Scenario discovery in Heterogeneously Typed Data

Jan Kwakkel
Faculty of Technology, Policy and Management
Delft University
j.h.kwakkel@tudelft.nl

Scenario discovery is a relatively novel approach aimed at addressing the challenges of characterizing and communicating deep uncertainty associated with simulation models (Dalal, Han et al. 2013). The basic idea is that the consequences of the various deep uncertainties associated with a simulation model are systematically explored through conducting series of computational experiments (Bankes, Walker et al. 2013) and that the resulting data set is analyzed to identify regions in the uncertainty space that are of interest (Bryant and Lempert 2010; Kwakkel, Auping et al. 2013). These identified regions can subsequently be communicated as scenarios. Scenario discovery is an analytical process which can be embedded in a participatory process supporting deliberation with analysis (National Research Council 2009).

A motivation for the use scenario discovery is that the available literature on evaluating scenario studies has found that scenario development is difficult if the involved actors have diverging interests and worldviews (van 't Klooster and van Asselt 2006; Bryant and Lempert 2010). Another shortcoming identified in this literature is that scenario development processes have a tendency to overlook surprising developments and discontinuities (van Notten, Sleegers et al. 2005; Derbyshire and Wright 2013).

Scenario discovery is an approach that aims at offering support for decision making under deep uncertainty. Deep uncertainty is encountered when the different parties to a decision do not know or cannot agree on the system model that relates consequences to actions and uncertain model inputs (Lempert, Popper et al. 2003), or when decisions are adapted over time (Hallegatte, Shah et al. 2012). In these cases, it is possible to enumerate the possibilities (e.g. sets of model inputs, alternative relationships inside a model, etc.), without ranking these possibilities in terms of perceived likelihood or assign a probabilities to the different possibilities (Kwakkel, Walker et al. 2010).

Although scenario discovery can be applied on its own (Gerst, Wang et al. 2012; Kwakkel, Auping et al. 2013; Rozenberg, Guivarch et al. 2013), it is also a key step in Robust Decision Making (RDM) (Lempert, Groves et al. 2006; Lempert and Collins 2007; Dalal, Han et al. 2013; Hamarat, Kwakkel et al. 2013). RDM aims at supporting the design of robust policies. That is, policies that perform satisfactory across a very large ensemble of future worlds. In this context, scenario discovery is used to identify the combination of uncertainties under which a candidate policy performs poorly, allowing for the iterative improvement of this policy. This particular use case of scenario discovery suggest that it could be used also in other planning approaches that design plans based on an analysis of the conditions under which a plan fails to meet its goals (Walker, Haasnoot et al. 2013).

Currently, the main statistical rule induction algorithm that is used for scenario discovery is the Patient Rule Induction Method (PRIM) (Friedman and Fisher 1999), although other algorithms such as Classification and Regression Trees (CART) (Breiman, Friedman et al. 1984) are sometimes used (Lempert, Bryant et al. 2008; Gerst, Wang et al. 2012). PRIM can be used for data analytic questions, where the analyst tries to find combinations of values for input variables that result in similar characteristic values for the outcome variables. Specifically, one seeks a set of subspaces of the model input space within which the values of a single output variable is considerably different from its average values over the entire domain. PRIM describes these subspaces in the form of 'boxes' of the model input space. To identify these regions, PRIM uses a lenient or patient hill climbing optimization procedure. The most frequently employed implementation of PRIM that is being used for scenario discovery is the one provided by Bryant in the scenario discovery toolkit, written in R (Bryant 2012).

There are two problems related to PRIM that are addressed in this paper. First, although originally presented as a regression based rule induction algorithm, in the context of scenario discovery PRIM is typically used on binary data. In contrast to e.g. CART, PRIM cannot be used directly for handling multiclass data (Gerst, Wang et al. 2012; Rozenberg, Guivarch et al. 2013). Second, the lenient hill climbing optimization procedure used in PRIM is not well adapted to cope with a situation where one or more of the inputs are integers or categorical. Friedman and Fisher (1999) outline how PRIM could be adapted to handle these data types. However, as will be argued in more detail, their suggested approach is defective because it partially defeats the lenient character of the algorithm and will not produce the most concise boxes in case of categorical data. The available support for scenario discovery currently does not support categorical data, and contains a work around for integer data.

To address these two problems, we first outline in more detail the PRIM algorithm. We will briefly discuss how Friedman and Fisher (1999) suggest that integer data and categorical data should be handled, and then present an improved version of this approach. To maintain the lenient character of PRIM, the objective function that is being used is adapted. To produce the most concise boxes, the way in which categorical data is handled is modified. To address the problem of multiclass problems, we draw on the way in which CART handles this and show how by adapting the objective function used by PRIM, it can be made applicable also to multiclass problems. The resulting modifications to PRIM are not affecting the efficacy of preprocessing steps such as employed in PCA-PRIM (Dalal, Han et al. 2013).We provide an open source implementation in Python for this modified version of PRIM.

We demonstrate and test the efficacy of the modified algorithm by applying it to several cases in order to compare the performance of the original version of PRIM with the modified version. In particular, we apply it to the same data as used in the original paper of (Bryant and Lempert 2010)[1], the case study of (Rozenberg, Guivarch et al. 2013), and an extended version of the case used by (Hamarat, Kwakkel et al. 2013).

[1] Rob Lempert has by E-mail expressed that he is willing to share the data. I have at the moment of writing not yet received this data. The data from Rozenberg et al and Hamarat et al have already kindly been made available.

References

Bankes, S. C., W. E. Walker, et al. (2013). Exploratory Modeling and Analysis. Encyclopedia of Operations Research and Management Science. S. Gass and M. C. Fu. Berlin, Germany, Springer.

Breiman, L., J. H. Friedman, et al. (1984). Classification and Regression Trees. Monterey, CA, Wadsworth.

Bryant, B. P. (2012). "sdtoolkit: Scenario Discovery Tools to Support Robust Decision Making." from http://cran.r-project.org/web/packages/sdtoolkit/index.html.

Bryant, B. P. and R. J. Lempert (2010). "Thinking Inside the Box: a participatory computer-assisted approach to scenario discovery." Technological Forecasting and Social Change **77**(1): 34-49.

Dalal, S., B. Han, et al. (2013). "Improving Scenario Discovery using Orhogonal Rotations." Environmental Modelling & Software **48**: 49-64.

Derbyshire, J. and G. Wright (2013). "Preparing for the future: Development of an 'antifagile' methodology that complements scenario planning by omitting causation." Technological Forecasting and Social Change.

Friedman, J. H. and N. I. Fisher (1999). "Bump hunting in high-dimensional data." Statistics and Computing **9**(2): 123-143.

Gerst, M. D., P. Wang, et al. (2012). "Discovering plausible energy and economic futures under global change using multidimensional scenario discovery." Environmental Modelling & Software.

Hallegatte, S., A. Shah, et al. (2012). Investment Decision Making Under Deep Uncertainty: Application to Climate Change, The World Bank.

Hamarat, C., J. H. Kwakkel, et al. (2013). "Adaptive Robust Design under Deep Uncertainty." Technological Forecasting and Social Change **80**(3): 408-418.

Kwakkel, J. H., W. L. Auping, et al. (2013). "Dynamic scenario discovery under deep uncertainty: the future of copper." Technological Forecasting and Social Change **80**(4): 789-800.

Kwakkel, J. H., W. E. Walker, et al. (2010). "Classifying and communicating uncertainties in model-based policy analysis." International Journal of Technology, Policy and Management **10**(4): 299-315.

Lempert, R. J., B. P. Bryant, et al. (2008). Comparing Algorithms for Scenario Discovery. Santa Monica, CA, USA, RAND.

Lempert, R. J. and M. Collins (2007). "Managing the Risk of Uncertain Threshold Response: Comparison of Robust, Optimum, and Precautionary Approaches." Risk Analysis **24**(4): 1009-1026.

Lempert, R. J., D. G. Groves, et al. (2006). "A general analytic method for generating robust strategies and narrative scenarios." Management Science **52**(4): 541-528.

Lempert, R. J., S. Popper, et al. (2003). Shaping the Next One Hundred Years: New Methods for Quantitative, Long Term Policy Analysis. Santa Monica, CA, USA, RAND.

National Research Council (2009). Informing decisions in a changing climate, National Academy Press.

Rozenberg, J., C. Guivarch, et al. (2013). "Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation." Climatic Change.

van 't Klooster, S. A. and M. B. A. van Asselt (2006). "Practising the scenario-axes technique." Futures **38**(1): 15-30.

van Notten, P. W. F., A. M. Sleegers, et al. (2005). "The future shocks: on discontinuity and scenario development." Technological Forecasting and Social Change **72**(2): 175-194.

Walker, W. E., M. Haasnoot, et al. (2013). "Adapt or perish: a review of planning approaches for adaptation under deep uncertainty." Sustainability **5**: 955-979.