What I Learned About Machine Learning --Revisited (Again)

Robert P Sheridan

Merck, Whitehouse Station, NJ, United States
eagseags@hotmail.com

Machine learning is a set of techniques to summarize "rules" from a number of examples and use the rules to predict new examples. QSAR is the name given to the application of machine learning to chemistry. In thirteen years Merck has gone from a place where QSAR was hardly ever used to one where it plays a very central role. Development of QSAR at Merck has three aspects:

1. Science: Finding the combination of descriptors and statistical methods that gives the best predictions for the least computational cost.
2. Infrastructure: Having the hardware and software to efficiently build and update models and have users be able to easily get and interpret predictions.
3. Applications: Finding new places where QSAR could be applied in discovery chemistry and, more recently, in process and analytical chemistry.

Much progress has been made on all three fronts, although there are many unresolved issues.