility data gathered by individual citizens can be used to improve the quality of life of these citizens. Among the novelties proposed by this group of papers, we find a method to manage tourist crowds based on real-time traffic and booking information [1], as well as methods to identify social activities in shopping malls [5] and in a larger metropolitan area [2, 8].

These case studies illustrate the multiple potentials that emerge when aggregated user data is captured and exploited to analyze, predict, and influence user behavior. At the same time, however, the existence of such personalized data also causes a threat to the privacy of individuals. One of the case studies of this Special Issue [3] focuses exactly on that exploring the privacy implications of smart city infrastructure.

We believe that finding the right balance between the privacy of individuals and the benefits that smart cities can offer is one of the key challenges that will become increasingly important in the future. In fact, the need to preserve user privacy has a direct impact on how we conduct research in this field. While other research fields often benefit from the release of shared open datasets, the case studies presented in this special issue are all based on unique datasets that cannot easily be shared with others. Apart from limitations to the reproducibility caused by this, this also can be seen as an additional burden for researchers. Instead of relying on well-known existing datasets, the authors have to go through the full process of capturing, cleaning, and annotating datasets that are suitable to study their underlying research questions.

References

1. Amritha AP, Sandeep J (2019) Dynamic route scheduler in vehicular ad hoc network for smart crowd control. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01224-1
2. Casella E, Ortolani M, Silvestri S, Das SK (2019) Hierarchical syntactic models for human activity recognition through mobility traces. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01319-9
3. Fernández JD, Sabou M, Kirrane S, Kiesling E, Ekaputra FJ, Azzam A, Wenning R (2020) User consent modeling for ensuring transparency and compliance in smart cities. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01330-0
4. Ibrahim LF, Salman HA, Taha ZF, Akkari N, Aldabbagh G, Bamasak O (2019) A survey on heterogeneous mobile networks planning in indoor dense areas. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01243-y
5. Jiang Z, Zhang J, Liu C, Huang C (2019) FloorSense: a novel crowdsourcing map construction algorithm based on conditional random field. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01229-w
6. Sivanantham G, Gopalakrishnan S (2019) A Stackelberg game theoretical approach for demand response in smart grid. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01262-9
7. Varona B, Monteserin A, Teyseyre A (2019) A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01234-z
8. Zinman O, Lerner B (2019) Utilizing digital traces of mobile phones for understanding social dynamics in urban areas. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01318-w