Semester Summary

Xiang

Dec. 13
summary

• Coursework: CS 412, STAT 542, etc.
• Review: WSDM’13, WWW’13
• Papers: IJCV(accepted), TIP (under review)
• Project (cont.)
Project Description

• Construction of Heterogeneous Information Network
  – Data Sources: news articles from multiple news agencies, telling about ones certain story/event.
    • Redundancy, incompleteness, conflict
  – Goal: a clean heterogeneous information network which can help summarize a concise and complete story from the news articles.
Current Progress

• Apply NLP tools on news articles (unstructured text)
  – Semantic role labeling
  – Information Extraction (name entity, type info., co-reference, time and location extraction, prediction)
• Sentence clustering – redundancy removal
  – SRL-based item set representation for sentence
  – Frequent-pattern-based feature
  – Sentence clustering with KL-divergence measure
• Other approaches
  – Relational co-clustering, etc
Next Step

• Sub-network Construction: construct HIN from each sentence cluster
• Integrate into complete HIN
• Multi-document summarization
• Submission to SIGIR, KDD, etc.
Open to discussion and collaboration on this project

• text refinement by low-rank approximation (utilize redundancy)

• Sentence clustering via multi-view techniques (SRL-based, more semantic)

• Interested at matrix-related problems 😊
Happy Holiday!

Thank you