Motivation

Current Knowledge Graph (KG) construction requires analysts to generate extensive schemas in order to be effective for any new data source. Automated KG construction approaches need to address issues such as:
- Extensive raw document pre-processing
- Missing relation inferencing
- Coherent relation building
- Scalable entity searching and tracking

Our efforts attempt to address these issues to facilitate an automated pipeline.

Enablement

These efforts enable analysts and knowledge workers to:
- Build Knowledge Graphs without extensive schema construction
- Infer attributes (group membership, sentiment, etc.) logically from text and graph data without analyst-intensive schema generation.
- Separately, these efforts allow for improved analysis in Natural Language Processing and Graph Analysis.

Contextual Embeddings
For KG Completion

Sample Knowledge Graph

Framework

Contextual Relation Types

Entity Prediction for Freebase Dataset

Challenge:
- Can we develop a scalable algorithm to infer missing entities and relationship in KGs?

Approach:
- **Contextual Embeddings (ContE)**, is able to capture interactions between relation types to improve both entity and relation type predictions.
- The key idea is to use contextual relation types of a triple in addition to the triple itself into the embeddings (i.e., latent feature vector) learning process.

Contributions:
- The reduced number of learning epochs are required
- Scalable to large KGs
- High accuracy for both entity and relation type predictions
- Higher accuracy (up to 27%) against state-of-the-art methods on two real-world KG datasets (Freebase and YAGO)

Deep Learning Enabled Sentence Modeling

Challenge: Can we efficiently pinpoint important text for KG building to decrease processing time?

Approach:
- Learn text from sequential and hierarchical aspect simultaneously
- Soft-align both aspects of text to intellectually model sentence iteratively

Contributions:
- Automated key sentence extractor: text filtering & doc. summarization
- Sentiment detection on Tweets

Building KG Using Topic Modeling

Challenge:
- Can we cluster entities based on topic modeling for building KG?

Approach:
- Use latent feature vectors to structure topic model
- Lifelong learning: Mine word correlation knowledge from topics generated and incorporate these knowledge to generate new topics, which are used to mine new knowledge iteratively

Contributions:
- Cold start KG builder
- Large corpus reduction and summarization

Scalable Multi-modal Information Fusion for KG

Challenge:
- Can we create a scalable data fusion pipeline that can allow for easier knowledge propagation between different data modalities (KGs, images, etc.)?

Approach:
- Model different data modalities as an arbitrary tuple (nodes in a multi-layered graph)
- Expand the loss function of KG embedding to capture relationships beyond pairwise relationships in the standard euclidean space (versus the kernel space)
- Use classification techniques to infer new relationships

Contributions:
- Scalable knowledge and relationship discovery algorithm for multi-layered graphs