

Industrializing Data Mining, Challenges and Perspectives

Françoise Fogelman-Soulié

KXEN, France

Business Intelligence is a very active sector in all industrial domains. Classical techniques (reporting and Olap), mainly concerned with presenting data, are already widely deployed. Meanwhile, Data Mining has long been used in companies as a niche-technique, reserved for experts only and for very specific problems (credit scoring, fraud detection for example). But with the increasing availability of large data volumes (in particular, but not only, from the Web), companies are more and more turning to data mining to provide them with high added-value predictive analytics. However producing models in large numbers, making use of large data volumes in an industrial context can only happen if solutions to challenges, both theoretic and operational, are found: we need algorithms which can be used to produce models when datasets have thousands of variables and millions of observations; we need to learn how to run and control the correct execution of hundreds of models; we need ways to automate the data mining process.

I will present these constraints in industrial contexts and show how KXEN has exploited theoretical results (coming from Vladimir Vapnik's work) to provide answers to the above-mentioned challenges. I will give a few examples of real-life applications and will conclude with some remarks on the future of data mining in the industrial domain.