The Frontier of Knowledge Embedding

Miao Fan

Ph.D. Candidate

Department of Computer Science and Technology

Tsinghua University

fanmiao.cslt.thu@gmail.com

Road Map

- 1. Knowledge Graph Embedding (KGE) (**Fundamental Theory**).

- 2. Distantly Supervised Relation Extraction via Exploiting Embedded Knowledge (**Application**).

- 3. Open Question Answering via Leveraging Embedded Knowledge (**Application**).

Precursors

To catch the main points of this presentation, the audience need to know those basic knowledge:

- 1. Stochastic Gradient Descent (SGD).

- 2. Derivative Formula (Vector & Matrix).

- 3. Hinge Loss Function.

- 4. Margin-based Rank Loss.

Knowledge Graph Embedding (Theory)

- **Reference Article:**
  - *Translating Embeddings for Modeling Multi-relational Data*

- **Download Link:**
  - http://www.thespermwhale.com/jaseweston/papers/CR_paper_nips13.pdf

- **Key points:**
  - Embedding **entities (h, t)** and **relationships (r)** of multi-relational data **(h, r, t)** in low-dimensional vector spaces.
  - The intuition is the **hierarchical** and **in-reflexive** relation instances (h, r, t).
  - The model ||h+r-t||.
  - The algorithm.

Distantly Supervised Relation Extraction with Embedded Knowledge (Application)

- **Reference Article:**
  - *Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction.*

- **Download Link:**
  - http://hal.archives-ouvertes.fr/docs/00/88/04/55/PDF/weston13emnlp.pdf

- **Key points:**
  - (Distantly) Weakly Supervised Relation Extraction.
  - Relation-mention and-ship scoring function $S_{m2r}(\cdot)$
  - Triplet scoring function $S_{(h,r,t)}$.
  - The combined function.

Open Question Answering with Embedded Knowledge (**Application**)

Reference Article:

Open Question Answering with Weakly Supervised Embedding Models

Download Link:

http://arxiv.org/pdf/1404.4326.pdf

Key points:

Open Question Answering.

Training Corpus Generation.

Question-Triplet Similarity.

The Promising Further Work

- 1. **Continuous representation (Embedded Encoding)** of knowledge (which is traditionally logical formed.)

- 2. Easily **conducting similarity computing** (Semantic Computing).

- 3. However, how to encode the word, entity, relation or the others (questions) into the same embedding space without losing the paraphrase?