

Active and Proactive Machine Learning Methods

Jaime Carbonell
Language Technologies Institute, CMU
<http://www.cs.cmu.edu/~jgc/>

Abstract:

Active learning is a “hot” research area because labeled data (or ranked data, or structured data) is relatively scarce in many applications, whereas unlabeled data abounds. We present some recent improvements in multi-criterial active learning and move to address some inherent limitations of most active sampling approaches, specifically the single, infallible, indefatigable, uniform-cost oracle one. Proactive learning copes with multiple “oracles” exhibiting different accuracies, different regions of expertise, variable costs, and non-uniform willingness to cooperate in a unified framework by casting the problem as joint optimization in selecting oracles and instances to maximize estimated information value within cost constraints. This is joint work with Pinar Donmez.

Short Bio:

Dr. Jaime Carbonell is the Director of the Language Technologies Institute and Allen Newell Professor of Computer Science at Carnegie Mellon University. He received SB degrees in Physics and Mathematics from MIT, and MS and PhD degrees in Computer Science from Yale University. His current research spans multiple aspects Language Technologies and Machine Learning: text mining, machine translation, active learning (where he coined the phrase “proactive learning” to include external experimentation and multi-source learning), automated summarization (where he invented the MMR search-diversity method), question answering, etc. Dr. Carbonell’s most recent work focuses on core challenges in machine learning and context-based machine translation, and in new opportunities where computational learning can have a major impact, e.g. computational immunology and computational proteomics. He edited or co-edited four machine learning books, and served as editor-in-chief of the Machine Learning Journal. Dr. Carbonell has served on multiple governmental advisory committees such as the Human Genome Committee of the National Institutes of Health, the Oakridge National Laboratories Scientific Advisory Board, the National Institute of Standards and Technology Interactive Systems Scientific Advisory Board, and the German National Artificial Intelligence (DFKI) Scientific Advisory Board. He has edited or co-edited 4 books in machine learning, and published some 250 scientific papers in various subjects.