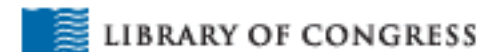




SPECIAL FEATURE



Computing Research that Changed the World: Reflections and Perspectives

March 25, 2009 | 8:45 am – 5:00 pm | Members' Room, Thomas Jefferson Building, Library of Congress

[Back to Main Page](#)

Learning to Improve Our Lives

[DAPHNE KOLLER – Stanford University](#)[Slides – 4.7 MB](#)[Download – 256 MB](#)[Watch the Talk \(17:34\)](#)

For a long time people believed that computers could do only what they were programmed to do. We now have many examples where computers can learn to predict and learn to improve their predictions. Computer learning – called Machine Learning (ML) by computer scientists – can help improve all our lives through practical applications such as speech understanding, fraud detection, intrusion detection, image search, automated surveillance for suspicious actions, and many more. There are many important applications of Machine Learning.

Spam constitutes 85–95% of all email, at an estimated cost of \$13B in 2007 in the US alone. Spammers are constantly adapting. How can we learn to detect spam? Typical filters will look for particular words or patterns – "!!!!", "offer," "pharmacy," etc. – then add parameters that weight each example, and then decide whether the item is spam or not by seeing if the weighted sum exceeds a threshold. These parameters are learned by an ML application. Much of this learning is adaptive, and can include personalization for a particular user.

Typical Mail-piece Address Block :

ZIP and Primary Number Recognition

ZIP Code: 14213 Database Query

Primary Number: 74

Lexicon Entry (Street Name)	Addon
BRADLEY ST	1024
COLONIAL CIR	1467
DEWITT ST	1534
GRANT ST	1905
HAMPSHIRE ST	2012
HAWLEY ST	1053
LAFAYETTE AVE	1347
LIVINGSTON ST	1653
MANCHESTER PL	1251
PO BOX	0074
POOLEY PL	1128
POTOMAC AVE	1146
PUTNAM ST	1650

Machine translation is a text-based application with much more complex functionality: the input cannot be converted to a "bag of words", and the output has to be a coherent sentence. Rule-based methods for machine translation floundered for decades. In a classic example, "the spirit is willing but the flesh is weak" was automatically translated into Russian and back and came out as "the vodka is good but the meat is rotten." Modern ML methods have allowed us to build sophisticated translation systems by using parallel text corpora, with millions of matching sentences in pairs of languages. These text corpora can also help systems learn what good text looks like in addition to what words/structures correspond, leading to much higher quality translations.

Automated handwriting analysis seems easy but there are many ways to write each number or letter. Using a learning-based system developed at SUNY Buffalo by Venu Govindaraju and colleagues, 25 billion letters a year are processed automatically by the US postal service – bar-coded for precise deliver – saving hundreds of millions of dollars.

Machine learning systems can also deal with more than a single sensor. One example application, developed at Microsoft by Eric Horvitz and colleagues, automatically combines sensors for weather and for traffic speeds with incident reports. This system can learn to predict current and future congestion on roads other than those that are directly sensed, and can also produce dynamically optimized routes for users.

In systems constructed at Stanford by Andrew Ng and colleagues, machine learning enabled a robot dog to learn to traverse complex terrain and climb stairs. Another system has learned to control helicopters in complex aerobatic maneuvers. A future application for this technology is to control the smart electric grid. The recent Northeast blackout was caused by primitive control systems sending power into already overloaded systems. ML can provide smarter control to get clean energy from where it's produced to where it's needed on a capacity-limited grid.

Current work on applying machine learning to medical care tackles diverse problems, for example: learning the standard of care in a hospital, warning when medical errors are being made, or tracking patients over time, whether in an intensive care unit or over longer periods in a home setting, to provide early warning of complications and allow early intervention. A prototype Microsoft-constructed system intended to help a parent worried about a child in the night would enable triage by less-experienced people. Evidence-based medicine aims to use electronic medical records to figure out which treatments work. But the factors that determine what works are complex and differ across individuals. ML systems learn the mapping from genetic and environmental factors to treatment response, to help determine which treatments are most likely to work for each individual. A system under development at MIT by John Guttag and colleagues automatically detects the onset of epileptic seizures with high accuracy. The system is connected to a device that introduces electric current into the brain of an epileptic patient just prior to a seizure, reducing its severity.

In the past few years, new technologies such as high-throughput sequencing, proteomics, and imaging have revolutionized biology, allowing us to obtain millions of data points that provide insight into biological systems. Machine learning is playing a critical role in extracting meaningful conclusions from these data. For example, only 0.1% of our DNA differs from person to person, giving rise to much of the variation in our appearance, health, and response to treatment. Only a few of the changes in the DNA are meaningful, and ML is critical in identifying the important ones. Machine learning systems are the only feasible approach for finding functional elements in our genome.

Machine learning is like computing on steroids. For any complex task that we might tackle, there are usually relevant data that we can use for training. A learning computer system is almost always better than one designed solely by a human. Machine learning with relevant data can improve just about any application. A little machine learning can go a long way!

[SHARE](#)