GhostMiner

GhostMiner is unique data mining software that not only supports common databases (or computer-based) and mature machine learning algorithms, but also assists with data preparation and selection, model validation, multimodels like committees or k-classifiers, and visualization. All of this and more is available in one package - a large range of data preparation techniques, a broad scope of selection of features methods and a choice of data mining algorithms and visualization techniques are integrated. This means that only one data format (project) is needed, and so trying out and comparing different approaches becomes extremely easy. The package also comes with an intuitive interface, which should make it easier to use even for non-technical users.

GhostMiner Developer

GhostMiner Developer is the tool for data-model designers and developers who, using databases, can train, test run and select new models. The use of this program requires a good knowledge of statistical analysis and some knowledge of methods of computational intelligence.

- Viewing of numerical and symbolic datasets, pre-processing of data (normalization or standardization), data charting, and recaps of statistical features.
- Assigning new cases to predefined categories using multiple co-variation criteria for the purposes of decision support, diagnosis, classification, evaluation, and heteroassociation of two sets of objects.
- Creating predictive neural, neurofuzzy, decision tree and nearest-neighbor models that are "trained" using a cross-validation procedure for optimal generalization.
- Discovering knowledge in the form of decision trees or logical rules, both crisp and fuzzy.
- Enhancing predictive accuracy of data models through the use of committees of models and the creation of K-classifiers, one for each class.
- Finding missing values in the data and complete patterns with missing parts.
- Automatic discretization of input variables in the case of fuzzy rules.
- Selecting variables important for a given decision or assigning weights to input variables, according to their importance.
- Making discrete continuous inputs providing intervals useful for symbolic description of data.
- Visualizing important relationships in data sets, showing data in relation to other known cases stored in the database.
- Reducing dimensionality of data, and preserving its predictive power.
- Finding clusters of similar data and assigning each record to the appropriate cluster.
- Calculating probabilities of the assignment of new cases to existing categories.
- Finding explanations for such assignments using models that provide logical rules.
- Showing similar cases using interactive visualization and reminding of other possibilities that may often be overlooked.
- Using projects based on committees of models to evaluate ambiguous cases.
- Classifying new datasets based on trained models.

Hardware and software requirements

- Suggested software configuration: Intel Pentium 4 2.0 GHz or higher, 512 MB RAM or more; 500MB HDD space or more.
- Apache/Software Foundation Tomcat 4.1.12 or higher.
- Java JDK 1.4.2 or higher.
- Client
  - Microsoft Internet Explorer 5.5sp2 or higher
  - Netscape Communicator 4.75 or higher.

ADME DB is a database containing the latest and most comprehensive data on interactions of substances with Drug Metabolizing Enzymes and Drug Transporters. It is designed for use in drug research and development, including drug-drug interactions and ADME (Absorption, Distribution, Metabolism and Excretion) studies. The information is organized by category (therapeutic area), drug name, enzyme, reaction, and type. ADME DB is supported by chemical/metabolite structures as well as kinetic values found in the literature. The database is available online and completely searchable by keywords or chemical structures. Advanced searches are also available to support investigational studies on drug-drug interactions.

Human Drug Metabolizing Enzymes Database

- Provides information on specific interactions for a given substance with Human Phase I Enzymes such as P450 (CYP), PMO, AHR, MFO and AO, Human Phase II Enzymes such as Oxidases, UDP, GST, and SULT.
- Contains over 33,000 entries.
- Contains more than 15,700 substances, a number of natural products and preparations, as well as other factors influencing Drug Metabolizing Enzymes activity.
- Contains data collected from more than 6200 citations.

Kinetic Metabolism Database

- Provides information on more than 250 Transporters (including ABCCs, OATPs, OCTs, OATs, BSEP, SLCs) involved in drug transport, physiological compounds, nutrients, and other chemicals and metabolites.
- Contains data collected from ~3280 citations.

Examples of applications of the system

- Evaluating and predicting clinical side effects and interactions of known and new drugs/chemicals/physiological compounds.
- Computer assisted development of new drugs and prodrugs.
- Selecting CYP selective substrates/inhibitors for in vivo/vitro studies.
- Studying properties of natural and artificial CYP mutants, using selective metabolic reactions.

To find additional information, please visit: www.fqs.pl/GhostMiner

To find additional information, please visit: www.fqs.pl/ADME_DB