
Article:
Armentano, M.G., Hopfgartner, F. orcid.org/0000-0003-0380-6088, Lykourentzou, I. et al. (1 more author) (2020) Special issue on knowledge discovery and user modeling for smart cities. Personal and Ubiquitous Computing, 24 (4). pp. 437-439. ISSN 0949-2054

https://doi.org/10.1007/s00779-020-01430-2

Preface to the Special Issue on Knowledge Discovery and User modeling for Smart Cities

Marcelo G. Armentano · Frank Hopfgartner · Ioanna Lykourentzou · Antonela Tommasel

Abstract Nowadays, multiple applications benefit from user modeling and personalization with different purposes. In the context of smart cities, this is particularly important for improving the citizens' daily experience in many areas such as transportation, traffic, energy consumption, urban infrastructure, leisure, public participation, etc. The ubiquitousness of social networking sites, and mobile and smart-devices offer new information sources, opportunities and challenges for changing the personalization paradigm. The focus of this special issue is to explore how new information sources, including social media and smart devices, can be exploited to provide intelligent services and solutions for the development of smart cities.

Keywords Smart Cities · User modeling · Special Issue

1 Introduction

User modeling and personalization are commonly used in multiple tasks, in which users are characterized based only on explicit information about their prior knowledge, behaviour, social relations, or preferences, aiming at adapting generic systems to the particularities of each user. The massive and 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 by the above mentioned 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 behaviour aiming at discovering knowledge that would allow them to better understand users, and thereby create more accurate models and personalization strategies. Hence, a significant need arises for further development of innovative methods and approaches that are able to mine and deal with such new data sources, which has important implications in the context of

inclusive eGovernment and Smart Cities. In this context, applications could leverage on the user's models to design and tailor services according to the characteristics and needs of each particular citizen.

The aim of this special issue was to explore recent advances in mining and understanding data generated by citizens, as well as how to tackle the new challenges that arise. For example, 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, amongst others.

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 twenty three 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 proposed 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] presented an approach to recognize human activities from mobility traces acquired through wearable devices, such as GPS loggers and smartphones. The 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 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] studied 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 citizens' needs. According to the authors, current systems try to comply with privacy regulations via anonymization or use 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 proposed an extension of the domain-agnostic SPECIAL policy language, which they applied to different case studies provided by Vienna's largest utility provider. Their extension aimed at reducing the semantic gaps between the specific use cases and the policy language definitions.

Ibrahim et al. [4] presented 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 highlighted 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] addressed the topic of indoor map construction by 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's activities, and build a semantic map exploiting Conditional Random Field prediction. Experiments conducted on a floor of a shopping mall of around 10000 m2 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] explored energy consumption patterns in smart grids aiming at reducing the peak load and alleviating the deviation between the demanded and supplied energy. In this context, the authors proposed 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] explored 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 against several suitable baseline algorithms.

Finally, Zinman and Lerner [8] presented an analysis of social function of urban areas using digital traces collected in a district of Tel Aviv. The urban area was divided in a grid-like manner and each cell was labeled with a leveled hierarchy of semantic categories of use (i.e. residential neighborhoods, commercial areas, industrial areas, etc.), including different levels of detail resolution. After extracting 158 features from call detail records collected during 62 days, the 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 the 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, more and more data is generated that leaves many opportunities to exploit this data for a 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.

These presented case studies can roughly be categorized into two main groups that can be described as follows. First of all, three papers explore how smart city infrastructure can be improved based on the analysis of sensor data. In [4], a case study is presented on how to determine the optimal location for indoor mobile networking base stations. In [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 recordings is presented in [7]. Secondly, several papers explore how individuals' mobility data can be exploited to improve citizens' lifestyles. This includes a method to manage tourist crowds based on real-time traffic and booking information [1] and methods to identify social activities in shopping malls [5] and in a larger metropolitan area [2, 8].

These case studies illustrate the manifold 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 individuals' privacy. One of the case studies [3] presented in this special issue acknowledges this by exploring privacy implications of smart city infrastructure.

We argue that finding the right balance between individuals' privacy and conveniences that smart cities can offer is one of the key challenges that will become increasingly important. In fact, the need to preserve users' privacy has a direct impact on how we can 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 the 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] A. P., A., J., S. Dynamic route scheduler in vehicular ad hoc network for smart crowd control. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01224-1

[2] Casella, E., Ortolani, M., Silvestri, S. et al. Hierarchical syntactic models for human activity recognition through mobility traces. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01319-9

[3] Fernández, J.D., Sabou, M., Kirrane, S. et al. User consent modeling for ensuring transparency and compliance in smart cities. Pers Ubiquit Comput (2020). https://doi.org/10.1007/s00779-019-01330-0

[4] Ibrahim, L.F., Salman, H.A., Taha, Z.F. et al. A survey on heterogeneous mobile networks planning in indoor dense areas. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01243-y

[5] Jiang, Z., Zhang, J., Liu, C. *et al.* FloorSense: a novel crowdsourcing map construction algorithm based on conditional random field. *Pers Ubiquit Comput* (2019). https://doi.org/10.1007/s00779-019-01229-w

[6] Sivanantham, G., Gopalakrishnan, S. A Stackelberg game theoretical approach for demand response in smart grid. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01262-9

[7] Varona, B., Monteserin, A. & Teyseyre, A. A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01234-z

[8] Zinman, O., Lerner, B. Utilizing digital traces of mobile phones for understanding social dynamics in urban areas. Pers Ubiquit Comput (2019). https://doi.org/10.1007/s00779-019-01318-w