

# Application: Boosting for Text Categorization

[with Singer]

- **weak classifiers**: very simple weak classifiers that test on simple patterns, namely, (sparse)  $n$ -grams
  - find parameter  $\alpha_t$  and rule  $h_t$  of given form which minimize  $Z_t$
  - use efficiently implemented exhaustive search
- “How may I help you” data:
  - 7844 training examples
  - 1000 test examples
  - categories: AreaCode, AttService, BillingCredit, CallingCard, Collect, Competitor, DialForMe, Directory, HowToDial, PersonToPerson, Rate, ThirdNumber, Time, TimeCharge, Other.