



PARDEE RAND GRADUATE SCHOOL

Supporting Robust Decisions with Classification and Data-Mining Algorithms

Benjamin Bryant

Advisor: Robert Lempert

Thanks to: Evolving Logic, Inc, RAND Pardee Center, National Science Foundation

useR! 2009

8 July

Outline

- **Policy analysis, robust decisions and the “scenario discovery” concept**
- **The PRIM algorithm as a means to implement scenario discovery**
- **Demo of the ‘sdtoolkit’ PRIM implementation**
- **Future directions**

We are interested in methods to support long-term, deeply uncertain decisions

- **For example:**
 - **Climate change adaptation**
 - **Terrorism risk**
- **Variety of techniques could be applied**
 - **Qualitative scenarios (no formalized mathematical model)**
 - **Probabilistic analysis (optimization and/or risk hedging)**
- **The “Robust Decision Making” (RDM) approach combines quantitative modeling with intuitive appeal of scenarios**
 - **Goal: Find policy options that are robust against all combinations uncertainties**

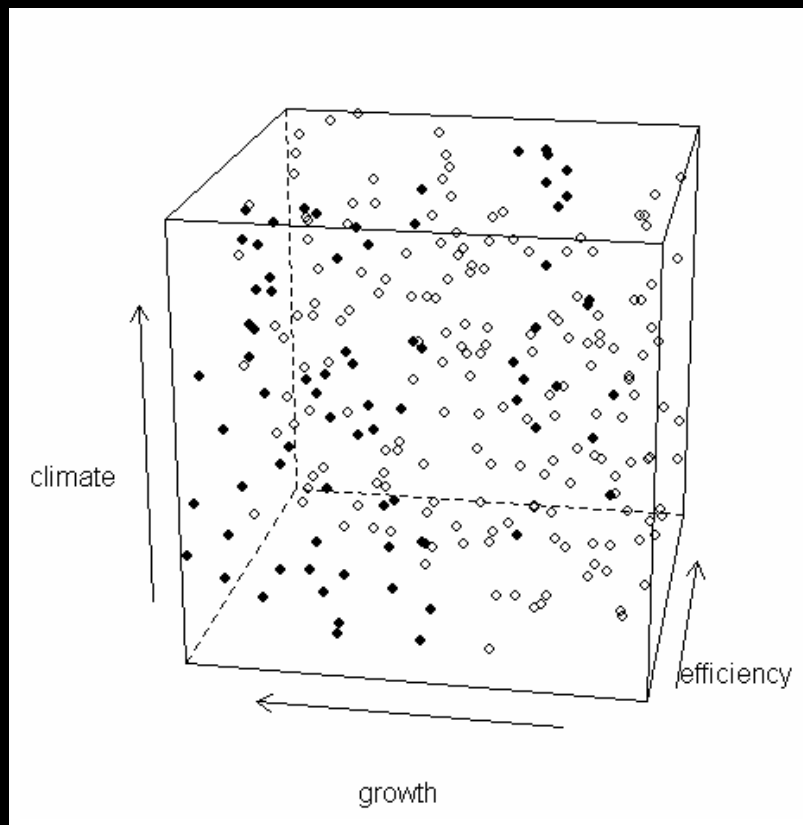
Scenario Discovery is one step in the RDM process



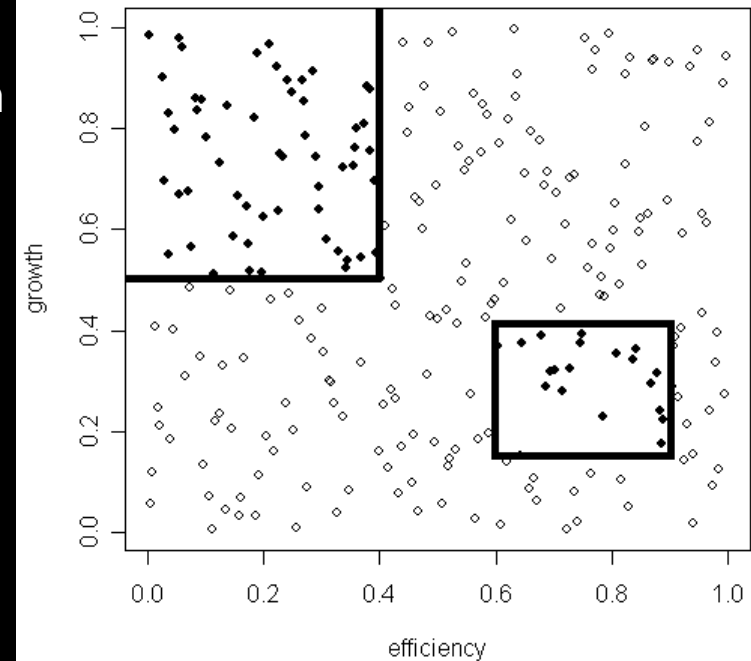
- Views “scenarios” as vulnerabilities of policies: States of the world where policy performs poorly
- Uses a simulation model to examine policy performance over many combinations of uncertainties
- Uses classification and/or data-mining algorithms to find regions of uncertainty space where the policy performs poorly
 - These regions represent possible future states of the world and become quantitatively defined “scenarios”

Current scenario discovery algorithms identify scenarios as 'boxes'

Box = restrictions of parameters describing region of input space



Algorithm
magic



RAND

(filled points = interesting)

*Dataset entirely contrived for illustration

5 8 July 2009

Boxes translate to concise sets of parameter restrictions

- In previous case:

Box 1:

growth > .5

efficiency < .4

Box 2:

.25 < growth < .4

.6 < efficiency < .9

Three measures characterize 'goodness' of box set

Density: Interesting cases (points) captured / Total captured

Coverage: Interesting points captured / Total interesting

Interpretability: Some decreasing function of the number of boxes & dimensions restricted

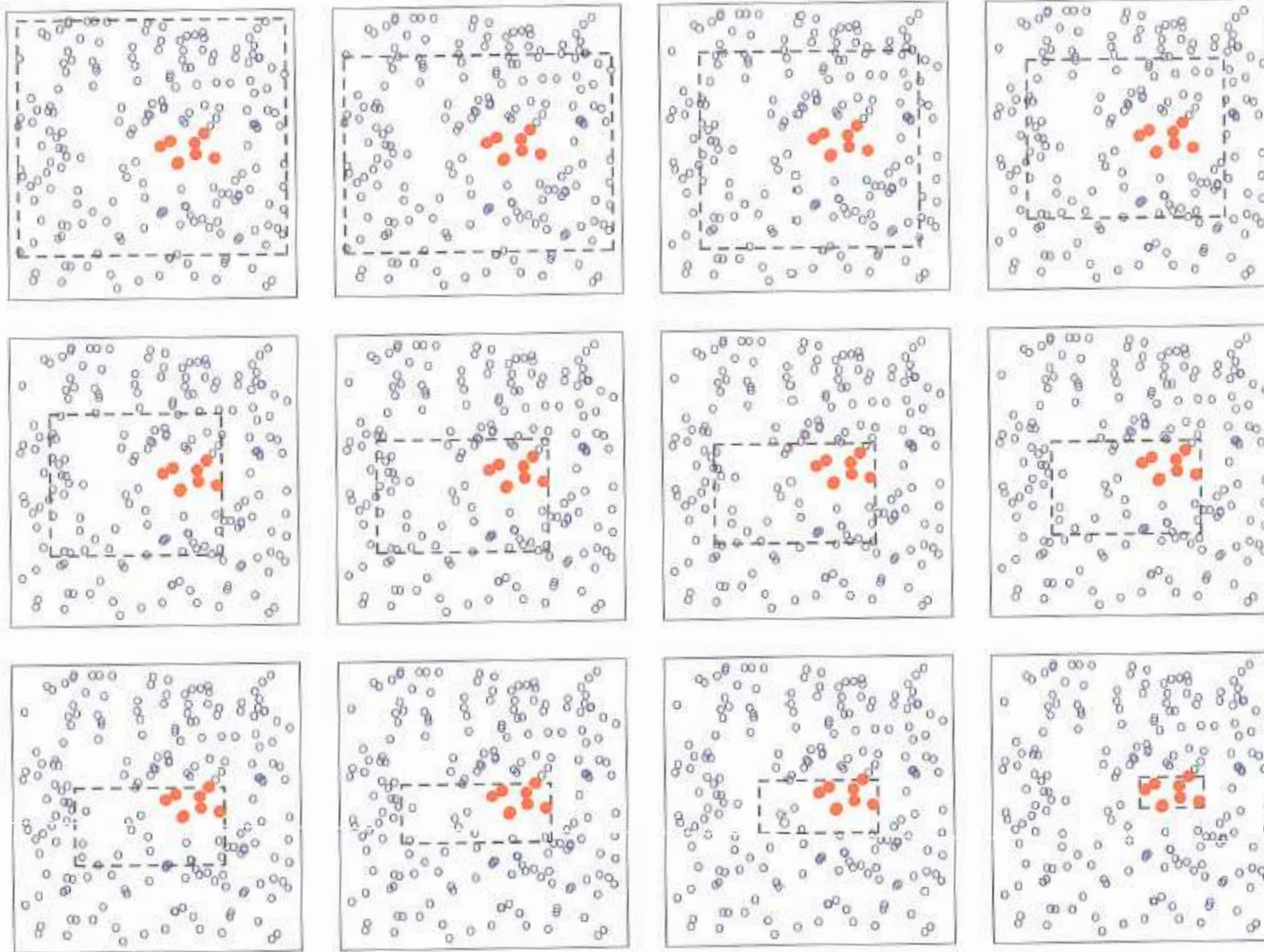
These measures are generally in tension and no all-purpose objective function exists, so:

Seek algorithms to populate an efficiency frontier relating measures.

We use the Patient Rule Induction Method to generate many candidate boxes

- PRIM is a “bump-hunter,” tries to find regions of input space with high output value
- Interactive by design
 - Produces many boxes, provides information to help the user choose among them
- Original version of PRIM not designed for scenario discovery specifically, but we made a few modifications

Prim works by peeling and pasting...



R package 'sdtoolkit' adapts PRIM for scenario discovery

- Long-term idea is to serve as environment for integrating functionality of multiple algorithms, post-processing, and visualization
- Currently implemented only with PRIM, but hopefully incorporate additional algorithms
- At present, toolkit provides the following features:
 - Coverage-oriented statistics and tradeoff curve (in addition to support)
 - Contour plots which indicate dimensionality on the peeling trajectory
 - Automatic generation of 'normalized restriction plots'
 - Automatic generation of color coded scatter plots with boxes drawn
 - Reproducibility and (quasi)-statistical significance tests

Demo of sdt toolkit

There are many potential additions to the scenario discovery interface

- Adding additional box-finding algorithms to toolkit
 - eg, CART
- Generate and sort approaches
- Improved search through box space
- Enhanced visualization of tradeoffs and boxes (3D!)

Even more theoretical work could inform and broaden scenario discovery implementations

- Sampling design
- Relationship of sampling to scenario significance
- Dataset and box diagnostics informed by other data-mining algorithms – esp clustering
- Non-box shapes that are still interpretable
- Interactive sampling/scenario-search for models with prohibitive run time

Thanks!

- Scenario discovery references:
- Bryant, B.P. (2009) “sdtoolkit: Scenario Discovery tools to support Robust Decision Making.” Contributed R package: <http://cran.r-project.org/web/packages/sdtoolkit/index.html>

Bryant, B.P. and R.J. Lempert (2009). Thinking Inside the Box: A participatory, computer-assisted approach to scenario discovery. In revision.

Groves, D.G. and R.J. Lempert (2007) A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, Vol. 17, No 1, 2007, pp 78-85. Available at: <http://www.rand.org/pubs/reprints/RP1244/>

Lempert, R.J, B.P. Bryant and S.C. Bankes. (2008) Comparing algorithms for scenario discovery. WR-557-NSF, RAND Working Paper Series, Santa Monica: Calif. Available at: http://www.rand.org/pubs/working_papers/WR557/

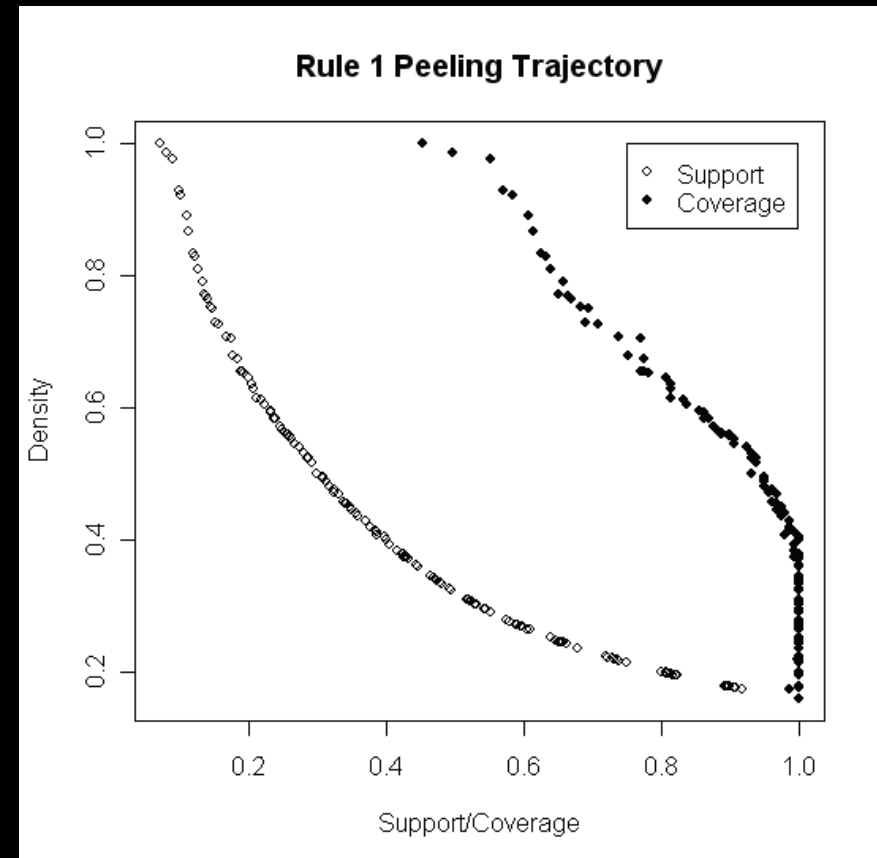
Lempert, Groves, Popper, and Bankes, 2006, A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios, *Management Science*, 52(4). Available at: http://www.rand.org/pubs/library_reprints/LRP20060412/

- PRIM reference:

Friedman, JH. and Fisher, N. (1999) Bump hunting in high dimensional data. *Statistics and Computing*. 9, 123-143.

Practical problems inhibit effective scenario assessment

- Existing algorithm interfaces lack:
 - Coverage oriented statistics and visualization
 - Means to assess significance of dimension restrictions
 - Sufficient interactivity



CART works by partitioning

