COMP90042 Web Search & Text Analysis

Workshop Week 8

Zenan Zhai

April 30, 2019

University of Melbourne

Outline

Information Extraction

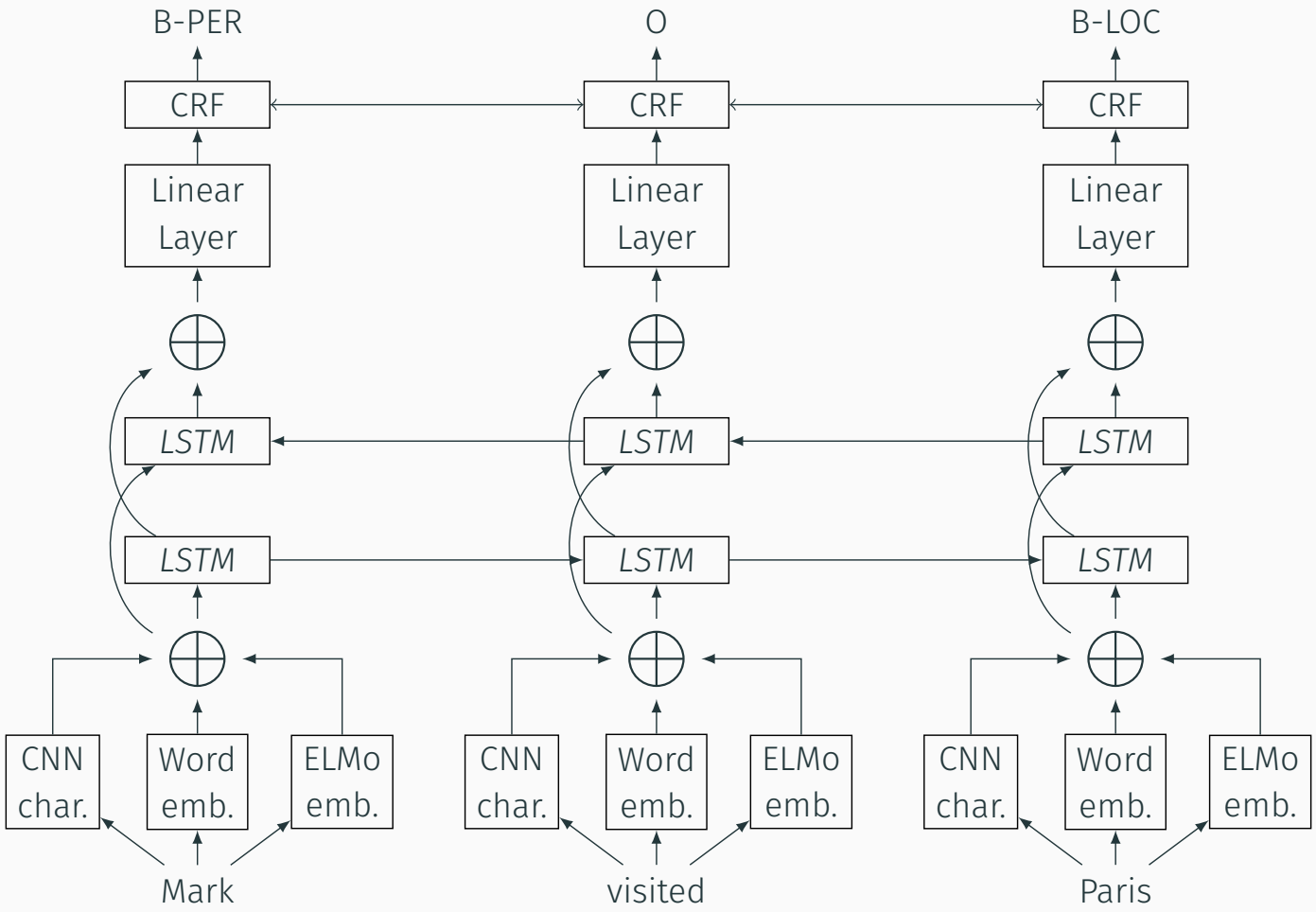
- Named entity recognition
- Relation extraction
  - Rule-based
  - Supervised
  - Semi-supervised
  - Distant supervision
  - Unsupervised
- Temporal/Event extraction
- Template filling

Question Answering

- Information retrieval
- Reading comprehension by Deep learning
- Form query to knowledge base

Named entity recognition (NER)

Sequence Labeling with named entity tags. (PER, ORG, etc)
Using IOB, IOBES tagging scheme.



Relation Extraction

Supervised settings

- Fixed set of relations.
- A sentence with multiple named entities.
- Why is rule-based method inadequate?
- Semi-supervised V.S. Distant Supervision

Unsupervised settings

- No relation set.
- How to group relations?

Other IE tasks

Event extraction

- What happened ?
- Timeline of events.

Temporal extraction

- Extract time related expressions (last month, next year, etc)
- Normalize to canonical form (e.g. next year $\rightarrow$ 2018)

Template filling

- Combination of above task.
- **[supermarket]** reduces price of **[item]** by **[number]**, **[time]**.

Outline

Information Extraction

- Named entity recognition
- Relation extraction
- Temporal/Event extraction
- Template filling

Question Answering

- Information retrieval
- Reading comprehension by Deep learning
- Form query to knowledge base

QA and information retrieval

QA steps:

- Retrieve relevant documents.
- Retrieve relevant paragraphs.
- Extract the answer from relevant paragraphs.

Which step is more difficult?

How can we do this?

Reading comprehension by Deep Learning

Bidirectional Attention Flow (BiDAF) (Seo et al., 2016)

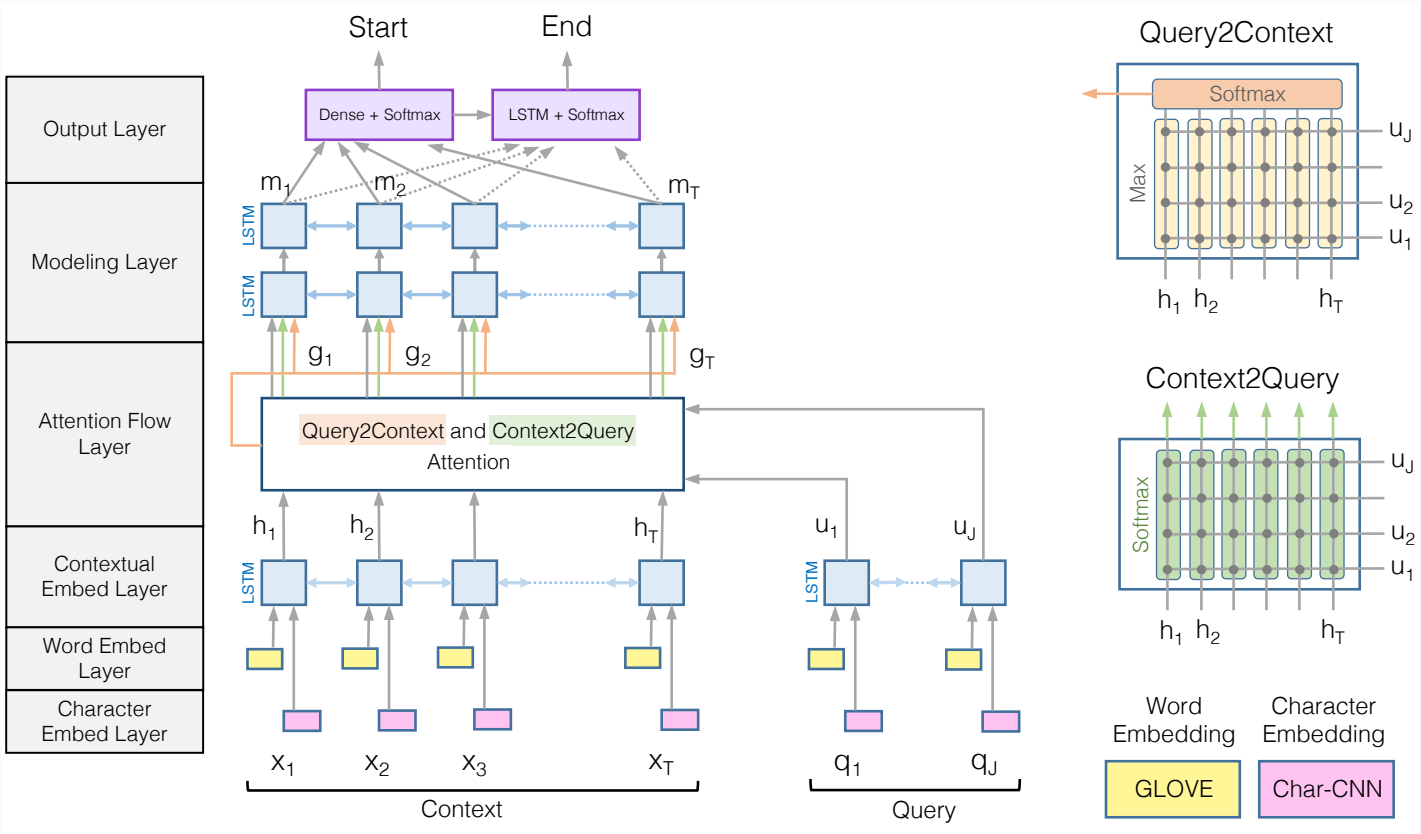


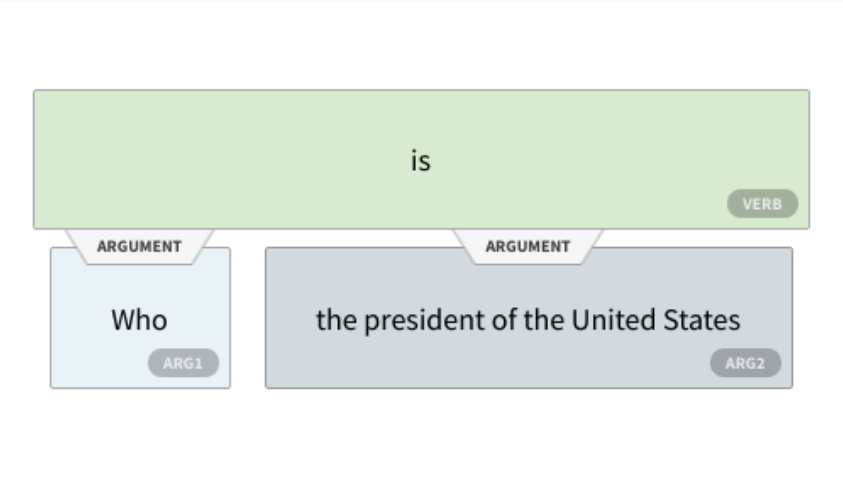
Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

QA in structured knowledge base

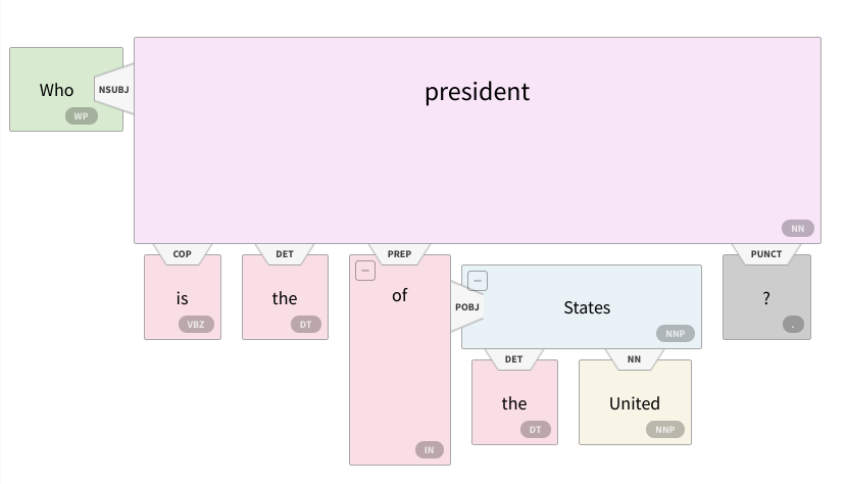
Process query to get logical form.

Compare logical form with structured data in knowledge base.

Derive logical form of query by semantic parsing.



Semantic parsing



Dependency parsing