Scribe Note: Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text

Presenter: Bill Zhang, Scribe: Ji Gao

June 1, 2019

1 Research Question

Research question: Finding answers to open domain questions in natural language.

Difficulties:

- Large amount of data
- Hard to combine the information from knowledge bases and text articles

2 Overview Figure

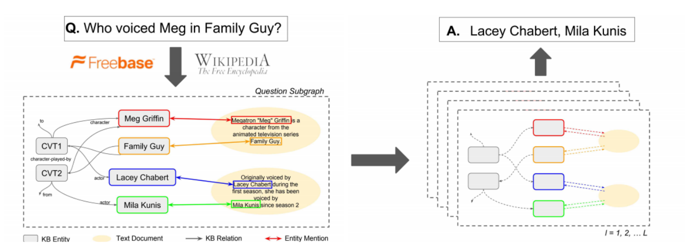

Figure 1: **Left:** To answer a question posed in natural language, GRAFT-Net considers a heterogeneous graph constructed from text and KB facts, and thus can leverage the rich relational structure between the two information sources. **Right:** Embeddings are propagated in the graph for a fixed number of layers (L) and the final node representations are used to classify answers.

3 Method Overview

- Overview:
  - Use graph structures to represent Knowledge base, and combine text articles into the graph
  - Effectively select a subgraph for a particular input.
  - Learn a heterogeneous graph network to represent the graph

- Before learning: Prepare a knowledge base. Also, create an article dataset from Wikipedia.
- Train a pair of question q and answer a in two steps:
  1. Extract a subgraph $\mathcal{G}_q \subset \mathcal{G}$ which contains the answer to the question with high probability.
  2. Use proposed model **GRAFT-Net** to learn graph representations in $\mathcal{G}_q$, and do the classification

4 GRAFT-Net

- In one sentence: A large Inductive Graph Recurrent model with multiple heterogeneous embedded inputs.
- Setting:
  - Knowledge Base is a graph where $\mathcal{K} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. $\mathcal{V}$ is the set of vertices. $\mathcal{E}$ is the set of edges. $\mathcal{R}$ is the set of relations, which is a property of every edge.
  - Articles $\mathcal{D} = \{d_1, d_2..d_D\}$, where each document is a sequence of words $d_i = w_{d_1}, w_{d_2}, ..., w_{d_{|d|}}$.
  - Query q: a sequence of words $w_{q_1}, w_{q_2}, ..., w_{q_{|q|}}$.
  - Answer $\{a\}$: a set of entities.
  - Problem: Give $(\mathcal{K}, \mathcal{D}, q)$, find a
- **Subgraph(Input to the GRAFT-Net):**

  A graph represents by triplet $\mathcal{G}_q = (\mathcal{V}_q, \mathcal{E}_q, \mathcal{R}^+)$.
  - $\mathcal{V}_q$ includes two parts:
    1. A subset of $\mathcal{V}$, retrieved from top-K PageRank in $\mathcal{K}$.
    2. A subset of $\mathcal{D}$, retrieved by bag-of-word model and search engine.
  - $\mathcal{E}_q$ is generally edges between $\mathcal{V}_q$:
    1. Subgraph of $\mathcal{V}_q$ in $\mathcal{K}$. That is, $(s, o, r) \in \mathcal{E} : s, o \in \mathcal{V}_q, r \in \mathcal{R}$
    2. Links between documents and entities in $\mathcal{V}_q$, represents by (v, d, r_L) : $v \in \mathcal{K} \cap \mathcal{V}_q, d \in \mathcal{V}_q$, and r_L is a special relation represents a link between document and entity.
  - $\mathcal{R}^+ = R \cup \{r_L\}$
- **Key equation:** Updating rule for entity vertices in the graph $\mathcal{G}_p$

$$h_v^{(l)} = \text{FFN}\left(\begin{bmatrix} h_v^{(l-1)} \\ h_q^{(l-1)} \\ \Sigma_r \Sigma_{v' \in N_r(v)} \alpha_r^{v'} \psi_r(h_{v'}^{(l-1)}) \\ \Sigma_{(d,p) \in M(v)} H_{d,p}^{(l-1)} \end{bmatrix} \right) \tag{1}$$

- Explanation of Equation 1 :
  - **FFN**: feed-forward network, i.e., $FFN(x) = \sigma(Wx + b)$ similar to other recurrent structure. The weights are shared among vertices.
  - l: A number indicate current step in recurrent neural network. It keeps increasing from 0 to L.
  - h_v: The hidden state of entity v. $h_v^{(0)}$ is initialized either randomly or with some knowledge base embedding.
  - h_q: The hidden state of whole query. $h_q^{(0)}$ is initialized with an LSTM. h_q is then updated with the combination of entities mentioned in the question.

$$h_q^{(l)} = FFN(\sum_{v \in S_q} h_v^{(l)}) \tag{2}$$

  - Words in documents $H_{d,p}$: The hidden state of a document is the combination over hidden states of every words in the document, and is uppdated using a different recurrent network synchronously. Each word is updated with the hidden state of linking entities.

$$H_{d,p}^{(l)} = \text{FFN}(H_{d,p}^{(l-1)}, \sum_{v \in L(d,p)} h_v^{(l-1)}) \tag{3}$$

  - Aggregate Neighbors: The third term in Equation 1 aggregates the neighbor hidden states.
    - Relations r: Each relation is modeled using a relation vector x_r
    - $a_r^{v'}$ is an attention weight, which is calculated by the query hidden state h_q and relation embeddings $a_r^{v'} = \text{softmax}(x_r^T h_q^{(l-1)})$
    - ψ_r is a weighted function:

$$\psi_r(h_{v'}^{(l-1)}) = pr_{v'}^{(l-1)} \text{FFN}(x_r, h_{v'}^{(l-1})) \tag{4}$$

.

    - $pr_{v'}$ is a scalar weight calculated using PageRank.

$$pr_v^{(0)} = \begin{cases} \frac{1}{|S_q|}, \text{ if } v \in S_q \\ 0, \text{ otherwise} \end{cases}$$

$$pr_v^{(l)} = (1 - \lambda) pr_v^{(l-1)} + \lambda \Sigma_r \Sigma_{v' \in N_r(v)} \alpha_r^{v'} pr_{v'}^{(l-1)}$$

- **Output of GRAFT-Net**:

After l iterations, a probability over all entities is calculated by

$$Pr(v \in \{a\}_q | \mathcal{G}_q, q) = \sigma(w^T h_v^{(l)} + b)$$

.

5 Evaluation:

5.1 Dataset

WikiMovies-10K and WebQuestionsSP.

- WikiMovies-10K is a subset of dataset WikiMovies[1] containing questions on movies.

  Knowledge Base and Document corpus is gathered from Wikipedia by [1].

- WebQuestionsSP[2] collects 4737 questions over *Freebase* entities.

5.2 Models in comparison

- KV-KB: Key Value Memory Networks[1] on only KB input.
- KV-EF: Key Value Memory Networks with both KB and text.
- GN-KB: GRAFT-Net only on KB.
- GN-LF: late-fusion GRAFT-Net, trained on text/KB seperately and then combine.
- GN-EF: Main model.
- GN-EF+LF: Ensemble over GN-EF and GN-LF.

5.3 Result

1. Get best result with full knowledge base on GN-EF+LF model, as in Table 2.

Model	Text Only	KB + Text			
		10 %	30%	50%	100%
WikiMovies-10K					
KV-KB	–	15.8 / 9.8	44.7 / 30.4	63.8 / 46.4	94.3 / 76.1
KV-EF	50.4 / 40.9	53.6 / 44.0	60.6 / 48.1	75.3 / 59.1	93.8 / 81.4
GN-KB	–	19.7 / 17.3	48.4 / 37.1	67.7 / 58.1	**97.0 / 97.6**
GN-LF		74.5 / 65.4	78.7 / 68.5	83.3 / 74.2	96.5 / 92.0
GN-EF	**73.2 / 64.0**	75.4 / 66.3	82.6 / 71.3	87.6 / 76.2	96.9 / 94.1
GN-EF+LF		**79.0 / 66.7**	**84.6 / 74.2**	**88.4 / 78.6**	**96.8 / 97.3**
WebQuestionsSP					
KV-KB	–	12.5 / 4.3	25.8 / 13.8	33.3 / 21.3	46.7 / 38.6
KV-EF	23.2 / 13.0	24.6 / 14.4	27.0 / 17.7	32.5 / 23.6	40.5 / 30.9
GN-KB	–	15.5 / 6.5	34.9 / 20.4	47.7 / 34.3	66.7 / 62.4
GN-LF		29.8 / 17.0	39.1 / 25.9	46.2 / 35.6	65.4 / 56.8
GN-EF	**25.3 / 15.3**	31.5 / 17.7	40.7 / 25.2	49.9 / 34.7	67.8 / 60.4
GN-EF+LF		**33.3 / 19.3**	**42.5 / 26.7**	**52.3 / 37.4**	**68.7 / 62.3**

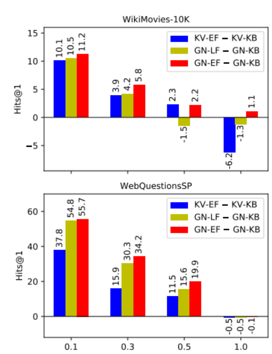

Table 2: **Left:** Hits@1 / F1 scores of GRAFT-Nets (GN) compared to KV-MemNN (KV) in KB only (-KB), early fusion (-EF), and late fusion (-LF) settings. **Right:** Improvement of early fusion (-EF) and late fusion (-LF) over KB only (-KB) settings as KB completeness increases.

2. Get comparable result to state-of-the-art with only text or KB.
3. Without heterogeneous update the performance is worse.

References

[1] Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. Key-value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126*, 2016.

[2] Wen-tau Yih, Matthew Richardson, Chris Meek, Ming-Wei Chang, and Jina Suh. The value of semantic parse labeling for knowledge base question answering. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 201–206, 2016.