Indirect Supervised Relation Extraction

• ReQuest: Indirect Supervision for Relation Extraction using QA Pairs.
  • Hypothesis from QA:
    • A positive QA entity mention pair’s embedding vector should be more similar (closer in the low-dimensional space) to any other positive QA entity mention pair, than to any negative QA entity mention pair of the same question.
  • Incorporate knowledge from both QA and RE datasets to jointly learn feature representations to infer label of test instances
  • Accepted to WSDM’18
Indirect Supervised Relation Extraction

• Follow-up: improve scalability & generalization.
  • Leverage hidden insights from more general QA datasets (e.g., SQuAD, TriviaQA, etc):
    • In a QA pair, the relation asked by the question should be the same as the relation answered by the answer sentence.
  
• Experiment with more RE models (tough baselines, Zhang et al., 2017) & datasets (NYT, KBP, TACRED)
• Current Status: no significant improvement observed yet; need more error analysis and experiments
Faceted Taxonomy Construction

• Extend SetExpan to HiExpan (hierarchy expansion): target KDD
  • Current Pipeline:
    • In each iteration:
      • Use SetExpan to expand the children set under each parent node in a hierarchy
      • Use a modified version of SetExpan to get the seeds for each parent node that’s not provided by the user
    • At the end of each iteration:
      • If conflicts detected, resolve conflicts and prune the tree
      • Store nodes being pruned and treat as negative examples for next iteration
    • Stop if no change or reach maximum number of iterations
  • Current Status:
    • Preliminary results on Wiki
    • Need to refine the current pipeline (conflict resolution, tree pruning, negative example handling)
Miscellaneous

• Submitted a paper to KBCOM workshop in WSDM’18:
  • Improving Question Answering Sentence Ranking by Rank Propagation
• Prepared code, project page & camera ready version for ReQuest
• Took courses in NLP, Information Retrieval
• PhD application