

Summary of Text Categorization Methods (cont.)

- How to design effective features? (application-specific)
 - Analyze the categorization problem and exploit domain knowledge
 - Perform error analysis to obtain insights
 - Leverage machine learning techniques (e.g., feature selection, dimension reduction, deep learning)
- How to obtain “enough” training examples?
 - Low-quality (“pseudo”) training examples may be leveraged
 - Exploit unlabeled data (using semi-supervised learning techniques)
 - Domain adaptation/transfer learning (“borrow” training examples from a related domain/problem)