Dekang Lin

Unsupervised Acquisition of ‘Deep’ Knowledge from Shallow Corpus Statistics

• Use ngram to acquire deep knowledge
  – Preposition (benefit * the community: of, to)
  – Context sensitive spell checker (cite, sight, site)
  – Non-referential pronoun recognition
    • make it in advance -> make them in advance
    • Make it in Hollywood -> make them in Hollywood*
  – Gender probability
    • Ababibu explained himself, Takeru bought his car
    • Noun * pronoun/possesive: Pat (58%M, 31%F)
  – Anti-reflexive
    • John needs his friend (his=John’s)
    • John needs his support (his /= John’s)
    • Count compatibility for “* needs * support”
  – Distributional similarity based on PMI -> clustering