Editorial

Rumen D. Andreev and Yuri P. Pavlov

Department of Communication Systems and Services,
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences,
str. Acad. G. Bonchev Bl. 2, Sofia 1113, Bulgaria
Email: rumen@isdip.bas.bg
Email: yupavlov15@isdip.bas.bg

Biographical notes: Rumen D. Andreev is currently an Associate Professor at the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia. He obtained his PhD in Computer Aided Design from Technical University, Sofia in 1987. He is Head of the Department "Communication Systems and Services" and Director of the Technology Transfer Office. His research interests include computer graphics, human–computer interactions, software engineering, computer aided design, personalised computer-enhanced learning, technology transfer, adaptive and complex systems. He has published in *Computers and Graphics*, *Computer Graphics Forum*, *Interacting with Computers*, *Communication and Cognition* and others and has presented his research at numerous national and international conferences. He is an author of two monographs: "Graphics Systems: Architecture and Realization", North Holland, Elsevier and "Decision Control, Management, and Support in Adaptive and Complex Systems: Quantitative Models", IGI Global. He has professional awards from Marquis' Who is Who in the World, International Biographical Centre, Cambridge, and American Biographical Institute.

Yuri P. Pavlov received his PhD in Cybernetics from the CLCS, Bulgarian Academy of Sciences in 1990 and he is currently an Associate Professor at the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia. His primary research interests include the use of utility theory, theory of measurement and stochastic approximation techniques in decision making and decision support systems. Different approaches for stabilisation and control, reduction and equivalent transformation of non-linear systems and optimisation of fed-batch cultivation processes are also of current interest. He has published in *Proceedings of Bulgarian Academy of Sciences*, *Encyclopedia of Business Analytics and Optimization-IGI Global*, *International Online Journal Bioautomation*, *European Journal of OR*, *Proceedings in Manufacturing Systems* – Romania, Global Journals Inc. (USA) and *Indian Journal of Applied Research* (India). He is an author of the monograph: Pavlov, Yuri P. and Rumen D. Andreev – "Decision Control, Management, and Support in Adaptive and Complex Systems: Quantitative Models", Hershey, PA, IGI Global, 2013. He is also Member of International Institute of Informatics and Systemics (IIIS) and Informing Science Institute.

The existing definitions of 'Big Data' use atleast one of the following critical factors to determine datasets as Big Data: *Size* – the volume of the datasets; *Complexity* – the structure, behaviour and permutations of the datasets; *Technologies* – the tools and techniques which are used to process a sizable or complex dataset. An extrapolation of these factors would therefore postulate the following: *Big data is a term describing the storage and analysis of large and/or complex datasets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning* (Ward and Barker, 2013).

The presented definition provides the following terms in relation to Big Data: data analysis, data storage, big data techniques and complex data. From them a number of trends for research in this area are evident. They focus on different aspects of Big Data. Firstly, that big data is intrinsically related to data analytics and the discovery of meaning from data. Secondly, it is clear that there are a number of related techniques. Finally, Big Data are used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information. This definition states that Big Data require the application of significant compute power. That is why we decide to concentrate this special issue on two aspects of Big Data – data analysis and computing infrastructure.

Domains that deal with Big Data include many of the sciences along with geographic information systems, geophysical data systems, medical systems, financial analysis, software development, social media or online systems. The amount of data makes the analysis task difficult.

The first five papers in this special issue are selected among the presented papers during an International Conference entitled "Big Data, Knowledge and Control Systems Engineering-BdKCSE'2014" organised by Institute of Information and Communication Technologies – Bulgarian Academy of Sciences and John Atanasoff Society of Automatics and Informatics, Bulgaria at 5th November, 2014 in Sofia, Bulgaria.

The sixth paper is an invited paper written by Velev and Zlatewa. Dimitar Velev is a Professor from the University of National and World Economy, Sofia, Bulgaria and Plamena Zlatewa is an Associate Professor from the Institute of System Engineering and Robotics – Bulgarian Academy of Sciences.

The first paper, "Scalable system with accelerators for financial option prices estimation" is written by Dimitrov and Atanassov. It describes a production-ready ESB system for estimation of option prices using stochastic volatility models. The system is capable of using different data sources and it is extendable from both infrastructure and software point of view with hot deployment. The main advantage of the system is that by incorporating GPGPU-computing and Intel Xeon Phi nodes it allows for the use of more accurate models that are otherwise unfeasible. This system can be useful for distributed processing of a large volume of option pricing tasks.

Gerunov's paper "Employment modelling through classification and regression trees" leverages a big dataset from the field of social sciences – the combined World Values Survey 1981–2014 data – to investigate what determines an individual's employment status. It proposes an approach to model this by first reducing data dimensionality at a small informational loss and then fitting a number of alternative machine-learning algorithms. A decision tree and a random forest model are studied in more detail. The main contribution of this paper is to outline a new approach for doing big data-driven

research in labour economics and apply it to a dataset that was not previously investigated in its entirety, thus achieving a more sophisticated process understanding.

The authors of the third paper "Estimation of flood risk zones of Maritza River and its feeders on the territory of Svilengrad municipality as part of smart water project web-GIS tool" are Dobrinkova and Slavov. They describe structured hydrologic estimation of the high flows, formed after intensive precipitations in Maritza River and its feeders on the territory of Svilengrad and neighbouring municipalities. The computed results will be implemented in the web-GIS tool, which structure will be also presented as modules for civil protection response capacity support for decision making by the responsible authorities. The high wave evaluation and its implementation in the web-GIS tool are part of the smart water project supported under DG 'ECHO' call for prevention and preparedness, will give illustration of the first attempts of application of INSPIRE directive on the Bulgarin–Turkish–Greek border zone.

The fourth paper "Residual bounds of the nonlinear matrix equation $X + A^{*}\mathcal{F}(X)A = Q$" is written by Popchev's and Angelova's and present a way of determine a residual bound using the method of Lyapunov majorants and the techniques of the fixed point principle. The first method is a widely used technique in control theory and the second is a technique of the functional analysis. Such approach has a more general significance and it is used for a class of non-linear matrix equations in the paper. Such types of equations arise when solving to discrete-time Riccati equations, arising in filtering and control theory. Riccati equations are a traditional topic in linear system theory. The linear system theory could be referred to the Big Data factor technologies – tools and techniques. The digital devices such as smartphones and sensors have led to an unprecedented rate of data creation and are driving a growing need for real-time analytics and evidence-based planning and even control. This data provide noised information about customers, such as geospatial location, demographics and past buying patterns, which can be analysed in real time to create real customer value. The structurisation and the modelling of the noised high-velocity and high-variety information could permit cost-effective and innovative forms of information processing and control that could be used for decision making.

The fifth paper "Cooperation of substantial agents in MAS" is written by Čapkovič et al. It concerns problems related to computing infrastructure consisting of agents that could be considered as service-oriented system. Each agent contains dataset that could be processed by its intelligent functions. The paper presents the usage of place/transition Petri nets (P/T PN) in order to model the behaviour of agents as well as the agent communication which is necessary for their cooperation and negotiation. Two kinds of the agent communication are compared. The first assumes that all agents are equal, i.e. no agent has a higher priority than others. The second kind of the communication has the hierarchical structure. Here, a supervisor on the upper level has higher priority than other agents. It coordinates activities of the agents being on the lower level. Here, the individual agents do not cooperate directly, but through a supervisor.

The sixth paper "An analysis of the relation between natural disasters and Big Data" is written by Velev and Zlateva. It concerns problems related to the needs of combinations of several major ICT developments to a more effective management of natural disasters. Nowadays, the combinations of several major ICT developments form a new trend to management of natural disasters – social networking, mobile computing, Internet of Things and *Cloud Computing*. The combination of these ICT

technologies leads to the situation that rescue organisations are trying out new ways to find value insight from both structured and unstructured data from internal and external sources. Such more complex point of view and more complex approaches are expected to complement but not replace long-standing information management programs and investments in data warehouses, business intelligence suites and relational database experience. In this paper, it is investigated and analysed the relation between natural disasters and Big Data and it is underlined the need of technology as Big Data and new software tools for handling this.

Reference

Ward, J.St. and Barker, A. (2013) *Undefined By Data: A Survey of Big Data Definitions*, arXiv:1309.5821v1 [cs.DB], 20 September, 2013.