Robust Function Discovery and Feature Selection for Life Sciences and Engineering

Mark Kotanchek
Evolved Analytics LLC
3411 Valley Drive
48640 Midland, MI, USA
+1-989-785-05-13

mark@evolved-analytics.com

ABSTRACT

Industrial process and product optimization is impossible without meaningful models and insights on significant features controlling process or product performance. Real-world modeling and feature selection problems have many issues - high-dimensional, non-linear, with unbalanced measurements, correlated features, missing experiments, etc., which makes it difficult for most people to know what the right approach is in any given situation.

We present a function discovery technology based on symbolic regression that routinely converts these problems into meaningful and insightful models with robust driver features identification. Without requiring a Ph.D. in Computer Science or Statistics, it is now possible to easily develop robust nonlinear models (complete with trust measures), identify data outliers and interactively explore the model dynamics and response sensitivities.

Our presentation will illustrate the ease and power of automatic conversion of a spreadsheet of data into an interactive data story report using examples drawn from life sciences and engineering.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: *Model Development.*

General Terms

Algorithms, Documentation, Performance, Design.

Keywords

Data analysis, data analytics, feature selection, variable selection, genetic programming