



Most of the data created in human history was actually generated in the past handful of years. “Every person in the world on average generates and consumes gigabytes of text, video, Internet media, images, and music every year,” Associate Professor Tony Jebara said.

Jebara’s specialty is developing machine learning programs that sift through massive amounts of data to discover underlying patterns and make accurate predictions. “I work at the intersection between statistics and computer science, applying machine learning tools to massive data sets where the relationship between variables is often not deterministic. Our algorithms must be fast, because computer speeds are not growing as rapidly as the amount of data they must handle.”

Computers slice through data that would take humans years to analyze. Yet their capabilities are only as good as the underlying algorithms, the set of rules used to classify and analyze data. Computers, for example, find it hard to identify faces, a task babies master within months.

This is an area where Jebara made his start by building one of the top face recognition algorithms. His approach to face recognition used probability distributions to calculate the likelihood that two images were of the same individual. Jebara also worked on extending the standard Bayesian algorithms to minimize error rather than maximize likelihood.

Most recently, Jebara has been working on matching and graph algorithms, two promising ways of learning from massive datasets, such as those generated by social networks and the Web. Viewing large amounts of data as a graph often provides a faster and powerful way to solve problems such as data labeling and partitioning.

Also, graphs allow algorithms to be implemented very efficiently by such techniques as message passing, which Jebara has worked on extensively. He has built programs that automatically visualize, label, partition and match data in large data sets, ranging from images to social networks.

Similar algorithms also power Sense Networks, a startup Jebara founded in 2006 to analyze data from telecommunications companies. By tapping smartphone calls and GPS data, Jebara’s algorithms can classify people by behavior patterns. Users can then query the network to see where people with similar tastes go to eat, drink, or shop. The phone company can use the data to filter recommendations and provide targeted advertising.

It is one more example of machine learning finding patterns in a world awash with data.

*B.S., McGill, 1996; M.S., Massachusetts Institute of Technology, 1998; Ph.D., 2002*

*Finding Patterns in  
a Complex World*

**TONY JEBARA**

Associate Professor of Computer Science