1-1-2013

Connecting big data with big decisions: Ideas for synthesizing analytics and decision analysis

Jeffrey Keisler
University of Massachusetts Boston, jeff.keisler@umb.edu

Connecting big data with big decisions: Ideas for synthesizing analytics and decision analysis.

Jeffrey M. Keisler,

Abstract: *An approach is developed to connect decision analysis models with outputs of analytic methods applied to various types of big data. Decision analysis models focus on issues of concern to a decision maker and incorporate use of a range of methods and axioms to develop insights about what the decision maker should do. In particular, decision analysis models typically use subjective judgments from the decision maker to describe beliefs about the likelihood of events and the desirability of outcomes. In order for human judgments to be improved by the availability of large amounts of data and processing power, it is necessary to define the right variables to interpolate between the data source and the decision model. Several applications are reviewed and suggest a more general approach.*

The issue:

Analytics, using combining massive computing power and massive data to identify patterns in the world, is coming of age. Still, the main applications are either directed toward general process improvement such as finding cost drivers, or toward rather small though very frequent automated decisions such as which popup ad to display.

Decision analysis developed in an earlier era of data scarcity, incorporating a modest number of subjective judgments about probabilities and preferences in order to find the best course of action in a given situation. The methods of applied decision analysis aim to structure these judgments and the questions to elicit them. Even at its genesis, however, decision analysis was also suited to incorporating statistical data, e.g., the sampling-based decision models in (1).

As decision analysis has evolved and expanded in the way it describes beliefs and preferences and possible actions, it has retained the ability to work with estimates of varied pedigrees. Thus, it has the capacity to serve as a needed front-end for analytics, connecting big data with big decisions. Carl Spetzler, Chairman and CEO of the consulting firm Strategic Decisions Group, has called decision analysis "analytics for the boardroom." Drawing on earlier efforts to link decision analysis with computer-derived values, we can articulate the process and means by which to do make this so. Making such connections has been perhaps a hidden theme in much of decision analysis. It has been an emergent theme of much of my own work, and this paper will focus on articulating that theme.

The basic idea:

We start with a decision model using only the usual material, e.g., an objectives hierarchy defined in the terms natural to the decision maker, or an influence diagram with chance nodes with events that are natural for the decision maker to think about. We then identify a loose universe of possibly relevant data sets and variables, and consider the possible analytics methods that might be applied, be they web mining, data mining, clickstream monitoring, web analytics, sensor tracking, etc. We then create a sampling of the type of results these methods might practically produce from the data. Using this sampling, the decision model is then expanded, adding nodes that will ultimately serve as links to the data universe.

The expansion takes the form of dummy nodes, e.g., evocative nodes in an influence diagram (variables the decision maker thinks about while making assessments, but does not explicitly quantify), or possible metrics in an objectives network for a multi-attribute utility function. These dummy nodes should be somewhat related to or reminiscent of the sample possible outputs of analytics methods. Using these nodes as templates, we then brainstorm for ways to get analytics outputs that are quite close to them. By having these templates as targets, it is usually possible to focus the analytics plan to produce relevant results through some grouping of data and series of steps. If not, we may still iterate on definition of the dummy nodes and repeat this step. Finally, we modify these dummy nodes to make them active nodes in the decision analytic model so that their successor nodes will be dependent on outputs of analytics methods – treating the output of the analytics (however much calculation and information processing underlies it) as an uncertainty or experimental result to be revealed. We may also leave some markers outside the usual DA model that serve as attractors for the pattern recognition of the analytics, i.e., adding in some additional evocative nodes representing values that we expect to be easily produced as outputs from data mining or other analytics methods.

Several efforts brought me to think about this aspect of decision modeling:

> Multi-attribute utility with GIS (2): This development takes map layer statistics, e.g., number of cells with values within some range, and uses this statistic as a measure for performance on one of the attributes. Using an iterative and interactive (perhaps we can combine these to call it *iteratactive*) process, a set of attributes is identified that is natural for both the assessment of stakeholder preferences (tradeoffs and single attribute utility curves) and for the generation of the statistics from map data.
>
> Web-agent populated decision analysis models (3): Decision tree models or influence diagrams involving variables of interest can be based on conditional probabilities, where the conditioning is on search analytic data. For example, the estimated probability of an economic recession in the next year might be based on the ratio of the number of hits for the words "growth" +"economy" to the ratio of the number of hits for "slowing" + "economy." Again, using an iteractive process, we can create all sorts of metrics based on what might generate interesting search results and what human judges might find to be useful conditioning variables for the uncertainties in their more intuitively workable decision model. This approach could certainly be expanded to encompass the much richer capabilities of Google Analytics currently available.

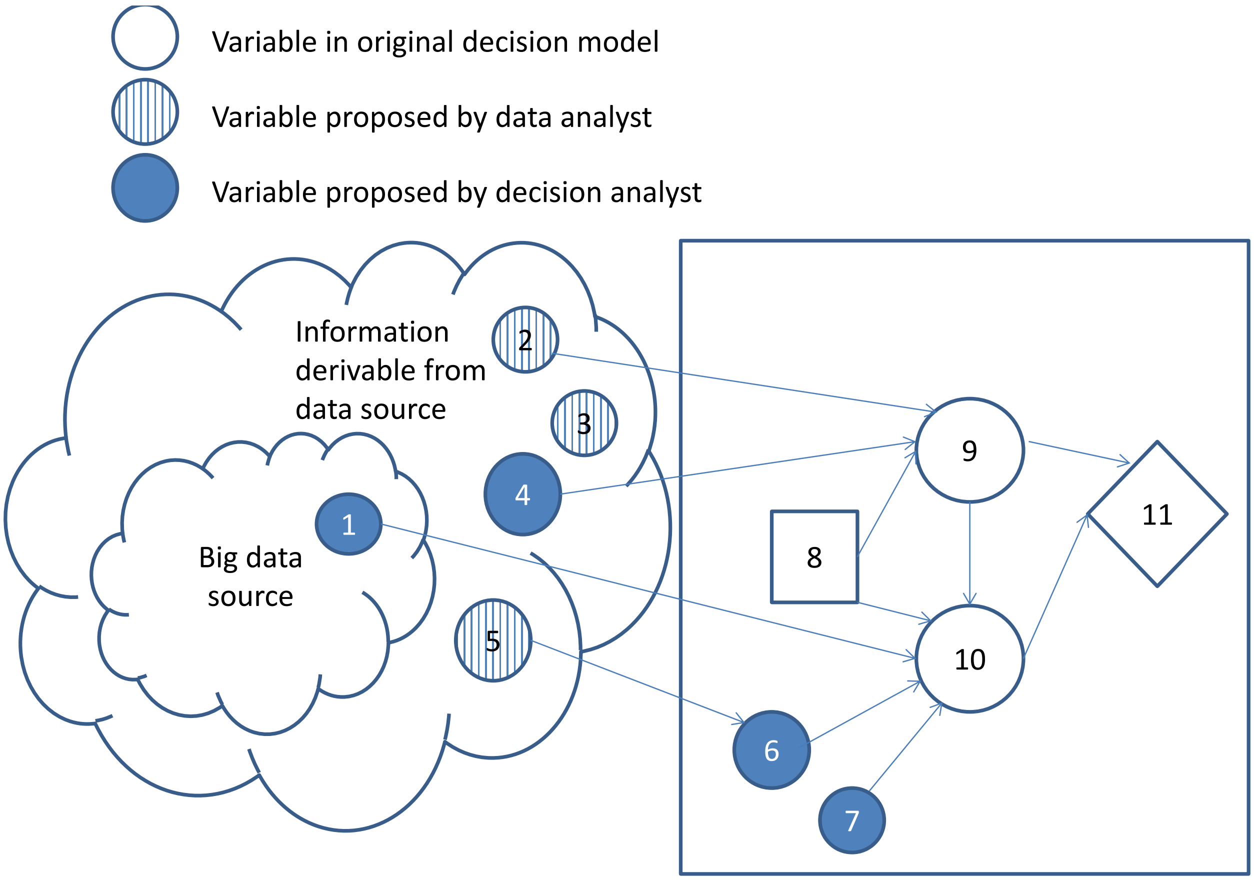
Many decision analysis applications likewise involve building the connections between the very large information space and the decision problem. Other theoretical work provides architecture for connecting big data with particular decision analytic model components. For example, machine learning approaches use Bayes nets to plow through massive data in a structured and dynamic way. Manual approaches to text mining, which could be automated, have been used be used to identify from communications or narratives possible objectives for use in multi-attribute models (4).

Toward a rigorous approach:

The steps above form a sort of process for creating a common language for communicating between a big data universe and the much smaller conceptual world in which the (rational) decision maker operates. In fact, these steps map to a recently developed set of mathematical results building on the Craig Interpolation Theorem (5), which can be stated as: if theory 1 expressed within language 1 and theory 2 expressed within language 2 imply some consequence, then there is some interpolant in the common language of 1 and 2 such that theory 1 implies the interpolant, and the interpolant and theory 2 imply the consequence. This kind of interpolation can be operationalized, as in (6) and (7), as a process of formulating the right language for connecting different bodies of knowledge and information sources.

A report plan, which will ensure the feasibility of decisions among alternatives, includes a set of questions provided by recipients of reports that it is hoped reporters will be able to answer. If the common language has been worked out sufficiently, it is possible to state with certainty situations where this will work. In the case here, sets of the decision analyst first puts out a set of possible questions – what would be the values for the dummy nodes for a particular situation. The data engineer then checks which questions might be answerable (i.e., reported back to the decision model) within the analytics structure if that new term is incorporated into it. Thus, the iterative approach described earlier for building connections between spheres is theoretically justified.

In the figure below, the decision model on the right is connected in this fashion to the data source in the small cloud on the left. The decision analyst first develops the model represented by nodes 8-11. Then the decision analyst proposes variables 1, 4, 6, and 7 as potentially useful and potentially provided by data analysis. The data analyst responds that there is no data to generate a value for variable 6 or variable 7, while variable 4 (derived information) and variable 1 (which turns out to be raw data) are both available and directly useful to the decision model. In addition, the data analyst proposes variables 2, 3 and 5 as potentially useful to the decision model. The decision analyst finds that variable 2 is useful and incorporates it into the model, while variables 3 and 5 are not directly useful. Through iteration, more variables might be proposed by either side or connections discovered. For example although that the new proposed variable 6 cannot be derived directly and the proposed variable 5 is not directly useful, it may be that variable 5 is useful for formulating judgments about variable 6, and together they can then be incorporated into the model.



In addition to drawing the most out of the analytics side, on the decision modeling side, more flexible approaches can ease the challenge of putting information in usable form. For example, models involving economic data might be formulated using random variables that map to a function space (e.g., the space of demand functions or supply functions), in addition to random variables that map to the real numbers or to a set of abstract elements (8).

With large and growing amounts of information coming on-stream all the time, the decision models might consist of modular elements that can be aligned in decision processes for medium size semi-repetitive decisions, e.g., incorporating similar big data in decisions for multiple pharmaceutical products with potential new results to be revealed by analytics about demographics, technical breakthroughs or competition.

Along with the notion of using decision analysis as a front-end for analytics, we can use it as a lens that allows business strategy to leverage analytic capability. Specifically, value of information methods, developed in an age of information scarcity, have potentially different uses in an age of information abundance. If we think of real options as a flip side of the value of information, where the potential to obtain information is combined with the potential to take on or abandon new efforts, we might then think about maximizing the value of those options through a variety of means (9). For example, an organization might develop organizational capabilities along with prospecting for decisions that will maximize the benefit from the information they plan to acquire, thereby specializing in certain kinds of decisions (10).

In sum, by explicitly expanding decision analysis to include connection to sources outside the decision model itself, it should be possible to realize far greater benefit from new developments in analytics by better bringing them to bear on the problems that are important to decision makers.

References:

1. **Raiffa, Howard, Robert Schlaifer.** *Applied Statistical Decision Theory.* Boston, MA : Harvard Business School, 1961.

2. *Combining multi-attribute utility and geographic information for boundary decisions: an application to park planning.* **Keisler, Jeffrey, Ronald Sundell.** 1997, Journal of Geographic Information and Decision Analysis, Vol. 1.2, pp. 101-118.

3. *Enhancing decision analysis models with web-agents.* **Keisler, Jeffrey, Wei Zhang.** 2006, Journal of Decision Systems, Vol. 4, pp. 453-473.

4. *Identifying and structuring the objectives of terrorists.* **Keeney, Gregory L., Detlof Von Winterfeldt.** 2010, Risk Analysis , Vol. 30.12, pp. 1803-1816.

5. *Linear Reasoning.* **Craig, William.** 1957, Journal of Symbolic Logic , Vol. 22, pp. 250-268.

6. *Craig interpolation for networks of sentences.* **Keisler, H. Jerome, Jeffrey Keisler.** 2012, Annals of Pure and Applied Logic, Vol. 163.9, pp. 1322-1344.

7. *Observing, reporting and deciding in networks of sentences.* **Keisler, H. Jerome, Jeffrey M. Keisler.** Annals of Pure and Applied Logic (to appear).

8. **Keisler, Jeffrey M., Erin Baker.** *Decision analysis with random functions and economic models.* unpublished manuscript.

9. **Keisler, Jeffrey M., Paul Mang.** VOI and Real options. *Wiley Encyclopedia of Operations Research and Management Science.* Wiley, 2011, Vol. 1, pp. 138-158.

10. **Keisler, Jeffrey M.** *A framework for organizational decision analysis.* Harvard University, 1992. Unpublished doctoral thesis.