NUL System at QA Lab-2 Task

NTCIR-12 Conference

Mio Kobayashi, Hiroshi Miyashita,
Ai Ishii, Chikara Hoshino
Technology Research & Innovation
Nihon Unisys, Ltd.
2016/6/9

Foresight in sight
Agenda

1. Introduction
2. System Architecture
3. Experimental Results & Discussion
4. Future Efforts & Conclusion
Introduction
Our past challenge

- NTCIR-11 QA-Lab
  - Convert to Textual entailment recognition (RTE)
  - Shallow approach
  - Apache Solr
    → 33 points (out of 100)
Previous studies

- Solve questions by keyword distributions
  - Kano (2014)
- Convert questions into factoid questions
- Convert questions into textual inference
  - Tian (2014)
Observation of the task

- True-or-False : 70% of questions
  - Not necessary to read knowledge resources widely
    - 99.3% of questions have a paragraph which contain all of named entities in correct choice
  - Time descriptions are more abstract
  - Location descriptions are more abstract
2 System Architecture
Knowledge Resources

- 4 sets of high school textbooks
- Wikipedia
- World history ontology (Kawazoe 2014)
- Web sites of world history
Dictionaries

- Named entity dictionary
  - Class of words
  - Approximately 45,000 entries

- Synonym dictionary
  - Wikipedia Redirect
  - Wikipedia Hyponymy extraction tool

- Hypernym-hyponym dictionary
  - Wikipedia Hyponymy extraction tool

- Antonym dictionary
  - World history ontology (Kawazoe 2014)

- Suffix dictionary

- Year conversion dictionary
  - Years of nations and historical events
Common Modules

- **Time expressions**
  - “Charlemagne defeated the Magyar at the 8th century.”
  - 794 ∈ 8th century

- **Matching of words**
  - Synonym, hypernym
  - Ignore suffixes
  - Exclusive relation (have same class & not matched)

- **Extracting named entities from questions**
  - “Charlemagne defeated the Magyar at the 8th century.”
Search engine settings

- Apache Solr
  - Index: paragraph / sentence
  - Weighting: Okapi-BM25
  - Morphological dictionary: UniDic (Den, 2008)
System Architecture

Exam data

Question analyzer

Solver 1
- PMI Score
- Search Rank

Solver 2
- Factoid Score

Solver 3
- Tree Match

Solver 4
- Factoid Score
- Ontology Search

Aggregate

Answer
<table>
<thead>
<tr>
<th>Type of questions</th>
<th>So.1</th>
<th>So.2</th>
<th>So.3</th>
<th>So.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative True-or-False Questions</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Relative True-or-False in Focus Word Questions</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Absolute True-or-False Questions</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Factoid Questions</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Slot-Filling Questions</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Time Reordering Questions</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>What-Time Questions</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Unique Image, Mixed and Other Questions</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Aggregate – weighted Borda count

Question: What is the name of banknote of the Yuan dynasty?
(元が発行した紙幣の名称として最も適当なものを探す)

Solver 1 (weight:0.6)
1st: [4], 2nd: [1], 3rd: [2], 4th: [3]

Solver 2 (weight:0.5)
1st: [4], 2nd: [2], 3rd: [1], 4th: [3]

Solver 3 (weight:0.4)
1st: [3], 2nd: [2], 3rd: [1], 4th: [4]

Solver 4 (weight:0.3)
1st: [1], 2nd: [3], 3rd: [4], 4th: [2]

[4] 5.4pt
[1] 4.8pt
[2] 4.2pt
[3] 3.6pt

“Answer is [4]”
Solver 1 – Score of PMI

- Pointwise Mutual Information (PMI) values
  
  - $S = \text{pairs}; \text{NE and next NE or CW (which has antonym)}$
  
  \[ PMIScore = \frac{1}{|S|} \sum_{(w_1, w_2) \in S} \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \]

- In correct choice, each word has relationship each other

17世紀後半、ロシアではピョートル1世が即位し、西欧化政策を開始した。

Late 17 century Russia Peter the Great enthronement Westernized policy start
Solver 1 – Score of Search Rank

**Search Rank**

- Calculate score about combinations of 3 or more words

\[
RankScore = \frac{1}{|Q|} \sum_{(ne_i,q_i) \in Q} -Rank(ne_i,q_i)
\]

\[
Rank(ne_i,q_i) = \begin{cases} 
rank_{ne_i,q_i} & (rank_{ne_i,q_i} < k) \\
2 \times k & (otherwise)
\end{cases}
\]

- Omit 1 NE and search the sentence
- Check the rank which contain omitted NE

17世紀後半、[    ]では ピョートル1世が 即位 し、西欧化政策を 開始 した。
Late 17 century [____] Peter the Great enthronement Westernized policy start
Solver 1 – Score of Time expression

- Time expression

\[ TimeScore = 10.0 - 20.0 \times \frac{\text{rank}(W_{time})}{\text{length}(\text{list}_{time})} \]

- Omit a time expression and search the sentence

- Check the search result and add some rules such as:
  - Time Expression is not found: -10.0
  - Time Expression and all NE are found in result: +10.0
  - All NE are found but it contains other Time Expression: -100.0
## Calculate total score of a choice

17世紀後半、イギリスのエリザベス1世が統一法を発布した。

<table>
<thead>
<tr>
<th>Time</th>
<th>PMIScore</th>
<th>RankScore</th>
<th>TimeScore</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.01</td>
<td>1.0</td>
<td>-10.0</td>
<td>2.79</td>
</tr>
<tr>
<td>2</td>
<td>1.56</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**total score:** -8.21
Convert to virtual factoid question

Charlemagne defeated the Magyar at the 8th century.
(8世紀にカール大帝はマジャール人を撃退した。)

Charlemagne defeated the Magyar at the (Time).
((時代)にカール大帝はマジャール人を撃退した。)

(Person) defeated the Magyar at the 8th century.
(8世紀に(人物)はマジャール人を撃退した。)

Charlemagne defeated the (Nationality) at the 8th century.
(8世紀にカール大帝は(民族)を撃退した。)
Solver 2 - Answering factoid questions

\[ QAScore(Q, d_k) = \sum_i \alpha_{q_i, a_k} \exp\{-\gamma l(q_i, d_k)^{\beta}\} \]

- \( Q \): virtual factoid question
  - \( q_i \): \( i \)th word of \( Q \)
  - \( d_k \): \( k \)th word of searched resource
- \( l(q_i, d_k) \): the nearest neighbor distance from \( d_j \) to \( d_k \)
  - \( d_j \): in synonym of \( q_i \), nearest word from \( d_k \)
Score about named entity “Avars”:

\[ QAScore(Q, Avars) = \alpha_{Charlemagne,Avars} \exp\{-\gamma l(Charlemagne, Avars)^\beta\} \]

There are 22 steps between “Avars” and “Charlemagne” which is nearest from “Avars”

\[ l(Charlemagne, Avars) = 22 \]
Solver 2 – cost of factoid questions

- **Calculate costs of each choice**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Word</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Avar</td>
<td>3.2</td>
</tr>
<tr>
<td>2</td>
<td>Mongolian</td>
<td>2.6</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Magyar</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Cost: $3.2 - 1.1 = 2.1$

- **Cost of a choice** = A mean of the each cost of virtual question
- **Answer** = Choice which has lowest cost
Convert to syntax tree

(クシャーナ朝はパータリプトラに都をおいた。)

pred establish

sbj Kushan Empire

obj capital

loc Pataliputra

hypothesis
An abstract expression for a choice is similar to an abstract expression in knowledge resources → the choice is correct.
The Kushan King Kanishka I moved capital to Peshawar.

クシャーナ朝の王、カニシカ1世は都をペシャワールに移した。
Matching of Syntax Tree

Choice

establish

Kushan Empire

Kushan

Kanishka I

capital

capital

Pataliputra

Peshawar

Mid-2nd Century

Resources

move

match

Matched

matched

Exclusive

Not matched

Same class & not synonym/hypernym

→ Pataliputra ∩ Peshawar = Φ
Convert to virtual factoid question (similar to solver 2)

Charlemagne defeated the Magyar at the 8th century.
(8世紀にカール大帝はマジャール人を撃退した。)

Charlemagne defeated the Magyar at the (Time).
(時代)にカール大帝はマジャール人を撃退した。)

(Person) defeated the Magyar at the 8th century.
(8世紀に(人物)はマジャール人を撃退した。)

Charlemagne defeated the (Nationality) at the 8th century.
(8世紀にカール大帝は(民族)を撃退した。)
Solve factoid from co-occurrence

Charlemagne defeated the Magyar at the (Time).
((時代)にカール大帝はマジャール人を撃退した。)

List of co-occur words

- **Charlemagne**
  - Franks, 768, 774, Itary, Carolingian Empire, Pepin the Short, Leo III, 814, Carolingian Renaissance, …
- **The Magyar**
  - Hungary, Ural Mountains, 862, 837, 5th centuries BC, Khazar Empire, …

Type filter

- 768, 774, 800, 814, Middle Ages, 9th century, …
- 862, 837, 5th centuries BC, 9th century, …

Answer

9th century
3 Experimental Results & Discussion
### Experimental Results

#### NUL scores of formal run

<table>
<thead>
<tr>
<th>Ph.</th>
<th>Exams</th>
<th>Run #1</th>
<th>Run #2</th>
<th>Run #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>National Center Test (1999)</td>
<td>43</td>
<td>49</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>Benesse mock exam (2015 Jun/All/out of 175)</td>
<td>121</td>
<td>121</td>
<td>118</td>
</tr>
<tr>
<td>(2)</td>
<td>Benesse mock exam (2015 Jun/Pattern 1)</td>
<td>76</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>(2)</td>
<td>Benesse mock exam (2015 Jun/Pattern 2)</td>
<td>64</td>
<td>64</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>National Center Test (2011)</td>
<td>65</td>
<td>65</td>
<td><strong>68</strong></td>
</tr>
<tr>
<td>3</td>
<td>Benesse mock exam (2014 Sep/All/out of 125)</td>
<td><strong>77</strong></td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>(3)</td>
<td>Benesse mock exam (2014 Sep/Pattern 1)</td>
<td>60</td>
<td>57</td>
<td>60</td>
</tr>
<tr>
<td>(3)</td>
<td>Benesse mock exam (2014 Sep/Pattern 2)</td>
<td>58</td>
<td>60</td>
<td>54</td>
</tr>
</tbody>
</table>
## Experimental Results

### Scores by each solvers of phase 3

<table>
<thead>
<tr>
<th>Ph.</th>
<th>Exams</th>
<th>So.1</th>
<th>So.2</th>
<th>So.3</th>
<th>So.4</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>National Center Test (1999)</td>
<td>46</td>
<td>56</td>
<td>30</td>
<td>40</td>
<td>52 / 100</td>
</tr>
<tr>
<td>2</td>
<td>Benesse mock exam (2015 Jun)</td>
<td>121</td>
<td>104</td>
<td>82</td>
<td>56</td>
<td>125 / 175</td>
</tr>
<tr>
<td>3</td>
<td>National Center Test (2011)</td>
<td>62</td>
<td>62</td>
<td>43</td>
<td>39</td>
<td>65 / 100</td>
</tr>
<tr>
<td>3</td>
<td>Benesse mock exam (2014 Sep)</td>
<td>58</td>
<td>65</td>
<td>36</td>
<td>33</td>
<td>76 / 125</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>287</td>
<td>287</td>
<td>191</td>
<td>168</td>
<td>318 / 500</td>
</tr>
</tbody>
</table>
Experimental Results

Numbers of correct (each solvers / type of questions)

<table>
<thead>
<tr>
<th>Type of questions</th>
<th>So.1</th>
<th>So.2</th>
<th>So.3</th>
<th>So.4</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative True-or-False Questions</td>
<td>53</td>
<td>54</td>
<td>48</td>
<td>32</td>
<td>56 / 75</td>
</tr>
<tr>
<td>Relative True-or-False in Focus Word Questions</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2 / 4</td>
</tr>
<tr>
<td>Absolute True-or-False Questions</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>-</td>
<td>9 / 17</td>
</tr>
<tr>
<td>Factoid Questions</td>
<td>8</td>
<td>7</td>
<td>-</td>
<td>4</td>
<td>6 / 10</td>
</tr>
<tr>
<td>Slot-Filling Questions</td>
<td>7</td>
<td>10</td>
<td>-</td>
<td>4</td>
<td>9 / 13</td>
</tr>
<tr>
<td>Time Reordering Questions</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>6 / 9</td>
</tr>
<tr>
<td>What-Time Questions</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0 / 3</td>
</tr>
<tr>
<td>Unique Image, Mixed and Other Questions</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7 / 13</td>
</tr>
</tbody>
</table>

By National Center Test(2011) and Benesse Mock Exam(2014 Sep, 2015 Jun)
Error Analysis

Patterns of incorrect

- Misjudgment of words dependency (solver 1)
- Confused by many-to-many relationship (solver 2 and 4)
- Error because of homonyms (solver 3)
- Failed to anaphora resolution (solver 3)
- Cannot match Syntax Trees by rephrasing (solver 3)
- Too frequent named entities (All solvers)
- Failed to extract necessary named entity from instruction (All solvers)
Conclusion & Future Works
## Conclusion & Future Works

### Conclusion
- Observe the task
  - Make 4 strategies from observation
- Combination strategy
  - To reduce variance in generalization error

→ Get 1st place (JA)

### Future Works
- Strict analysis of sentence
- Determination of important words
- Optimal strategy of solver combination
- Question analysis