Data from multiple sources

In many situations large amounts of data are collected. In process industries, e.g., a large number of process parameters are recorded. Other examples are medical studies on patients, planned chemical experiments, or customer-relationship analysis. Often, the data may originate from several different sources. To fully draw advantage of the data, it must be organized in a proper way and interpreted to give the user the required information. The interpretation may then permit some corrective action from the user, e.g., change a certain process parameter.

Situation and current solution - In a typical case, a user has collected data from different data sources into one or more files. The task is to analyze this data and draw the correct conclusions, e.g., to identify some required action. Today, software packages in this area are often specialized for one type of data source, one area of application, or one type of data analysis method. This forces a user to use several different software packages to complete one task.

The limitations – Data from different sources usually come in many different formats, either stored in files or in database tables. Prior to analysis, the user must arrange data from the different sources properly to allow a meaningful interpretation. This often requires a significant amount of manual work, especially in cases where the sampling rates of the data sources differ. To perform the analysis, the user has to know which of all the available methods should be used in the specific situation. For example, a specific statistical test has to be selected to compare two sets of measurements and determine if they are similar, or a specific algorithm has to be selected to create a reliable classification model. This often requires significant expert knowledge in areas that may not be of primary interest to the user.

Proxedra's solution – Proxedra offers fully integrated software that covers the entire chain from data acquisition, management and analysis to data interpretation and on-line control.

The software's modular architecture allows the direct interface of multiple data sources with analysis modules that can be set up to handle a multitude of data analysis problems such as prediction of quantitative variables or class relationship. Proxedra's novel problem-oriented concept builds on that the user chooses between available problem types (scenarios) instead of between available methods/algorithms. A built-in guide in the form of wizards leads the user towards the ideal solution to the problem.

Since the software can easily be directly interfaced to various instruments, there is no need to use different instrument-specific software packages for data collection that provide data in different formats. Instead, all data is collected in one global database, easily accessible by all users in an organization. Together with optional user-security features, this facilitates the implementation of FDA and GMP related procedures.

When multiple data sources are interfaced, they are automatically merged to create one common dataset ready for subsequent analysis. If necessary, the merging step can implement synchronization between data sources by resampling and/or time shifting as illustrated below. It is also possible to extend the dataset with attributes that are calculated using any of the available variables.

Data source #1

Time	Sensor 1
0	0.3471
1	0.4789
2	0.5363
3	0.6371
4	0.7432
5	0.6543
6	0.5423
7	0.4924
8	0.3523

Data source #2

Time	Sensor 2
0	0.0233
2	0.4554
4	0.3456
6	0.1234
8	0.0576

Data source #3

Time	Attribute
0	0.43
3	0.54
6	0.88
9	0.49

Merged resampled data

Time	Sensor 1	Sensor 2	Attribute
0	0.3471	0.0233	0.4300
2	0.5363	0.4554	0.5033
4	0.7432	0.3456	0.6533
6	0.5423	0.1234	0.8800
8	0.3523	0.0576	0.6200

Extended data

Time	Sensor 1	Sensor 2	Attribute	Calculated Attribute
0	0.3471	0.0233	0.4300	1.2300
2	0.5363	0.4554	0.4789	2.4400
4	0.7432	0.3456	0.5533	3.1300
6	0.5423	0.1234	0.6533	4.4400
8	0.3523	0.0576	0.8044	3.2600

Proxedra's software has several key features that facilitate the handling of data from multiple sources. This significantly reduces the costs involved in handling large amounts of data, but the main benefit is that it becomes easier to extract the valuable information that is present (often hidden) in the data.

Proxedra
Teknikringen 7
SE-583 30 Linköping, Sweden
info@proxedra.com
www.proxedra.com