

Machine Learning for Text Classification Applications

David D. Lewis
Independent Consultant
Chicago, IL, USA

Research on machine learning for text classification has exploded in the past decade. Practical applications have followed, in areas such as knowledge management, process improvement, customer service automation, text mining, alerting, intelligence and law enforcement, information feeds, spam and porn filtering, bioinformatics, and survey research.

After introducing text classification as an information processing task, I will discuss how machine learning can be used to reduce human effort in this task. Text classification poses particular challenges for machine learning, and I will look in detail at how algorithms for learning both rule-based and statistical/numerical classifiers are affected. I will also emphasize techniques for data preparation and transformation that are critical to success in operational settings.

Text classification is far from a solved problem, and I will conclude by discussing areas where further progress is likely to depend on contributions from computational linguistics.