ADMEWORKS ModelBuilder Specification

Introduction

ADMEWORKS ModelBuilder is a tool dedicated for building QSAR/QSPR models that can later be used for predicting various chemical and biological properties of compounds. A set of data on molecular structures and their respective experimental values of the property of interest is a prerequisite for every model building. Two classes of models (Qualitative and Quantitative) can be built using various algorithms. The models are based on values of physicochemical, topological, geometrical, and electronic properties derived from the molecular structure.

Models created in ADMEWORKS ModelBuilder are easily imported into ADMEWORKS Predictor (optional product) which is a high-speed virtual (In Silico) screening system intended for simultaneous evaluation of the ADMET properties of compounds. Simultaneous evaluation of the pharmacological as well as the ADMET properties of compounds is useful in the discovery phase to produce balanced quality hits, and also in the lead optimization phase to lessen the occurrence of faulty leads.

Functionality

- Quantitative/qualitative structure-activity relation's analysis.
- Multiple statistical methods for generating predictive models: Perceptron, Iterative Least Squares Method, Multiple Linear Regression, Stepwise Regression, Leap-and-Bounds Regression, Genetic Algorithm, Fuzzy Adaptive Least Squares, k Nearest Neighbors, ADA Boost and Support Vector Machine.
- Customizable cross-validation of models.
- Multivariate/pattern recognition data analysis.
- ADMET data analysis.
- Near 400 predefined descriptors and an unlimited number of substructure-related descriptors.
- Interactive graphs that display property distribution, predicted vs. actual activity, property correlations, clustering of samples and properties and others.
- Interactive graphical feature and outlier selection tools.
- Automated statistical tools for feature and sample selection.
- SDF and CSV file import/export.

Descriptors

The set of descriptors (values of physicochemical, topological, geometrical, and electronic properties derived from the molecular structure) come mostly from ADAPT (the software system created by Professor Peter Jurs and coworkers at the Pennsylvania State University to study molecular structure - biological activity relationships SAR and molecular structure - physicochemical property relationships SPR) and MOPAC (general-purpose semiempirical molecular orbital package developed by Professor James Stewart).

For every descriptor, an additional mathematical operation can be done on its value. Math functions can be applied in the Math functions page.

ADMEWORKS ModelBuilder includes 400 predefined descriptors (2D and 3D) and an unlimited number of substructure-related descriptors. Detailed list of all descriptors is available in separate document.

Feature selection

Feature selection is a technique of selecting a subset of descriptors for building models. ADMEWORKS ModelBuilder offers several tools useful in process of feature selection:

- Missing Value Test
- Zero Test
- Correlation Test
- Automated Correlations Test
- Class Zero Test
- Multicollinearity Test
- Fisher's Ratio Test
- Variance Test
- Genetic Algorithm
- PSO
- Parameter Set Refinement
- Iterative SVMFS
- Iterative SVMFS MIXED
- Boost-LDA FS

QSAR Models

Two types of models can be built in ADMEWORKS ModelBuilder: Qualitative (Discriminant Function) or Quantitative (Multiple Linear Regression). A qualitative model categorizes molecules into classes; an example of such is a model that predicts if a molecule fits in a carcinogenic or noncarcinogenic class. Conversely, a quantitative model produces a numerical value for its prediction. A water-solubility model is a good example of a quantitative model.

Quantitative

- MLR
- Stepwise MLR
- Leaps-and-Bounds MLR
- Interactive MLR
- Logistic Regression
- PLS
- SVM Regression

Qualitative

- Iterative Error-Correction Feedback Perceptron
- Stochastic Gradient Perceptron
- Iterative Least Squares Linear Decision Surface
- KNN
- SVM
- FuzzyALS
- LDA
- RFSBoost-LDA
- Interactive DF

Other tools

- Logical Filter
- Model Wizard

Graphs

ADMEWORKS ModelBuilder offers interactive graphs that display property distribution, predicted vs. actual activity, property correlations, clustering of samples and properties:

- Parameter Correlation
- Class Division
- Radar Chart
- Structure Chart
- Line Graph
- Sample Clustering
- Parameter Clustering
- Principal Components Analysis Graph

Hardware and Software Requirements

Minimum requirements that must be met to install and run ADMEWORKS ModelBuilder:

- Intel Pentium 4 2.0 GHz; 512 MB RAM; 1GB available hard disk space, CD-ROM drive, mouse or pointing device.
- Operating Systems: Windows XP/VISTA/7.