Grand Challenges in Data Mining: From the Technical to the Pragmatic

Usama Fayyad
Chief Data Officer & Sr. VP
Yahoo! Inc.

Data Mining has received much attention as companies and organizations started to ask how they can better utilize the huge data stores they built up over the past two decades. While some interesting progress has been achieved over the past few years, especially when it comes to techniques and scalable algorithms, very few organizations have managed to benefit from the technology. This paradoxical situation of having too much data and be unable to utilize it or mine it arose because of both technical and business challenges. We will cover these challenges, paint a picture for where the data problems are, and cover some of the pragmatic issues. While emphasis in the academic community has been focused primarily on developing new algorithms, few people have paid attention to the problem that algorithms cannot get to data today in a usable form. Of particular interest are the challenges of how to make the technology really work in practice: considering the business setting and the realities of the possible — including how to evolve the role of data into a more strategic position. We shall also cover applications in the EBusiness setting to illustrate the challenges and contributions of data mining as an example of the type of organization where the role of data has become so central that it has gained a strategic role that is critical to the business. I will also discuss some of this evolution which brings challenges to all of us on how to think about data and how to present it in a context that transcends the traditionally purely technical framework and extends it into the core business strategy realm. Finally, since there are still many unsolved deeper technical and scientific problems in this field, we conclude by revisiting the technical challenges facing the field.