Interview with Daniel Dolk and Christer Carlsson on "Decision Analytics"

Daniel Dolk is Professor Emeritus in the Department of Information Sciences at Naval Postgraduate School in Monterey, CA (USA). He helped pioneer the discipline of model management which lies at the intersection of computer science, operations research/management science, and decision support. His current research interests are predictive analytics and service science, decision support for mobile and social computing, "big data" architectures for real-time bidding and programmatic marketing, and evolutionary computing. He has been the Decision Technologies Track Chair for HICSS from 1998–2012 and is Vice Chairman of the IFIP Working Group 7.6 on Optimization-Based Computer-Aided Modeling and Design.
Christer Carlsson is Director of the Institute of Advanced Management Systems Research, and a professor of management science at Abo Akademi University in Abo, Finland since 1985. He is the President of the International Fuzzy Systems Association [IFSA]. He is a Fellow of the IFSA, a Senior Member of the IEEE and a member of the SG of the European Center for Soft Computing. He received the VII Kaufmann Prize and Gold medal from SIGEF. He has worked in industry-sponsored projects with major Finnish multinational corporations for more than 25 years and developed solutions for strategic planning with intelligent tools, fuzzy real options planning instruments, logistics optimization systems, knowledge mobilization, foresight with soft computing agent tools and predictive maintenance with fuzzy ontology and possibilistic Bayes models.

DOI 10.1007/s12599-014-0328-2

Daniel R. Dolk
Department of Information Science
Graduate School of Operational and Information Sciences
Monterey, CA 93943
USA
drdolk@nps.edu

Christer Carlsson
IAMSR at Åbo Akademi University
Joukahainengatan 3-5 A
20520 Åbo
Finland
christer.carlsson@abo.fi

Interview by

Prof. Dr. Stefan Voß (✉)
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
20146 Hamburg
Germany
stefan.voss@uni-hamburg.de

Published online: 2014-04-25
This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Voß S (2014) Interview mit Daniel Dolk und Christer Carlsson zum Thema „Entscheidungsanalyse". WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-014-0419-1.

BISE: Sunny Hawaii! The HICSS conference (Hawaii International Conference on System Sciences) is always among the nicest venues to talk about the future of information systems and beyond. With Dan Dolk as the previous chair and Christer Carlsson as one of the current chairs of the decision sciences and decision analytics track (with "Decision Analytics, Mobile Services and Service Science" the name is now slightly different but that should not concern us here) we have most prominent representatives of the discipline to chat with – it is, as always, a pleasure.

Dan, Christer, let's talk about the future of decision analytics and where we should go next. As academics we usually have a challenging road ahead, but what are the real challenges? And what could we learn from the past? In a nutshell, what is the analytics challenge?

Carlsson: The Analytics Challenge, a good topic! We should address real-world problems in industry, society and business that are worth the efforts to tackle and solve, that is, the outcome and the consequences are important enough to motivate the use of resources to find solutions. Typically these problems have some joint characteristics: they are large and complex, that is, they are formed by numerous interacting factors with numerous relations; they are dynamic, that is, both factors and relations change over time; they are buried in big data, that is, the data sets that could offer a handle for tackling the problems are so large that algorithms and computational technology are ineffectual for working on them.

There are some lessons learned from operations research a couple of generations ago (and thus mostly forgotten by the new generation of analysts that rely on massive computer support):

- Use analytics to find the core logic and effective variables that create and run problems we have to tackle; working out the core logic means that we can clean away irrelevant data and significantly reduce the big data problem.
- Develop analytics methods that can match the complexity of the problem and process the data in reasonable time; increasing complexity can be matched with more effective and faster algorithms, which appear to be metaheuristics based and geared to find solutions that are not necessarily optimal but "near-optimal."
- "There is a trade-off between precision and relevance: at some point you cannot increase relevance without losing precision or increase precision without losing relevance" [a nice quote from Lotfi Zadeh]. For real-life problem solving it is often beneficial to work out "sufficiently good" solutions, especially if this can be done in "almost real-time" – we have good examples of highly automated industrial production processes where the cost effects of production stops over several hours quickly over-shadow the cost effects of not having an optimal solution.
- Develop technology to handle the "big data" challenge and the dynamics of the digital society; that is, we have created the "big data" problem by generating huge amounts of data – simply because the computer scientists came up with the technology to do it, and we believed in the slogan the more data the better – and we will have to develop the technology to cope with it; this has happened a number of times in the decades of operational research and is basically not anything new, but we should stay focused on technology development that is necessary and relevant for the types of problems we try to solve and not get carried away by "nice to have" technology contributions.

BISE: I remember an older lesson to be learned for the operations research specialists. If something in the real world looks like a traveling salesman problem, do not make it a traveling salesman problem. Instead carry your algorithmic and problem solving expertise towards the problem rather than the other way around. And not because we can model a problem, (we should not promise that) we can solve it to optimality. Good points well taken.

While decision analytics may seem more oriented to the operations research or OR community, what do the information systems or IS people have to do? Is INFORMS with its slogan "Science of Better" trying to find answers regarding these key points or questions or are they assuming that they are the more important discipline to go ahead?

Dolk: My perception is that decision analytics is the answer to the OR community's long standing desire for a full recognition from the business community that the kinds of modeling they do are directly important and relevant. Rather than a largely academic exercise in mathematical ingenuity and problem solving, the portfolio of OR technology is now in the critical path of business processes and practices across a much wider spectrum of applications than before. I'm not sure how INFORMS sees themselves and the OR discipline in the context of BDDA, *Big Data and Decision Analytics*, but they are certainly a key player without any doubt. My hope is that BDDA will strengthen the relationship between the information sciences and operations research, a union which historically has not been leveraged effectively.

Carlsson: IS as a discipline is sharing the background in Positivism with OR, that is, IS researchers build models – artifacts – of real-world events and problems, validate and verify the correctness of these models, use algorithms to find solutions and then implement and verify the usability of these solutions to deal with real-world problems. The difference between IS and OR researchers is that the modeling is more supported with information technology in IS research than in OR research. The Analytics movement – now strongly endorsed by INFORMS, which was not always the case – is relying strongly on contributions from information technology. Hence, IS people have more of the relevant know how than OR people.

BISE: Uppss, so we should focus on being a sparring partner between the worlds. Then, the next question should be as follows: What will be the future role of a multidisciplinary person on the interface between IS and OR?

Carlsson: In the next 10 years the multidisciplinary people are going to be in high demand in the industrial and business job market. Knowledge of IS and OR – and the skill and experience to use it for actual planning, problem solving and decision making – is sought after by the major corporations and this demand will continue to grow over the next 10 to 15 years. The "big data" wave has now created a hunt for statisticians – it seems to have been some kind of surprise that you need statistical methods for handling masses of data.

Dolk: This question follows closely on to the previous one. IS professionals now have a great deal to do in applying the science of design to the challenging requirements of these new environments. Take the case of real-time bidding for advertising exchanges where the placement of on-line ads is done on an individual-by-individual basis. Real-time bidding architectures must be able to access an individual's demographic and purchasing profile (often multiple files with millions of rows), match it to one or more ads being considered for delivery, and determine a bid price when a match looks promising, all in less than 100 ms. Further, this process may be repeated millions of times a second depending upon the reach of the advertising campaign (typically millions of recipients) and the number of advertising campaigns being conducted simultaneously (typically a few hundred). This extreme computing environment demands a new kind of IS/Computer Science professional, one who is not only fluent in advanced database processing and high performance parallel processing, but also able to traffic comfortably in optimization, Markov chain, machine learning, information retrieval, naive Bayesian, and simulation models, to name just a few.

More generally, a new breed of IS professional is emerging, now branded the "data scientist," who can cross-walk between business processes, statistical and OR modeling, and high end computer architecture. A McKinsey 2011 quarterly report stated that "large-scale data gathering and analytics are quickly becoming a new frontier of competitive differentiation." A NY Times article this year reported that McKinsey Global Institute projected that the US needs 140,000 to 190,000 more workers with deep analytical expertise and 1.5 million more data-literate managers. This new category of data scientist is geared towards solving BDDA kinds of problems and will probably necessitate the creation of multidisciplinary curricula in universities to handle and advance the synthesis between IS and OR.

BISE: To which extent do we need to combine classical operations research methods and information systems and is "soft computing" the only way (of course not when it comes to my mind, but let's be provocative)?

Carlsson: "Soft Computing is developing the theory, methodology and instruments to bridge analytics and intuition."

In my mind IS and OR are overlapping methodologies and we need not think in terms of "combining" the methods, we simply combine OR modeling and IS technology as needed to handle and deal with the problems we need to find solutions to. Soft Computing is a development of methods and technology to handle imprecision and imprecise data, that is, soft computing deals with problems where classical OR is failing.

Dolk: Soft computing, which emanated originally from Zadeh's fuzzy logic, is a discipline recognized much more widely in Europe than in America. Because of this cultural difference, I don't think one can say it is the only way to combine OR and IS. However, "soft computing" in its broadest sense, which I take to mean heuristic (that is, not crisp, "hard" algorithmic), non-deterministic approaches to solving problems, is vital to analytics. For example, as decision cycles become even more time-constrained (see the real-time bidding discussed before), adaptive modeling becomes more and more critical. "Hard" models may serve as initial starting points for these applications, but quickly give way to adaptive modeling feedback loops as real-time data streaming occurs. This emergent, adaptive nature of modeling is more receptive to "soft computing" paradigms and approaches as seen in the biological sciences compared with the "hard computing" models of the physical sciences.

BISE: Occasionally we find ongoing discussions on what big data is. Common to most definitions is that big data is a collection of data (sets) so large and complex that they are difficult or impossible to process with traditional database management tools or data processing applications. To which extent is Decision Analytics the driver to let the field of big data flourish? Or is it the other way around?

Carlsson: Decision Analytics is a means to let the field of "big data" flourish and also to open up the business opportunities that are advertised to be in the tens of billions of Euro on an annual base. DA will help us to quickly make sense of what is relevant data and how to employ it for optimal problem-solving, planning and decision making.

Decision analytics is defined as the scientific process of transforming data into insight for making better decisions. That is, it is the opposite of "fast and bad quality decision making." Decision analytics offers the only way to come to terms with "big data" and to get a return on investment of the resources (knowledgeable people, modeling tools, technology) that are needed to deal with the "big data" challenges.

Dolk: "Big data" and decision analytics comprise a dynamic feedback loop which is generative in nature. For example, the proliferation of sensors to finer and finer degrees of granularity creates gigantic data sets, requiring advanced analytics for interpretation and sense-making. Insights from this knowledge discovery process oftentimes result in substantial changes to business practices which, in turn, drive the need for further analyses and models and the (hopefully) virtuous cycle continues. So I don't see one or the other as being the prime driver but rather both involved in a co-evolutionary process.

BISE: What is the future of "big data" and decision analytics and what impact is it likely to have on business and society?

Dolk: Decision analytics has made a huge impact in a relatively short period of time. As the business world is becoming more aware of the potential value that analytics can contribute, there is a natural rush towards explaining how companies should leverage this technology for "transforming data into insight for making better decisions" and getting a full ROI on BDDA investment. This consultant-flavored enthusiasm inevitably accompanies the advent of new technologies and will run its course accordingly. However, there are very real challenges that confront business with respect to the design, development, and management of BDDA systems. Model management, for example, will become a big issue as companies find that severely time-compressed decision loops result in running thousands of models in just a few seconds or minutes. The efficient management of data ETL (extract, transform, load) processes will be a critical success factor for BDDA since roughly 70–80% of a project's overall effort is likely to be spent in this arena. (This will come as no surprise to OR modelers.)

At the societal level, I believe BDDA constitutes a transformative and generative force which truly is exciting. One can see whole new fields and enterprises arising from BDDA applications which render "business as usual" obsolescent. For example, the BDDA application developed by the Climate.com Corporation which combines massive amounts of weather, historic crop yields, and soil data allows farmers to manage their fields at a much more microscopic level than before, even to the point of determining whether a field on one side of the road is better than one on the other side. The results are higher crop yields and less likelihood of loss resulting from extreme climate phenomena. Climate.com then sells insurance to the farmers based upon the risk assessment of their respective fields. This brings agriculture directly into the BDDA arena (the NewYorker was reporting on that;

http://www.newyorker.com/reporting/2013/11/11/131111fa_fact_specter).
A major driving force behind BDDA is the proliferation of sensors, especially location-based sensors, at finer and finer levels of granularity. These sensors generate massive amounts of data at all scale levels (for instance, from satellite data to nano-sensors in human bodies) providing a nearly limitless set of application domains in which BDDA can thrive.

BISE: Yes, this is big on big data. Then let's start one more time at the OR scenery in mentioning the framework of matheuristics. *Matheuristics* are optimization algorithms made by the interoperation of metaheuristics and mathematical programming techniques. An essential characteristic is the exploitation of features derived from the mathematical model of the problems of interest. Metaheuristics and related templates were, in fact, usually proposed in the past, when Mixed Integer Programming (MIP) was seldom a viable option for solving hard real-world problems. However, research in mathematical programming, and in particular in discrete optimization, has led to a state of the art where MIP solvers or customized MIP codes can be used effectively even in a heuristic context, either as primary solvers or as sub-procedures. So, we had been somewhat successful in coining this term matheuristics. But how to do it when it comes to decision analytics?

What about *Meta-Analytics*? Let's assume we have this as a new buzzword; how would you define it? Or is it just a crazy idea?

Carlsson: I am not sure about this – we will probably over time form a Decision Analytics paradigm with good practice solutions (in the same way that we got families of OR methods) that may build a generic theory base that can identify good paths for effective decision support in the framework of complex problems and big data.

Dolk: Meta-analytics can be approached from two perspectives: specifically, metaheuristics from the OR point of view, and more generally in the overall BDDA landscape, meta-modeling. Metaheuristics are of course very familiar to OR researchers as search strategies for generating approximate, sub-optimal solutions to problems in a computationally reasonable time. As such there is an affinity to the "soft computing" philosophy. Metaheuristics-driven solvers certainly comprise part of the analytics landscape but it seems to me too narrow a context in which to contain the term "meta-analytics." Metaheuristics are primarily used to solve optimization problems, whereas the analytics field is much broader than that. "Meta-analytics" implies to me a higher level and more encompassing purview of strategies for applying analytics to particular problems. This brings into play the concept of meta-models as well as the grander issues of developing analytics methods that match the complexity of specific problems while still processing the huge volumes of data in reasonable time.

BISE: Great insights. Thanks very much, Dan and Christer. This is a lot of food for thought for going back to work on decision analytics. There seem to be many challenging opportunities ahead. Let's face them.