Mining Urban Data (Part A)[☆]

Ioannis Katakis[1]

National and Kapodistrian University of Athens

Abstract

Modern cities are flooded with data. New information sources like public transport and wearable devices provide opportunities for novel applications that will improve citizens' quality of life. From a data science perspective, data emerging from smart cities give rise to a lot of challenges that constitute a new interdisciplinary field of research. This article introduces the first part of a special issue on the topic 'Mining Urban Data' published in the journal Information Systems.

Keywords: urban data, smart cities, data mining, data management, sensor networks, social networks

We are moving towards a smart city era. The fact that the modern urban environment is flooded with data poses a great number of challenges and opportunities. Smart phones record our trajectories and activities. Buses and vehicles moving around a city store their own paths, delays and issues they come across. Sensors embedded in the streets capture the volume of traffic. At the same time, citizens submit information about where they are, what they do and how they feel in social media. They constitute what is now known as 'social sensors'. Moreover, modern weather stations are transmitting a large volume of weather and environmental information and buildings broadcast data about energy consumption, leaks or security issues.

[☆]A preface to the special issue - Mining Urban Data
URL: `www.katakis.eu` (Ioannis Katakis)

[1]Managing Guest Editor

All these information sources open up multiple opportunities for the development of applications that benefit the citizens' quality of life. Municipalities are already managing traffic by taking advantage of information sent to their headquarters by sensors embedded on the streets or in vehicles. Based on smart city data, citizens will be able to follow not only the shortest or fastest route, but also the most beautiful one, the most environmental friendly, or even the one with the most interesting historical background. Other applications will have significant impacts on health, building design and civic protection.

From a data science perspective, there are a lot of challenges. The toughest obstacles are that such data are heterogeneous, noisy, and unlabeled. In addition urban data can include massive and high-speed data streams (for example video feeds). On top of that, urban data are many times related to human activity that is particularly complex. For such data, privacy and security should also be a concern.

The utilization of urban data requires a lot of *interdisciplinary* effort. The area attracts interest from multiple research communities: Data Mining and Machine learning, Energy and Environmental Science, Optimization, Urban Planning, and Transportation. The readers will observe that this diversity is reflected in the papers published in this special issue.

The purpose of this Special Issue is first to collect the state-of-the-art so that readers can understand and take advantage of different perspectives. Another goal is to disseminate recent research results and attract more people into this exciting and vital for the next generation scientific topic.

The editors received an impressive collection of 44 submissions. This number definitely underlines the emergence of a new interdisciplinary research field. The reader will have the opportunity to study a variety of topics including mobile data management, environmental analysis, social media mining, trip planning and a lot more.

In this Issue. In their article 'Dynamics of urban lightscape', Dobler et al. focus on studying the 'pulse of the city' by applying physics (astronomy) tech-

niques on light images analyzing the patterns of light. They identify recurring sequences and discuss their correlations. Potential applications include emergency response and environmental monitoring.

Next, the special issue includes two papers related to social media. Social networks are an important source of urban data since they are correlated with human activity. Interestingly, both articles focus on utilizing the sentiment expressed in social media. In their article, 'CrowdPulse: A framework for real-time semantic analysis of social streams', Musto et al. present a system that consists of components for semantic processing and sentiment analysis designed to process social streams. They evaluate their system on two scenarios. The fist one is to identify risky areas of the Italian territory based on content analysis. The second one monitors L'Aquila, a city recovering from an earthquake. On a similar direction Vasdeval at al. try to estimate the urgency of urban issues by analyzing emotions that are expressed in messages describing the problems.

The last two papers utilize data related to moving vehicles. In 'Mining moving patterns for predicting next location', Chen et al., predict the next location of vehicles. They evaluate three approaches that are based on Markov Models. Vehicle data are obtained from traffic surveillance cameras in the City of Jinan. Finally, 'Real-time Traffic Incident Detection Using a Probabilistic Topic Model' regards traffic congestion. Traffic models capture the states of normality that aid in identifying anomalies. The analysis is based on car data moving into the Shuto Expressway in Tokyo.

This special issue will continue in a following issue of Information Systems.

Acknolwedgements. We would first like to thank the team of guest editors involved in this special issue: Gennady Andrienko, Dimitrios Gunopulos, Vana Kalogeraki, Katharina Morik, and Olivier Verscheure. We would also like to thank Yannis Ioannidis, area editor of Information Systems, that agreed to team up with us on this endeavor.

This special issue would not be possible without the support and extreme care of Dennis Shasha, editor in Chief of Information Systems. The profession-

alism of the administrative people in Information Systems and Elsevier made everything much easier for everyone. Special thanks to Hilda Xu (Publishing Content Specialist), Jeyalakshmi Ramasaamy (Journal Manager) and Santhoshi Basuthkar (Journal Manager).

Of course, we would like to express our gratitude to the numerous reviewers that agreed to evaluate the papers of the special issue. Their detailed review enabled the significant improvement of the submitted manuscripts.

The Mining Urban data Workshops were the first steps leading to this special issue. The first one was co-located with EDBT/ICDT in Athens in March 28 (`www.insight-ict.eu/mud`). The second one was organized in ICML 2015 in Lille France at the 10th of July (`www.insight-ict.eu/mud2`). This special issue was inspired and partially funded by the EU project INSIGHT (`www.insight-ict.eu`).