Motivation


Goal: Extract structured relations from web text.

Uses: Smarter search, question answering, analytics, etc.

Distant Supervision

How can we leverage existing knowledge bases (say from Wikipedia) to expand them?

Problem: We don’t have ‘alignments’ between sentences and facts.


Approach: Model relation expressed by sentences as a latent variable. Train via EM.

System Performance

Our system was in the top 5 at the TAC-KBP competition organized by NIST with an F1 score of 31.7% (versus the best submission which had a score of 37.27%).

Error analysis on the 2010 dataset:

Relational Inference

Idea: Can neighbouring relations to help us infer new relations?

Approach: Bayesian logic program style factor graph inference.

Results: 1% F1 improvement from 31.7% to 33%.

Other Directions

Entity Linking: plays a major role in accuracy of predictions.

Harry S. Dent Jr is an American financial newsletter writer.

Harry S. Dent Sr. as American political strategist.

Harry Dent is also the son of the fictional Batman character Harvey Dent.

Approach: Train a classifier using web-page context and Wikipedia links.

Joint Inference: using the DeepDive system.

Solve the problem of entity linking, coreference, named entity recognition and relation extraction jointly.