Multi-Choice Question Answering System of WIP at the NTCIR-12 QALab

Bingfeng Luo, Yuxuan Lai, Lili Yao, Yansong Feng, and Dongyan Zhao

Institute of Computer Science and Technology, Peking University

Introduction

World History Questions of Japanese University Entrance Exam (English Subtask)

From1-4 below, choose the one sentence that is correct in regard to the person/people that it describes.

(1) Ouyang Xiu and Su Shi are writers representative of the Tang period.
(2) Yan Zhenqing is a calligrapher representative of the Song period.
(3) Wang Anshi, who lived during the Song period, carried out reforms called the New Policies (xin fa).
(4) Qin Hui came into conflict with the party in favor of war, concerning the relationship with the Yuan.

… More than 20 modern countries were part of the realm of the Ottoman Empire during its heyday, …

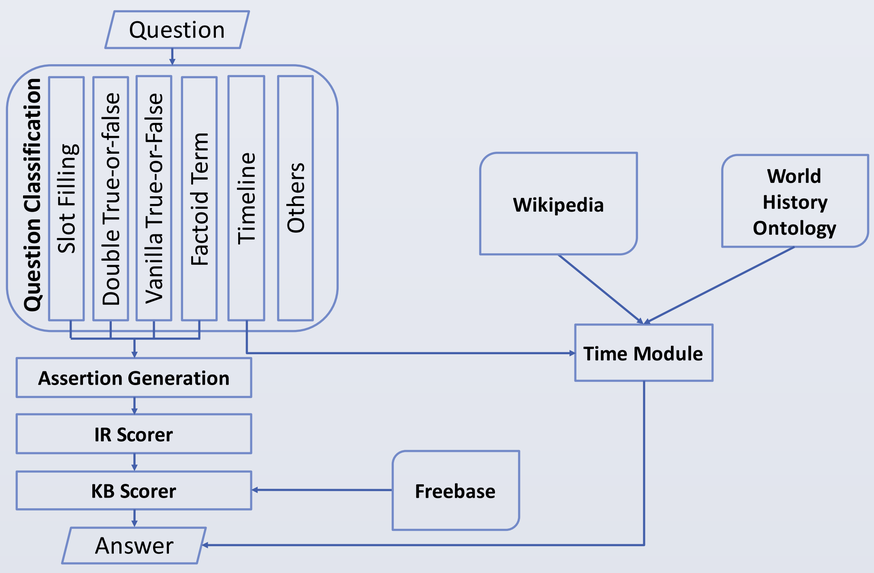
Choose the option that is incorrect as the name of a country among those referred to in the underlined portion.

(1) Cyprus
(2) Hungary
(3) Bulgaria
(4) Czech Republic

PROBLEM DEFINITION: Given a multi-choice question, output the correct option. The questions include: true-false question, slot-filling question, factoid question, timeline question and etc.

CHALLENGE: Apart from IR approach, how knowledge base can help with this task?

Framework



Question Classification

- Slot Filling
  - There is at least one blank mention in the question instruction
- Double True-or-False
  - Problem Definition: Given 2 sentences, judge the correctness of them
  - There is a list of 2 sentence items in the question data section of the xml data file
- Timeline
  - Problem Definition: Sort several events according to their time
  - There is a list containing more than 2 items in the question data section
- Vanilla True-or-False
  - Problem Definition: Given 4 sentences, pick the true (or false) one
  - Questions don't belong to the 3 types above and has at least 5 words in each option
- Factoid
  - Questions don't belong to the 4 types above

Assertion Generation

- Slot Filling
  - Fill the blank. The filled sentence is the assertion
- Double True-False
  - Each sentence forms an assertion
- Timeline
  - Treated specially by timeline question solver
- Vanilla True or False
  - Each sentence forms an assertion
- Factoid
  - Concatenate the question instruction and the option
  - If the instruction refers to an underlined sentence, we also concatenate the underlined sentence

Information Retrieval (IR) solver

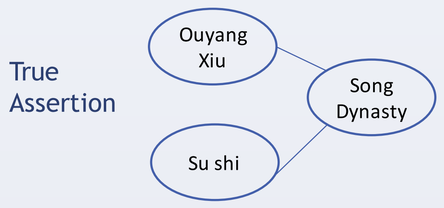
- Corpus: all Wikipedia articles
- Query: assertion
- Use Lucene for sentence level index
- Return the score of the highest ranked sentence

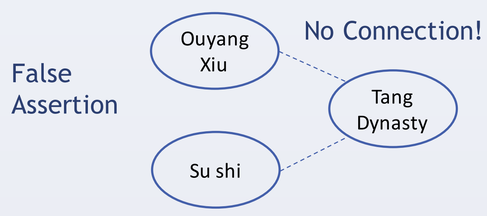
Knowledge Base Solver

INTUITION

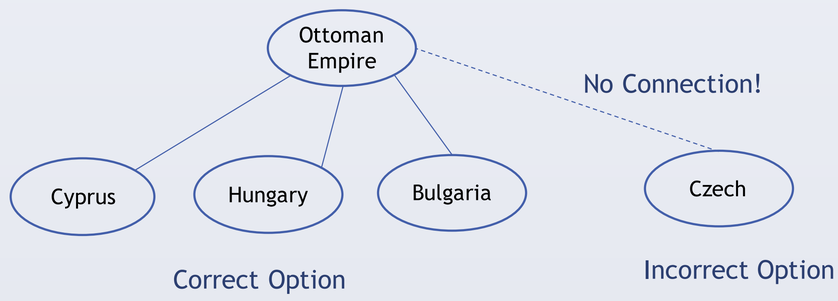
- Entities within a true assertion is more related to each other than a false assertion
- Entities in a true option is more related to the entities in the question instruction and relevant underlined sentences of the question
- The connectivity of entities in knowledge base can reflect relatedness
  - Connectivity: number of edges divided by number of entities

Assertion Connectivity: See the 1st example in the introduction





Question-Option Connectivity: See the 2nd example in the introduction



METHOD

- GDBT to incorporate IR score and knowledge base features
- Learning to rank
  - Rank each option pair
- KB features:
  - Assertion connectivity
  - Connectivity between question side (question instruction + underlined background) and option side
  - Number of entities in question, option, and assertion
  - Length of option and assertion
  - Features indicating the type of the question

Timeline Question Solver

- Detect entities and events in each item
  - String matching and Illinois Wikifier
- Use the start time of the entities and events to rank the items
  - Start time comes from World History Ontology and Wikipedia

Experiments

EXPERIMENTAL SETUP

- Data
  - 1997, 2001, 2003, 2005, 2007, 2009 National Center Test dataset for training
  - Cross validation
  - 2011 National Center Test dataset for test

EXPERIMENTAL RESULTS

Method	Cross Validation Precision
Gradient Boost Decision Tree (GBDT)	**51.0**
Pure Information Retrieval (IR)	49.5

- 34 points (precision 33.3%) in *2011 test dataset*
- Our timeline module can answer 2 out of 3 timeline questions in training set.
  - Our method can't detect any entities or events in some items of the failed one

Conclusions

- Entity connectivity in knowledge base is a useful feature
- Using the start time of entities and events is sufficient to generate timeline for some questions