

From Mining the Web to Inventing the New Sciences Underlying the Internet

Usama Fayyad¹

Yahoo!, California, USA

Summary. As the Internet continues to change the way we live, find information, communicate, and do business, it has also been taking on a dramatically increasing role in marketing and advertising. Unlike any prior mass medium, the Internet is a unique medium when it comes to interactivity and offers ability to target and program messaging at the individual level. Coupled with its uniqueness in the richness of the data that is available for measurability, in the variety of ways to utilize the data, and in the great dependence of effective marketing on applications that are heavily data-driven, makes data mining and statistical data analysis, modeling, and reporting an essential mission-critical part of running the on-line business.

However, because of its novelty and the scale of data sets involved, few companies have figured out how to properly make use of this data. In this talk, I will review some of the challenges and opportunities in the utilization of data to drive this new generation of marketing systems. I will provide several examples of how data is utilized in critical ways to drive some of these capabilities. The discussion will be framed with the More general framework of Grand Challenges for data mining : pragmatic and technical.

I will conclude this presentation with a consideration of the larger issues surrounding the Internet as a technology that is ubiquitous in our lives, yet one where very little is understood, at the scientific level, in defining and understanding many of the basics the Internet enables : Community, Personalization, and the new Microeconomics of the web. This leads to an overview of the new Yahoo ! Research organization and its aims : inventing the new sciences underlying what we do on the Internet, focusing on areas that have received little attention in the traditional academic circles. Some illustrative examples will be reviewed to make the ultimate goals more concrete.

¹Dr. Usama Fayyad is Yahoo!'s executive vice president of Research & Strategic Data Solutions. He also oversees the Yahoo! Research organization that includes offices in Sunnyvale, Burbank and Berkeley, CA as well as New York, Europe, and S. America.

Prior to joining Yahoo!, he co-founded and led the DMX Group, a data mining and data strategy consulting company. In early 2000, he co-founded and served as CEO of digiMine Inc. (now Revenue Science, Inc.). He's professional experience also includes five years spent leading the data mining and exploration group at Microsoft Research. From 1989 to 1996 Fayyad held a leadership role at NASA's Jet Propulsion Laboratory (JPL) where his work garnered him the top research excellence award that Caltech awards to JPL scientists, as well as a U.S. Government medal from NASA.

Fayyad earned his Ph.D. in engineering from the University of Michigan, Ann Arbor (1991), and also holds BSE's in both electrical and computer engineering (1984); MSE in computer science and engineering (1986); and M.Sc. in mathematics (1989). He has published over 100 technical articles in the fields of data mining and Artificial Intelligence, is a Fellow of the American Association of Artificial Intelligence, has edited two influential books on the data mining and launched and served as editor-in-chief of both the primary scientific journal in the field and the primary newsletter in the technical community published by the ACM. He regularly delivers keynotes and talks at government, industry and academic conferences around the world.