SAP HANA SPS 09 - What’s New?
Predictive Analysis Library
(Delta from SPS 08 to SPS 09)
SAP HANA Product Management
November, 2014
Agenda

Release Theme

List of Algorithms

Framework Changes

New Algorithms

• Back-Propagation (Neural Network)
• Top-K Rule Discovery (KORD)
• Brown’s Exponential Smoothing
• Linear Regression with Damped Trend and Seasonal Adjust
• Forecast Accuracy Measure
• Croston Method
• K-Medians
• Principal Component Analysis

Enhancements

Documentation
Release Theme
Release Theme

The SPS 09 version of the Predictive Analysis Library includes many new algorithms as well as several enhancements to existing algorithms.

These new features were chosen based on the prioritization of customer and other stakeholder requests.
List of Algorithms
SAP HANA In-Memory Predictive Analytics

Predictive Analysis Library (PAL) - Algorithms Supported

- **Association Analysis**
  - Apriori
  - Apriori Lite
  - FP-Growth *
  - KORD – Top K Rule Discovery **

- **Classification Analysis**
  - CART *
  - C4.5 Decision Tree Analysis
  - CHAID Decision Tree Analysis
  - K Nearest Neighbour
  - Logistic Regression
  - Back-Propagation (Neural Network) **
  - Naïve Bayes
  - Support Vector Machine

- **Regression**
  - Multiple Linear Regression
  - Polynomial Regression
  - Exponential Regression
  - Bi-Variate Geometric Regression
  - Bi-Variate Logarithmic Regression

- **Cluster Analysis**
  - ABC Classification
  - DBSCAN
  - K-Means
  - K-Medoid Clustering *
  - K-Medians **
  - Kohonen Self Organized Maps
  - Agglomerate Hierarchical
  - Affinity Propagation

- **Time Series Analysis**
  - Single Exponential Smoothing
  - Double Exponential Smoothing
  - Triple Exponential Smoothing
  - Forecast Smoothing
  - ARIMA *
  - Brown Exponential Smoothing **
  - Croston Method **
  - Forecast Accuracy Measure**
  - Linear Regression with Damped Trend and Seasonal Adjust**

- **Probability Distribution**
  - Distribution Fit *
  - Cumulative Distribution Function *
  - Quantile Function *

- **Outlier Detection**
  - Inter-Quartile Range Test (Tukey's Test)
  - Variance Test
  - Anomaly Detection

- **Link Prediction**
  - Common Neighbors
  - Jaccard's Coefficient
  - Adamic/Adar
  - Katzβ

- **Data Preparation**
  - Sampling
  - Random Distribution Sampling *
  - Binning
  - Scaling
  - Partitioning
  - Principal Component Analysis (PCA) **

- **Statistic Functions (Univariate)**
  - Mean, Median, Variance, Standard Deviation
  - Kurtosis
  - Skewness

- **Statistic Functions (Multivariate)**
  - Covariance Matrix
  - Pearson Correlations Matrix
  - Chi-squared Tests:
    - Test of Quality of Fit
    - Test of Independence
  - F-test (variance equal test)

- **Other**
  - Weighted Scores Table
  - Substitute Missing Values

* New in SPS 08
** New in SPS 09 – Names to be finalized
Framework Changes
Framework Changes

New Wrapper Generator Functions
SYS.AFLLANG.WRAPPER.PROCEDURECREATE
SYS.AFLLANG.WRAPPER.PROCEDUREDROP
Allow/require schema specification

Deprecated
SYSTEM.AFL.WRAPPERGENERATOR / SYSTEM.AFL.WRAPPER_ERASER

New Role
AFLPM_CREATOR_ERASER_EXECUTE

Detailed changes can be found in SAP Note 2046767 and in the HANA PAL manual.
Neural Network

**Back-Propagation (BP)**

Back-propagation is a common method of training neural networks. PAL BP function creates a multi-layered network and supports continuous or categorical input, different activation functions, and generates models for both classification and regression problems.
KORD

KORD algorithm returns top K association rules with highest measures (e.g. lift, leverage). Unlike traditional association rule mining algorithms such as Apriori and FP-Growth, which return all the rules with given measure threshold, the computational complexity of the algorithm is linear with respect to the data size.
Brown Exponential Smoothing

Brown Exponential Smoothing is another exponential smoothing method which forecasts time series with trend but without seasonality. The smoothing factor alpha is used to doubly smooth the original series data. An adaptive alpha option is provided which updates the alpha according to the “new” points in the series.

Currently, only Brown Linear Exponential Smoothing is supported.
Linear Regression with Damped Trend and Seasonal Adjust

Linear regression with damped trend and seasonal adjust is an approach for forecasting when a time series has a trend. In PAL, the algorithm provides a dampened smoothing parameter for smoothing the forecasting value. In addition, if the time series has seasonality and the user provides the length of periods, it can deal with the seasonality by the user specified value, or it can help detect the seasonality and determine the periods.
Forecast Accuracy Measure

Given two time series, the function calculates the measures of the error. In the PAL, the following measures are supported:

- Mean percentage error (MPE)
- Mean square error (MSE)
- Root mean square error (RMSE)
- Error total (ET)
- Mean absolute deviation (MAD)
- Out-of-sample mean absolute scaled error (Out-of-sample MASE)
- Weighted mean absolute percentage error (WMAPE)
- Symmetric mean absolute percentage error (SMAPE)
- Mean absolute percentage error (MAPE)
Croston method

The Croston method is an approach for intermittent demand forecasting. The algorithm consists of two steps. First, exponential smoothing estimates are made on non-zero items. Second, the intervals between zero items are smoothed to predict the future zero intervals.
K-Medians

K-Medians is a variation of k-means clustering. Instead of calculating the mean for each cluster to obtain the center, the algorithm uses the median to help minimize the impact of outliers on the results of the algorithm.
Principal Component Analysis

PCA is a multivariate algorithm aimed at reducing the dimensionality of multivariate data while accounting for as much of the variation in the original data set as possible. This technique is especially useful when the variables within the data set are highly correlated.
Enhancements
Application Function Modeler for SAP HANA
Graphical Re-Design and DataFlow Enhancements with HANA SPS09

Major Application Function Modeler Update

- Enhanced graphical dataflow modeling and programming editor in HANA Studio
  - Support multi-step application function flows
  - Integrates new SAP HANA EIM Services
- Transportable flowgraph design-time model
- Application function flow scenario
  - Integrates Predictive Analysis Library (PAL), Business Function Library (BFL), …, Custom AFLs, R-Scripts
- Information Management scenario
  - Smart Information Management data-flows (batch or real-time) and smart data quality
- Generated run-time objects as stored procedures or tasks, executable via SQLScript
HANA Predictive Analysis Library – What’s New in SPS 09?
Enhancements (1 of 5)

All (Single, Double Triple Exponential Smoothing)
• Provide one additional output table containing measurement of the error
• Add parameter “EXPOST_FLAG” to output forecast without smoothed value of historic data

Single Exponential Smoothing
• Provide an option to automatically adjust the smoothing factor alpha to better fit the points in the time series.

Triple Exponential Smoothing
• Support additive seasonality modeling (in addition to multiplicative seasonality)
• Add parameter “INITIAL_METHOD”, which offers additional initialization method using moving averages to get trend and seasonal component by decomposition
Forecast Smoothing

• Support additive seasonality modeling besides multiplicative seasonality
• Support automatic cycle determination

ARIMA

• Autoregressive integrated moving average with intervention (ARIMAX) algorithm is an extension for ARIMA. Compared with ARIMA, an ARIMAX model provides the internal relationship not only from former data, but also takes external factors into consideration.
• Note: Use the new PAL function (ARIMAXFORECAST) to perform forecasts based on ARIMAX models.
Multiple Linear Regression
• Additional metrics as output: Akaike information criterion (AIC) and BIC (Bayesian information criterion)

Logistic Regression
• Additional output table for calculating Akaike information criterion (AIC)

Linear and Logistic Regression
• Attribute list to allow selected feature columns of training data to be involved in the learning
• Another parameter “DEPENDENT_VARIABLE” could be set to specify the column of the dependent variable.

Random Distribution Sampling
• Support standard deviation (SD) as input
Decision tree
- Support “MIN_RECORDS_OF_PARENT” to control the minimal training records in every non-leaf node
- Support “MIN_RECORDS_OF_LEAF” to control the minimal training records in every leaf node
- CART supports pruning method
- CART supports “USE_SURROGATE” to control the behavior of surrogate split
- C4.5 treats null as a special value

FP-Growth
- Performance improvement
HANA Predictive Analysis Library – What’s New in SPS 09?
Enhancements (5 of 5)

K-Means
- Additional output table for calculating Slight Silhouette of each cluster
- Additional output table for calculating Slight Silhouette for the whole dataset and clusters

DBSCAN
- Support categorical variables
- Performance improvement

Anomaly Detection
- Additional output table for calculating distance between outlier and centers
- Additional output table for calculating centers
Disclaimer

This presentation outlines our general product direction and should not be relied on in making a purchase decision. This presentation is not subject to your license agreement or any other agreement with SAP.

SAP has no obligation to pursue any course of business outlined in this presentation or to develop or release any functionality mentioned in this presentation. This presentation and SAP’s strategy and possible future developments are subject to change and may be changed by SAP at any time for any reason without notice.

This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP intentionally or grossly negligent.
How to find SAP HANA documentation on this topic?

In addition to this learning material, you can find SAP HANA platform documentation on SAP Help Portal knowledge center at http://help.sap.com/hana_platform.

The knowledge centers are structured according to the product lifecycle: installation, security, administration, development:

**SAP HANA Platform SPS**
- What’s New – Release Notes
- Installation
- Administration
- Development
- References

Documentation sets for SAP HANA options can be found at http://help.sap.com/hana_options:

**SAP HANA Options**
- SAP HANA Advanced Data Processing
- SAP HANA Dynamic Tiering
- SAP HANA Enterprise Information Management
- **SAP HANA Predictive**
- SAP HANA Real-Time Replication
- SAP HANA Smart Data Streaming
- SAP HANA Spatial
Thank you

Contact information

Mark Hourani
SAP HANA Product Management
AskSAPHANA@sap.com