SCIENCE CHINA
Information Sciences

• INSIGHT •
Special Focus on Advanced Technology of Software
May 2016, Vol. 59 050103:1–050103:3
doi: 10.1007/s11432-016-5545-5

Intelligent CPSS and its application to health care computing

Dayou LIU[1,2], Bo YANG[1,2*], Shang GAO[1,2*], Yungang ZHU[1,2] & Yong LAI[1,2]

[1]*College of Computer Science and Technology, Jilin University, Changchun* 130012*, China;*
[2]*Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun* 130012*, China*

Received November 5, 2015; accepted January 18, 2016; published online April 8, 2016

Citation Liu D Y, Yang B, Gao S, et al. Intelligent CPSS and its application to health care computing. Sci China Inf Sci, 2016, 59(5): 050103, doi: 10.1007/s11432-016-5545-5

The rapid development of cyber-physical systems (CPS) had a tremendous impact on human behavior and interactions. With human involvement, CPS has naturally evolved into cyber-physical social systems (CPSS) [1]. The top five technology breakthroughs of 2013 were closely related to CPSS and peripherals according to McKinsey's report [2].

CPSS takes full account of crowd intelligence, and requires effective integration of data, information and knowledge from diverse dimensions of computing, networking, physical/social systems, etc., all with the purpose to integrate and analyze complex human behaviors. Thus, data/information/knowledge processing in CPSS based systems is challenging: not only do we need to solve massive, high-dimensional and noisy data problems, but also need to deal with further difficulties originating from data uncertainty, inconsistency, heterogeneity and spatio-temporal effects. With these challenges, the effective integration of distributed data mining, network data mining, data fusion, spatio-temporal reasoning and many other artificial intelligence methods is expected to play important roles. In this paper, we exemplify such integration in CPSS by using intelligent health-care computing in patient-oriented systems.

Current data analysis methods mostly require the entire input data, not suitable for handling large-scale, distributed, decentralizd, and temporal data sets, which often characterize CPSS. Therefore, the integration of big and complex data sets is crucial. We believe that distributed data mining is one of the key areas, and suggest two basic strategies in this regard: divide-and-conquer and self-organization. For example, in applications with a large number of variables, which can be pair-wise correlated or independent to some degree or none, we can divide variables into different sets based on their associative relationships, then apply traditional techniques independently and merge the distributed results. Alternatively, one can take advantage of multi-agent systems, and build a self-organized, adaptive, and distributed data processing system, where agents in different CPSS components can gather and process local information and communicate through networks as well as collaborative protocols. By using multi-agent systems, locally collected and possibly hidden information can be effectively utilized in discovering global knowledge patterns.

To solve complex problems in the real world, humans play a leading role. It has been controver-

* Corresponding author (email: ybo@jlu.edu.cn, shanggao@jlu.edu.cn)
The authors declare that they have no conflict of interest.

sial whether to describe intelligent behaviors by structure based methods or statistics based methods, the former of which emphasizes knowledge based methods and the latter focuses on machine learning techniques. Here another integration issue arises, which is to integrate structure based and learning based knowledge with human involvement. We argue that only with effective integration, one could reach the ambitious goal of using CPSS to tackle complex problems. As an example, the interdisciplinary research on graphical models and probabilistic methods is a good candidate for such integration. Not only does it take into account the structured knowledge (i.e., easy to understand), statistical meaning and causal relationships, but it also has the potential to handle massive amounts of data [3]. Probabilistic graphical models (PGMs) can be decomposed into "smaller and simpler" problems using conditional independence, hence they provide a fundamental toolbox and theoretical basis for distributed data mining using divide and conquer strategy. Currently, we are facing two major challenges when using probabilistic graphical models in CPSS: First, most learning and reasoning problems are NP-hard, requiring exponential growth of computing capabilities with a large number of variables; Second, there is a gap between learning and reasoning in PGMs, which may result in a huge waste of time spent in computation. For the first challenge, we need to tackle the curse of dimensionality and to adopt and integrate diverse reasoning algorithms in both online and offline scenarios. For the second challenge, we suggest combined and integrative framework in learning and reasoning phases for building a practical model [4].

CPSS comprises a wide variety of networks, including communication networks, control networks, computation networks and social networks, etc. In order to gain deep understanding of the system behaviors, one needs to analyze these networks and the data generated by them [5]. This highlights the importance of network mining and analysis in CPSS research. Taking four key elements of CPSS (society, control, communication, and computation) into account, we highlight the following research areas and questions: (1) Social network analysis, i.e., how do individuals compete and collaborate, and how does the information flow through networks? (2) Control network analysis, i.e., what types of network are controllable and how should they be controlled? (3) Communication network analysis, i.e., how should we build effective and robust communication networks? (4) Computation network analysis, i.e., how does the microscopic behavior affect macroscopic behavior? (5) Multiplex network analysis, i.e., how should we collaboratively analyze and integrate different networks in CPSS?

Overall, network mining and analysis is multidisciplinary in nature. Therefore, to answer the above questions, cross-area efforts must be pursued in both inter (physics, sociology, and mathematics, etc.) and intra (graph theory, probabilistic methods, and distributed data processing, etc.) computer science fields.

Intelligent health care computing (IHCC) is a typical and significant application of CPSS. From technical aspects, building IHCC becomes technically mature after 20 years of research and development in machine learning and data mining. Furthermore, information technologies and crowd knowledge bases become influential with the wide deployment of cloud platforms. Among these, our aim is clear: we need to bridge the gap between intelligent human behavior and machine intelligence.

Providing better treatment for patients is one of the major goals of IHCC. IHCC involves three major steps in providing intelligent diagnosis: first, patients consult experts; second, experts give advice with comprehensive knowledge bases and intelligent diagnostic systems; and finally, knowledge and data are consolidated for conclusive diagnoses and treatments. Data fusion is demanded in the last two steps. It is practically effective when using multiple sources of data and different knowledge bases in designing intelligent health care systems. Needless to mention, knowledge consolidation is a critical phase to safe-guard effective data fusion process in the last two steps [6].

IHCC consists of three major components. Firstly, the system requires shared and high-performance computing environment, including Internet technologies, medical equipments, sensor networks and cloud computing platforms. Secondly, IHCC demands effective data integration of domain knowledge bases and representation, as well as reasoning (under uncertainty) methods. Thirdly, the system includes patient profiles, comprehensive diagnoses and treatment history, targeted expert knowledge bases, aiming to providing intelligent systems, such as data retrieval system and intelligent diagnostic solutions.

CPSS-based health care computing is still in its infancy with substantial challenges, and we hope the system can improve the health care systems in China: patients can have a systemic way of seeking expert advice. By directing patients to the right Medical Doctors and providing them with the right treatments, medical resources can be utilized efficiently and consistently. However, as in other CPSS applications, huge efforts are needed in building a large number of system components, such as to intelligently consolidate distributed knowledge bases, to achieve decision making and verification, and to comply different ethical standards and protocols.

Government, hospitals, doctors and individuals are all potential IHCC users. IHCC is coordinated by government on research and development phases, in order to achieve standardization and data/knowledge sharing protocols. Hospitals can build customized patient databases via IHCC, and M.Ds can use IHCC to aid their diagnosis treatments. Patients can constantly consult the IHCC through public portals in different phases regardless of locations for getting better diagnoses and treatment advice.

To conclude, CPSS is an important extension of CPS by the deep integration of computing, networks, and social/physical systems. Using the health care as an example, it is clear that CPSS is set to promote the scientific revolution and create new computing capabilities, which provides a tremendous opportunity for the next generation artificial intelligence research.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61133011, 61373053, 61572226, 61502198, 61402195), Jilin Province Natural Science Foundation (Grants Nos. 20150101052JC, 20150520066JH), China Postdoctoral Science Foundation (Grants Nos. 2013M541303, 2014M561292), and Opening Fund of State Key Laboratory of Applied Optics.

References

1 Wang F Y. The emergence of intelligent enterprises: from CPS to CPSS. IEEE Intell Syst, 2010, 25: 85–88
2 Manyika J, Chui M, Bughin J, et al. Disruptive Technologies: Advances That Will Transform Life, Business, and the Global Economy. Technical Report. McKinsey Global Institute. 2013
3 Bielza C, Moral S, Salmern A. Recent advances in probabilistic graphical models. Int J Intell Syst, 2015, 30: 207–208
4 Dietterich T G, Domingos P, Getoor L, et al. Structured machine learning: the next ten years. Mach Learn, 2008, 73: 3–23
5 Hussein D, Park S, Han S, et al. Dynamic social structure of things: a contextual approach in CPSS. IEEE Int Comput, 2015, 19: 12–20
6 Bleiholder J, Naumann F. Data fusion. ACM Comput Surv, 2008, 41: 1–41