

**Dr. Jaime Carbonell**

**Proactive Machine Learning and its Application**

**Abstract**

We are in the golden age of statistical machine learning, with new powerful methods proliferating and an ever-increasing array of useful applications. However, the bottleneck to applying machine learning to many practical problems is the paucity of accurately-labeled training data and the cost in time and money to obtain human expert judgments or conduct definitive experiments. Active learning strives to find the most informative instances to label by an external “oracle”, but to achieve widespread practicality we must do more, i.e.: cope with multiple external information sources (experts, crowds, experiments, observations, etc.), estimate their reliability, their availability and their cost, and jointly optimize selections of instances and sources of expertise in an amortized setting to maximize learning in any given time horizon. This joint-optimization process is called Proactive Learning; we discuss how to do it for increasingly complex cases. Then we touch on several applications, including crowd-source learning for machine translation, wind-turbine farm optimization, and proteomic analysis.

**Bio**

Dr. Jaime Carbonell,  
Director Language Technologies Institute  
Carnegie Mellon University

CV see under [www.cs.cmu.edu/~jgc](http://www.cs.cmu.edu/~jgc)