Paper Review

Question Answering on Knowledge Bases and Text using Universal Schema and Memory Networks

Summarizing:
The paper used the Universal schema, the combination of knowledge base (KB) and context and align and jointly embed these two parts in a key-value memory network as a matrix, with each slot containing one question-answer pair. The model has better result on SPADES fill-in-the-blank question answering dataset than only training with text or KB. It also outperforms the current state-of-the-art model.

Key Scientific Question and Main Contributions:
The paper try to solve Question Answering (QA) by answering questions on a knowledge base and text with memory network. Specifically, given a question with a blank, how can we fill the blank with accurate answer based on KB and text.

The main contribution:
1. The combination model gives better result than just use KB or text alone.
2. The model outperforms the current state-of-the-art model with higher F1 points on SPADES dataset.
3. The author shows how individual source helps each other and leads to a more comprehensive result.

Technical contributions and relevant work:
The model uses distributed representation to encode the entities and relation. In addition, it includes key-value MemNN to gives the model greater flexibility for encoding knowledge sources and helps shrink the gap between directly reading text and answering from a KB. The attention weight is computed by key and the value is for predicting the answer. Bidirectional LSTM is applied to the training set(key) and question to ensure both forward and backward training.

For the attention part, the model only consider the memory cells that contain at least one entity in the question to save time. It then calculates the context vector iteratively by using the attention weights and values of memory cells. The final contextual vector is used to determine the answer by softmax.

Evaluation setup and Outcomes:
Four models are compared, ONLYKB, ONLYTEXT, ENSEMBLE and UNISCHEMA. The word, entity and relation embeddings, and LSTM states dimension were set to $d = 50,$ and initialized with word2vec. The network weights were initialized using Xavier initialization. Optimization method is Adams with the default hyperparameters, and regularization is L2 norm. Mini-batch and batch normalization is applied.

UNISCHEMA outperforms the other three. Specifically, it is robust even in poor resource situation. It also outperforms the current state-of-the-art model.

Strength and weakness:
Strength: Combination of KB and text jointly is vital here. It is clever to use distributed representation to encode the textual relations into structured data. Key-value MemNN makes the model more flexible.

Weakness: UNISCHEMA model is no better than ONLYTEXT in resource-scarce scenario.