2016 Spring Summary

• Outlier Document Detection

  – **Objective:** mining outlier documents in a corpus
  – Developed an Embedded vMF Allocation model to discover topics based on word embedding
  – Designed an outlierness measure to effectively find outlier documents

  – Collaboration with Fangbo
• Comparative Exploration in Multi-dimensional Text Data

  – **Objective:** Discover insightful comparison of text cells in a multi-dimensional text data set

  – Conducting comparison based on topic distributions inferred from word embedding

  – A series of efficiency optimization for finding pairs in massive data

  – Collaboration with Fangbo
• Other collaboration
  – Truth Finding
    • Aspect mining (with Meng Jiang)
  – Crowdsourcing
    • Leveraging correlation between question types (with Aditya and Dan)