Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms

by **Thorsten Joachims**, *Dept. of Computer Science, Cornell University, NY, USA*

Text Classification, or the task of automatically assigning semantic categories to natural language text, has become one of the key methods for organizing online information. Since hand-coding classification rules is costly or even impractical, most modern approaches employ machine learning techniques to automatically learn text classifiers from examples. However, none of these conventional approaches combines good prediction performance, theoretical understanding, and efficient training algorithms.

Based on ideas from Support Vector Machines (SVMs), *Learning To Classify Text Using Support Vector Machines* presents a new approach to generating text classifiers from examples. The approach combines high performance and efficiency with theoretical understanding and improved robustness. In particular, it is highly effective without greedy heuristic components. The SVM approach is computationally efficient in training and classification, and it comes with a learning theory that can guide real-world applications.

Learning To Classify Text Using Support Vector Machines gives a complete and detailed description of the SVM approach to learning text classifiers, including training algorithms, transductive text classification, efficient performance estimation, and a statistical learning model of text classification. In addition, it includes an overview of the field of text classification, making it self-contained even for newcomers to the field. This book gives a concise introduction to SVMs for pattern recognition, and it includes a detailed description of how to formulate text-classification tasks for machine learning.

Learning To Classify Text Using Support Vector Machines is designed as a reference for researchers and practitioners, and is suitable as a secondary text for graduate-level students in Computer Science within Machine Learning and Language Technology.