FROM THE EDITORS

A Data-Centric Framework for Cyber-Physical-Social Systems

Bin Guo, Zhiwen Yu, and Xingshe Zhou, *Northwestern Polytechnical University, China*

Cyber-physical systems (CPSs) enable the physical world to merge with the virtual world by integrating computation and physical processes. A CPS facilitates tight integration between computation, communication, and control in its operation and interactions with the environment in which it's deployed. Now, in addition to embracing cyber and physical features, interest is growing in harnessing human and social factors in CPSs. For instance, one recently proposed cyber-physical-social system (CPSS) embraces the cyberspace-enabled parallelism: a real system and its artificial counterparts run in parallel and interactively through cyberspace.[1] Another proposal is a human-in-the-loop CPS, which infers users' intent by measuring human cognitive activity through body and brain sensors.[2] These studies take human and social features as important elements in CPSs mainly from the system automation and control perspective. However, to grasp the fuller potential of a CPSS, data-centric realization is necessary.

Consider an urban big data system that's a typical CPSS system. It integrates and derives information from cyberspace (for example, data from governments and institutions), physical space (surveillance cameras, smartphones, and so on), and social space (such as mobile crowd sensing[3] and mobile social networks). This kind of *data-driven CPSS* (D-CPSS), however, is yet to be fully studied and investigated.

In this article, we study CPSS from a data-centric perspective, characterize its core features, and present a layered framework (architecture). In particular, we focus on multispace collaborative sensing and cross-space data fusion, and identify areas that require further research and development.

Data-Centric CPSS

Data-driven development will likely be a promising software paradigm in the coming decades. This will also lead to a revolution in the design and development of cyber-physical-social applications and services. With D-CPSS, we can leverage the cross-space, multimodal data from heterogeneous data sources to better characterize the target (for instance, an event or object). The combined effects of tri-space data will also nurture numerous, novel applications or services in urban environments. From a data-centric viewpoint, each CPSS follows a generic life cycle, consisting of data collection, processing, and usage. A four-layered architecture of a D-CPSS that follows this life cycle is shown in Figure 1.

The inputs of the architecture are cyber-physical-social data sources, whereas outputs are various applications on data usage. Looking at the four vertical layers, the leftmost two are about data collection, whereas the rightmost two are about data processing. As a human-in-the-loop sensing and computing system, a D-CPSS should maintain data privacy during its entire life cycle. We examine the details of each layer in the upcoming sections.

Collaborative and Augmented Sensing

D-CPSS leverages a combination of online and offline sensing resources. Different sources have distinct sensing or perception abilities; thus, they must be orchestrated to achieve collaborative sensing. In collaborative sensing, the input is the sensing task and tri-space sensing sources, and the output is highly relevant data about the sensing task. The resource management and collaborative sensing layers in Figure 1 depict the collaborative and augmented sensing process.

Resource management layer. The *resource management layer* manages various sensing resources from cyber, physical, and social spaces. When a sensing task arrives, this layer can quickly locate the sources relevant to its needs. This layer has the following functionalities:

 Published by the IEEE Computer Society

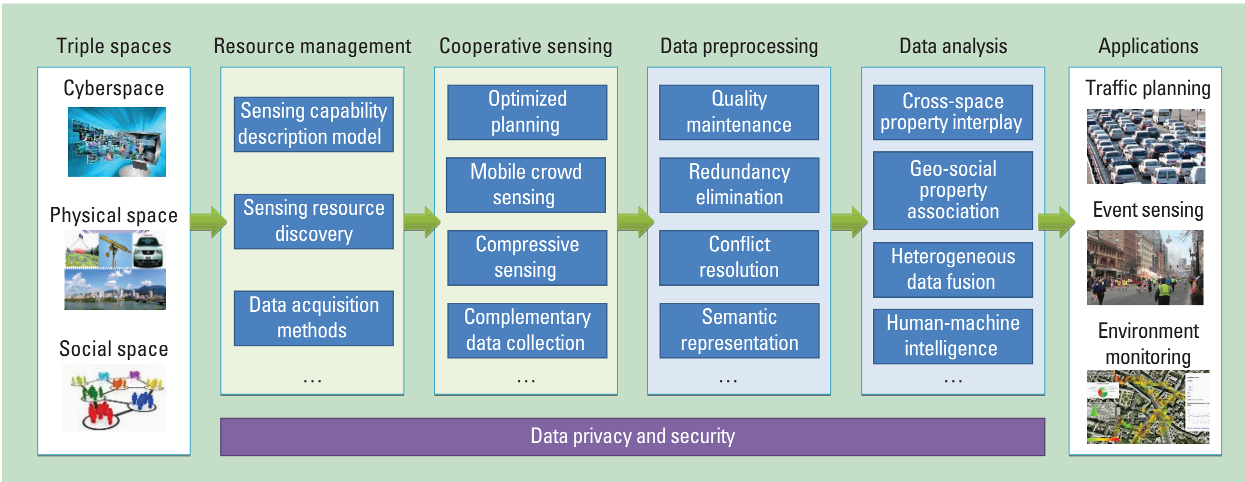

Figure 1. A layered architecture for data-driven cyber-physical-social systems (D-CPSS). The resource management and cooperative sensing layers deal with data collection, whereas the data preprocessing and data analysis layers deal with data processing.

- *Sensing capability description model.* A unified model should be defined to describe the capability of each sensing source. Numerous sensing capabilities should be depicted in the model, including its spatial-temporal coverage, sampling frequency, accuracy, range of error, and data type.
- *Sensing resource discovery.* This component discovers the relevant sensing resources according to the spatial-temporal needs of a task and the capability description of each sensing source.
- *Data acquisition methods.* The resource management layer contains a repository that involves various methods for collecting data from different forms of communities (using available Web APIs, crawling, or pervasive sensing techniques).

Cooperative sensing layer. The *cooperative sensing layer* manages the scheduling and cooperation of the selected sensing sources according to the dynamics of the sensing task. It has the following functionalities:

- *Optimized planning.* Due to the dynamics of a sensing target, active sensing resources often need to be rescheduled. For example, when the sensing target moves or a close shot of the target is obtained, new available sensing sources need to be discovered and activated.
- *Mobile crowd sensing.* The proliferation of smartphones leads to mobile crowd sensing, which leverages heterogeneous user participation in data collection, including offline sensing and online data generation.[3]
- *Compressive sensing.* The data collected by sensors are most often spatially and temporally correlated. Much data can be eliminated or compressed to facilitate data storage and processing.
- *Complementary data collection.* Different sources capture different aspects of a sensing target, and thus complementary data should be collected from each source to generate a complete picture.

Cross-Space Data Fusion

A D-CPSS aggregates heterogeneous and complementary data from different sources to better characterize the sensing target. The data processing and data analysis layers of the system architecture are about backend data processing and mining.

Data preprocessing layer. The *data preprocessing layer* is responsible for data cleaning and quality maintenance, and has the following functionalities:

- *Quality maintenance.* Numerous factors can damage data quality. For example, user-generated data in social media might be fake, humans can provide biased information, or sensor readings can be faulty. We thus need to find appropriate ways to measure data quality.
- *Redundancy elimination.* The data contributed by distributed sources usually contain redundancy in semantics or content. To facilitate data collection and processing, we should identify data redundancy with regard to task requirements and eliminate redundant data.
- *Conflict resolution.* Sensor readings or learned information can

be in conflict. For example, due to environmental differences, a group of co-located smartphones sensing the same event can obtain different recognition results. A more complex issue is the inconsistency of semantics derived from cross-space, multimodal sensory data (such as audio clips, images, videos, and texts). New methods for conflict resolution need to be studied.

- *Semantic representation.* Data from different spaces are heterogeneous while referring to the same individual. Heterogeneous data presentations can provide valuable depth, but machine processing generally expects homogeneous data. We should thus study semantic models to bridge the gap of multispace data sources.

Data analysis layer. The *data analysis layer* fuses the data from different spaces and studies the correlation and interplay among them. This layer has the following functionalities:

- *Cross-space property interplay.* Our society is founded on the interplay of human relationships and interactions. Under D-CPSS, investigations should explore the interplay between online interactions and real-world phenomena and events, such as interaction patterns and emergency events, as well as social interaction and economic development.
- *Geo-social property association.* The analysis of cross-space data indicates community properties from different dimensions: temporal, geographical, and social. These properties are highly associated under the bridge role of "human." For example, one study investigates the correlation between human social ties and geographic coincidences.[4]
- *Heterogeneous data fusion.* Data from online social media, physical sensors, and crowdsensed pictures might represent the same object (an event or activity, for example), but with distinct representation methods and from different aspects. However, information from these sources is complementary and lets the system attain augmented object understanding. Existing learning methods are homogeneous, but there are startups on heterogeneous data fusion[5] and cross-media computing.[6]
- *Human-machine intelligence.* The involvement of human intelligence in D-CPSS brings benefits to data processing. As studied in our previous work,[7] human and machine intelligence are complementary. The coexistence of human power and machine capacities, however, needs to be orchestrated in an appropriate manner to enhance them both. We have identified three fusion modes in mobile crowd sensing.[7]

Although research on D-CPSS is still in its early stages, there are startup applications that demonstrate its features and benefits.

D-CPSS Application Areas

D-CPSS will change the way that we do business, management, and research. Although research on D-CPSS is still in its early stages, there are startup applications that demonstrate its features and benefits. These applications range from social networking and intelligent transportation to public sensing and environment monitoring.

Social Contact Manager leverages a combination of online data crawling and offline smartphone sensing techniques to auto-collect rich information about human social contacts.[8] Tripplanner demonstrates the power of using a combination of participatory sensing and social media data for personalized, traffic-aware travel route recommendation.[9] FlierMeet allows people to repost the fliers posted on urban surfaces and circulate them in cyberspace.[10] Another proposed model infers fine-grained air quality information throughout a city, leveraging heterogeneous crowd-contributed data sources, such as traffic flow and points of interest in location-based social networks.[11]

Potential Research Issues

As a new research area, D-CPSS faces numerous challenges or issues before it can become reality. We've identified several areas that deserve further study:

- *Methods on collaborative sensing.* The study of cross-space collaboration and augmented sensing is still in its initial stage. Methods on how to efficiently discover relevant data sources and effectively leverage their complementary capabilities in the dynamic sensing process are still under investigation.
- *Data linkage and semantic presentation.* The linkage of data over different spaces should be identified. Semantic presentation models should also be studied to aggregate learned features from different spaces to the same sensing target.
- *Cross-space data mining.* With the increase in large-scale, interlinked data collected from

heterogeneous data sources, advanced techniques on complex network modeling, geosocial property association, and heterogeneous data fusion will become more important.
- *Privacy protection.* There is considerable public fear regarding the inappropriate use of personal data, particularly through the linking of data from tri-space data sources. Therefore, new privacy-preserving methods should be studied to deal with the information leakage issues that result during cross-space learning and prediction.
- *Novel application areas.* To propel the development of D-CPSS, we should keep exploring its wide usage and unique benefits in different application areas.

D-CPSS leverages the aggregated power of cyber, physical, and social spaces to improve the efficiencies of organizations and the quality of people's lives. The potential benefits of D-CPSS are real and significant, and some initial successes have already been achieved. However, several technical challenges must be addressed to fully realize its potential. This will bring many new opportunities for researchers, system developers, and engineers.

Acknowledgments

This work was partially supported by the National Basic Research Program of China 973 (no. 2015CB352400), and the National Natural Science Foundation of China (nos. 61332005, 61373119, and 61222209).

References

1. F.Y. Wang, "The Emergence of Intelligent Enterprises: From CPS to CPSS," *IEEE Intelligent Systems*, vol. 25, no. 4, 2010, pp. 85–88.
2. G. Schirner et al., "The Future of Human-in-the-Loop Cyber-Physical Systems," *Computer*, vol. 46, no. 1, 2013, pp. 36–45.
3. B. Guo et al., "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm," *ACM Computing Surveys*, vol. 48, no. 1, 2015, pp. 1–31.
4. N. Eagle et al., "Inferring Social Network Structure using Mobile Phone Data," *Proc. Nat'l Academy of Sciences*, vol. 106, no. 36, 2007, pp. 15274–15278.
5. J. Choo et al., "Heterogeneous Data Fusion via Space Alignment Using Nonmetric Multidimensional Scaling," *Proc. SIAM Int'l Conf. Data Mining* (SDM), 2012, pp. 177–188.
6. Z. Shi, "Cross-Media Knowledge Discovery," *Proc. Int'l Conf. Knowledge Discovery and Data Mining* (KDD), 2012, pp. 816–816.
7. B. Guo et al., "Building Human-Machine Intelligence in Mobile Crowd Sensing," *IT Professional*, vol. 17, no. 3, 2015, pp. 46–52.
8. B. Guo et al., "Enhancing Memory Recall via an Intelligent Social Contact Management System," *IEEE Trans. Human-Machine Systems*, vol. 44, no. 1, 2014, pp. 78–91.
9. C. Chen et al., "TripPlanner: Personalized Trip Planning Leveraging Heterogeneous Crowdsourced Digital Footprints," *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 3, 2014, pp. 1259–1273.
10. B. Guo et al., "FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing," *IEEE Trans. Mobile Computing*, vol. 14, no. 10, 2015, pp. 2020–2033.
11. Y. Zheng, F. Liu, and H.P. Hsieh, "U-Air: When Urban Air Quality Inference Meets Big Data," *Proc. Int'l Conf. Knowledge Discovery and Data Mining* (KDD), 2013, pp. 1436–1444.

Bin Guo *is a professor of computer science at Northwestern Polytechnical University, China. His research interests include pervasive computing, mobile phone sensing, and mobile social networks. Guo received a PhD in computer science from Keio University. He is a senior member of IEEE. Contact him at guob@nwpu.edu.cn.*

Zhiwen Yu *is a professor and director of the Department of Discipline Construction at Northwestern Polytechnical University, China. His research interests include pervasive computing and human-computer interaction. Yu has worked as an Alexander Von Humboldt Fellow at Mannheim University, Germany, and a research fellow at Kyoto University, Japan. He is a senior member of IEEE. Contact him at zhiwenyu@nwpu.edu.cn.*

Xingshe Zhou *is a professor of computer science at Northwestern Polytechnical University, China. His research interests include embedded and pervasive computing. Zhou is the director of the Shaanxi Key Laboratory of Embedded System Technology. He is a member of IEEE. Contact him at zhouxs@nwpu.edu.cn.*