Front. Comput. Sci., 2017, 11(1): 1–3
DOI 10.1007/s11704-016-6907-2

PERSPECTIVE

Urban computing: enabling urban intelligence with big data

Yu ZHENG[1,2,3]

1 Microsoft Research, Beijing 100080, China
2 School of Computer Science and Technology, Xidian University, Xi'an 710071, China
3 Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Urban computing is a process of acquisition, integration and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings and humans, to tackle the major issues that cities face, e.g., air pollution, energy consumption and traffic congestion. Urban computing connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytics models, and novel visualization methods, to create win-win-win solutions that improve urban environment, human life quality, and city operation systems. Urban computing also helps us understand the nature of urban phenomena and even predict the future of cities. Urban computing is an interdisciplinary field fusing computer science and information technology with traditional city-related fields, like urban planning, transportation, civil engineering, economy, ecology, and sociology, in the context of urban spaces [1].[1)]

Figure 1 presents a general framework of urban computing which is comprised of four layers: urban sensing, urban data management, urban data analytics, and service providing. The following paragraphs discuss main challenges of each layer and key techniques needed.

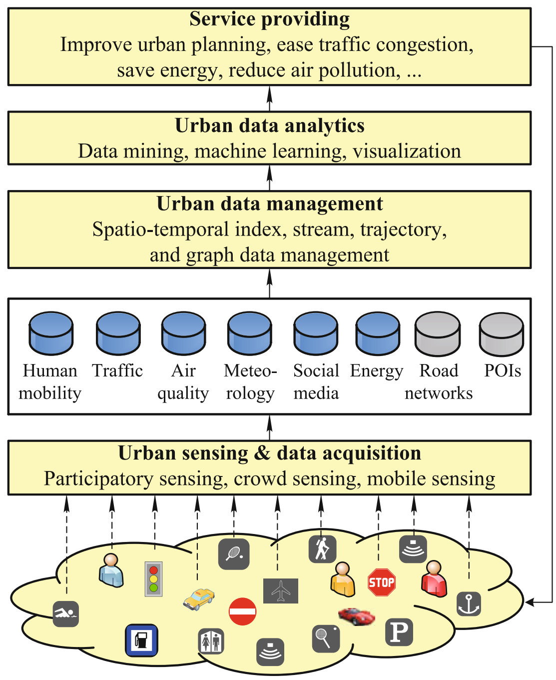

Fig. 1 A general framework of urban computing [1]

- ***Urban sensing*** collects data from different sources through sensors or humans in a city. There are two main urban sensing modes, consisting of sensor-centric sensing and human-centric sensing. The former mode deploys a collection of sensors in fixed locations, e.g., at meteorological stations, or with moving objects, such as buses or taxies. Those sensors continuously send readings to a backend system without involving people in the loop, once they have been deployed. The human-centric sensing mode leverages humans as sensors to probe urban dynamics when they are moving around in cities. The information collected by individuals is then used to solve a problem collectively. The challenges of urban sensing are four-fold: 1) skewed sample data, 2) data sparsity and missing, 3) implicit and noisy data,

E-mail: yuzheng@microsoft.com

1) For the reader's convenience, this paragraph is directly extracted from Ref. [1] to keep the definition consistent

and 4) resource deployment [2].

- ***The urban data management*** layer manages large-scale and dynamic urban data, which is usually associated with a spatial coordinate and a timestamp (i.e., spatio-temporal properties), using cloud computing platforms, indexing structures, and retrieval algorithms. Current cloud computing platforms cannot support spatio-temporal data very well for three reasons: 1) the data structure of spatio-temporal data, e.g., trajectory data [3], is very different from texts and images; 2) Queries on spaito-temporal data, e.g., finding a vacant taxi around me in the past two minutes, are different from key words matches; 3) Because an urban computing application needs to simultaneously consume multiple datasets from different domains, we need to organize different datasets organically and in advance. However, the three techniques are missing on current cloud computing platforms. To address these issues, this layer first devises different storage mechanisms on the cloud for different types of urban data. Second, this layer designs unique indexing structures and retrieval algorithms for spatial and spatio-temporal data as well as hybrid indexing structures for organizing multi-modality data across different domains [4]. Those indexing and retrieval techniques are foundations of upper-level data mining and machine learning tasks. Third, this layer enables some advanced data management functions, such as map-matching, finding the Maximum K-coverage setand dynamic dispatching, which can solve many urban computing problems by themselves.
- ***The urban data analytics layer*** applies a diversity of data mining models and machine learning algorithms to unlock the power of knowledge from data across different domains. This layer *adapts* basic data mining and machine learning models, such as clustering, classification, regression and anomaly detection algorithms, to handling spatio-temporal data's unique properties, consisting of spatial distance, spatial hierarchy, temporal closeness, period and trend. This layer also fuses the knowledge from multiple disparate datasets based on cross-domain data fusion methods [5], such as the Deep Learning-based [6], multi-view-based, probabilistic dependency-based, similarity-based and transfer learning-based data fusion. As many urban computing applications need instantaneous services, it is also important to combine database techniques with machine learning algorithms in a data mining task. Based on the aforementioned components, advanced topics on this layer include filling missing values in spatio-temporal data, predictive models, object profiling, and causality inference.
- ***The service providing layer*** offers interface that allows domain systems to call the knowledge from an urban computing application, through cloud computing platforms. As urban computing is an inter-disciplinary field, the knowledge from data must be integrated into existing domain systems to inform their decision making. In addition, it is imperative to enable interactive visual data analytics [7], which combine human wisdom with machine intelligence by keeping domain experts in a learning loop. In terms of the timing that a service is created for, this layer provides three categories of services, consisting of understanding current situation, predicting future, and diagnosing history. For instance, inferring the real-time and fine-grained air quality throughout a city based on big data belongs to the first category [8], while forecasting air quality over future time is an example of the second category [9]; diagnosing the root cause of air pollution based on data accumulated over a long period belongs to the last category. Based on the domain that a service is created for, the services provided by this layer range from transportation to environmental protection, to urban planning, energy saving, social and entertainment, and public security.

Acknowledgements The work was supported by the National Natural Science Foundation of China (Grant Nos. 61672399 and U1401258), the China National Basic Research Program (973 Program) (2015CB352400).

References

1. Zheng Y, Capra L, Wolfson O, Yang H. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology, 2014, 5(3): 222–235
2. Ji S G, Zheng Y, Li T R. Urban sensing based on human mobility. In: Proceedings of the 18th ACM International Conference on Ubiquitous Computing (UbiComp 2016). 2016, 1040–1051
3. Zheng Y. Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology, 2015, 6(3)
4. Bao J, Li R Y, Yi X W, Zheng Y. Managing massive trajectories on the cloud. In: Proceedings of the 24th ACM International Conference on Advances in Geographical Information Systems (ACM SIGSPATIAL 2016). 2016
5. Zheng Y. Methodologies for cross-domain data fusion: an overview. IEEE Transactions on Big Data, 2015, 1(1)
6. Zhang J B, Zheng Y, Qi D K. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the 31st AAAI