Semester Summary

Lucas Liu @ DMG
May 5
Research – Heterogeneous Supervision for RE

• Collaboration with Xiang, Qi, Zhi and Huan. Submitted to EMNLP-2017.

• Goal:
  • Heterogeneous Supervision for Relation Extraction:
    • Relation Extraction: Hussein \((e_1)\) was born in Amman \((e_2)\) on 14 November 1935. \(\rightarrow\) born_in
    • Heterogeneous Supervision: Utilizing Labeling function to generate annotation and integrate different supervision sources (i.e. knowledge base and heuristic patterns).

    \[
    \begin{align*}
    \lambda_1 & \text{return } born\_in \text{ for } <e_1, e_2, s> \text{ if } \text{BornIn}(e_1, e_2) \text{ in KB} \\
    \lambda_2 & \text{return } born\_in \text{ for } <e_1, e_2, s> \text{ if } \text{match}('* \text{ born in } *', s) \\
    \lambda_3 & \text{return } died\_in \text{ for } <e_1, e_2, s> \text{ if } \text{match}('* \text{ killed in } *', s) \\
    \lambda_4 & \text{return } died\_in \text{ for } <e_1, e_2, s> \text{ if } \text{DieIn}(e_1, e_2) \text{ in KB}
    \end{align*}
    \]
Research – Heterogeneous Supervision for RE

• Collaboration with Xiang, Qi, Zhi and Huan. Submitted to EMNLP-2017.

• Challenge:
  • Relation Extraction itself is challenging;
  • Heterogeneous Supervision often conflicts with each other:
Research – Heterogeneous Supervision for RE

• Collaboration with Xiang, Qi, Zhi and Huan. Submitted to EMNLP-2017.
• Task:
  • Inferring true labels from noisy labels;
  • Conducting relation extraction;
Research – Heterogeneous Supervision for RE


- Task:
  - Inferring true labels from noisy labels:
    - Avoid global source consistency assumption. (since labeling functions may be more accurate in a subset comparing to the rest);
  - Conducting relation extraction:
    - Utilizing representation learning to handle overwhelming text features (~1 million).

- Our solution:
  - Conducting both task with representation learning
Research – Heterogeneous Supervision for RE

• Collaboration with Xiang, Qi, Zhi and Huan. Submitted to EMNLP-2017.
• Our solution:
  • Conducting both task with representation learning
Research - Other Collaboration

• Embedding-based local event detection method:
  • Joint work with Chao, Dongming, Quan, Honglei. Submitted to KDD-2017.

• Expert Finding with Locally-trained Embeddings:
  • Joint work with Huan, Qi, etc. Submitted to PKDD-2017.
Research - Ongoing

• Automatically Fine-grained Relation Extraction:
  • Fine-grained structure may exists for some relation types (e.g., None):
    • For KBP, the following should be identified as Not-Target-Type (None):
      • *Michelle Obama* was the 44th first lady of the United States and wife of U.S. President *Barack Obama*.
      • *President Obama* made a historic trip to *Cuba* this week.
  • Solution:
    • Conducting Clustering (generative model) and Classification (discriminative model) in an unified framework

• Efficient Online Short Text Topic Modeling:
  • Embedding based text clustering method demonstrates great potential on short text;
  • Further improved it and extend it into a paper.
Misc

• IE 521: Convex Optimization

• Optimization Reading Group organized by Prof. Ruoyu Sun & Prof. Niao He from IE:
  • Covering many related machine learning / optimization techniques, e.g. GAN.