accelrys®

SCITEGIC®
PIPELINE PILOT

Access methods for Recursive Partitioning (RP) and Multi-objective Pareto Optimization.

Learn on single or multiple response variables.

Address multiobjective optimization problems.

Advanced Modeling Collection

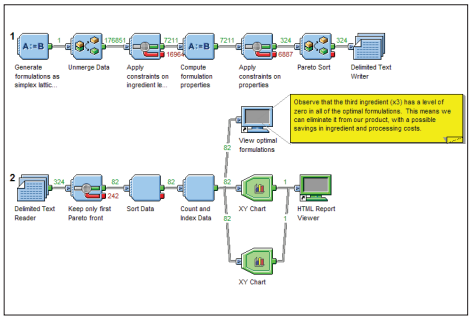
The Pipeline Pilot Advanced Modeling component collection provides methods for Recursive Partitioning (RP) and Multi-objective Pareto Optimization. A variety of RP methods are available in the collection including both single tree and forest of tree learners. The methods can learn on single or multiple response variables. The Pareto Optimization components include methods for multiobjective optimization problems to provide solutions whose criteria trade off amongst two or more partially conflicting goals.

With the Recursive Partitioning Components you can:

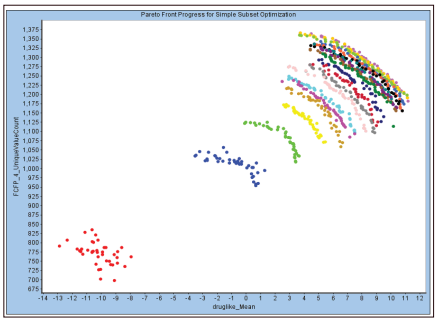
- Perform very rapid learning and data mining experiments on very large datasets with very large number of descriptors
- Learn molecular datasets using fingerprints as descriptors
- Visualize trees to understand the relationships between descriptors and responses
- Analyze descriptor usage to identify the most discriminating descriptors
- Rapidly apply models to new predict new data sets

With the Pareto Optimization Components you can:

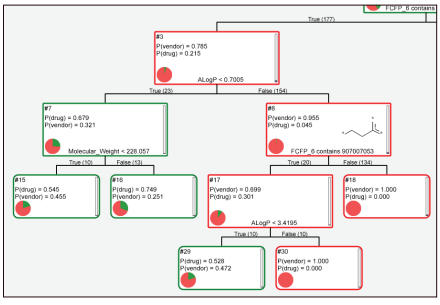
- Optimize solutions for problems as diverse as combinatorial library design, formulation ingredient optimization, or stock portfolio risk management
- Find individual samples with a dataset that have the best tradeoff of desired property values
- Find subsets of samples from a larger dataset that collectively have the best trade-offs between desired property values



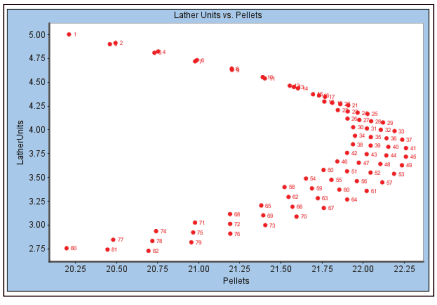
A protocol to optimize formulations with constraints on ingredient levels and properties



ParetoFront progress for simple subset optimization



Visualize decision trees to understand the relationships between descriptors and responses



Optimized formulations with constraints on ingredient levels and properties

Advanced Modeling Component Collection

Learners

Components are provided to learn single tree models, cross-validated single tree models and forest models. Parameters within the components provide extensive control over the learning method, allowing the section of small forests or large random forests as well as the size, depth, splitting, pruning and weighting methods to be applied to the trees. The components allow the specification of single or multiple response (Y) variables to be learned within the tree or forest.

Viewer

The Tree Model viewer provides a highly interactive, web-based visualization of a forest of trees. Each tree allows the user to browse and navigate to the trees within a forest and within parts of a specific tree. The tree display shows the descriptors in use (including graphical display of fingerprint fragments), the proportion of observations within each class and the splitting rules. Drill down from a node shows member observations and the rules that lead to that node. These rules can be automatically translated into PilotScript to be used in a custom filter to identify more records that would obey the same rules.

Optimizers

Components are provided for performing a Pareto sort, a Pareto subset optimization and a Pareto combinatorial optimization. The sort does not select observations but simply ranks observations according to their Pareto score. Criteria that contribute to the score are defined by the user in terms of properties, goals (max, min) and weight.

The Pareto subset optimizer provides a set of subset selections, each of which is on the Pareto front: the optimal trade off between the user defined properties and goals being optimized.

The combinatorial optimizer, as its name suggests, adds in the combinatorial constraint, so rather than select a subset of observations, the user defines a number of contributions from groups that define the product. As an example in combinatorial chemical library design, the user specifies the selection of e.g. an 8x12x20 fully combinatorial library from all possible products of the larger sets of e.g. 100x100x100 reagents in the three R-groups.

SCITEGIC® PIPELINE PILOT

accelrys®

Accelrys Corporate Headquarters
10188 Telesis Court, Suite 100
San Diego, CA 92121–4779, USA
Tel: +1 858 799 5000

Accelrys European Headquarters
334 Cambridge Science Park
Cambridge, CB4 OWN, UK
Tel: +44 1223 228500

Accelrys Asia Headquarters
Nishi-shimbashi TS Bldg 11F
Nishi-shimbashi 3–3–1, Minato-ku,
Tokyo, 105–0003, Japan
Tel: +81 3 3578 3861

0308