NUL System at QA Lab-2 Task

Mio Kobayashi, Hiroshi Miyashita, Ai Ishii and Chikara Hoshino
Nihon Unisys, Ltd.

Introduction

In this poster, we describe our solvers for "Relative TF" type question which are key questions in this contest.

Previous Studies

Previous studies have different compromising characteristics. So, we assumed that if we can combine them properly, we can get better result.

Methods	Characteristic
Use distribution of words. [Kano 14]	Coverage↑Exactness↓
Convert to QA. [Kanayama 13, Okita 14]	Some questions can not be converted.
Use abstract expression. [Tian 14]	Coverage↓Exactness↑

Observations

We examined questions from following view point
All of NEs in true sentence exists in a single paragraph?
Time/Location expressions in the choice are wider than knowledge base's one?

Locality of information

Entailment for Time/Loc expression

Item	Result
NEs in a single paragraph	99.3%
Wider location expression	11.9%
Wider time expression	28.9%

Sufficient only Locally information. But need recognizing entailment for Time/Loc expressions.

Number of questions: 135
Knowledge resource: Textbook(4sets), Wikipedia, World History ontology, Chronological table

Dictionaries & Modules

We made following dictionaries and modules.

Dictionaries/Modules	Explanation
Named Entity	With types like "person", "nation".
Synonym	Synonyms.
Hypernym, Hyponym	Hypernyms, Hyponyms. This dictionary include inclusion information about location.
Suffix	Suffixes like "人(-ese)", "系(-ian)".
Event to year converter	Historical event to time.
Recognize entailment for time expression	Module for judging time inclusion relations.

Examples

Choice	Knowledge	Judge	Explanation
ブッダ	仏陀	Match	From synonym dictionary
Crusades	4th Crusades	Match	"Crusades" is hypernym of "4th Crusades".
France	Paris	Match	France ⊇ Paris(Geometrically inclusion)
Iranian	Iran	Match	Stem is same.
6th Century	AD 515	Match	6th Century ⊇ AD 515
America	England	Exclude Match	Both words are nation type, but no synonym or hypernym relations are exists.

Use distribution of words

Hypothesis

If True sentence, co-occurrence of two words is high, and three or more words relations are strong.

PMIScore/TimeScore/RankScore

- PMIScore to evaluate co-occurrence of two words
- TimeScore, RankScore for considering the relation between one NE and the other words

Late 17th century (TIME), Elizabeth I (NE) issued (CW) Act of Uniformity (NE) in United Kingdom (NE) .

Pair:(Elizabeth I , United Kingdom), (Elizabeth I , Act of Uniformity)
TIME:Late 17th century

PMIScore

We define the following Score.

$$PMIScore = \frac{1}{|S|}\sum_{(w_1,w_2)\in S}\log\frac{P(w_1,w_2)}{P(w_1)P(w_2)}$$

where, S is all pair, $|S|$ is number of S, $P(w_1)$ is $hits(w_1)/N$, $P(w_1,w_2)$ is $hits(w_1\ AND\ w_2)/N$, $hits(w_1)$ is the number of sentences retrieved when the query "w_1" with synonyms is given to Solr, N is total number of sentence units in search index.

TimeScore

We get list of time expressions list up to 20 from top 30 search results by the choice excluded the $time_i$ (Late 17th century) as the query.
$rank(w_{time})$ is $time_i$'s rank in the time expression list, and $length(W_{time})$ is the length of the list. TimeScore is defined as:

$$TimeScore = 10.0 - 20.0 * \frac{rank(w_{time})}{length(list_{W_{time}})}$$

RankScore

We make pair of an NE ne_i and a query q_i excluded ne_i from the choice, and define the following RankScore.

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

Where,

$$Rank(ne_i,q_i) = \begin{cases} rank_{ne_i,q_i} & (rank_{ne_i,q_i} < k) \\ 2*k & (otherwise) \end{cases}$$

where, $rank_{ne_i,q_i}$ is the rank of search results by q_i that include ne_i. k is hyper-paramerter, and we set $k = 30$.

We sum up PMIScore, RankScore and TimeScore as the score for Judgment True, False.

Convert to Factoid

Hypothesis

We convert to the sentence to the factoid question. If True sentence, we expected hidden word, otherwise not.

QAScore/Cost/AveCost

We hide Named entity, and make a virtual factoid question, then we calculated the cost to the returned answer.

Charlemagne (person) attacked Magyar (personType).

Q1: (person) attacked Magyar.
Q2: Charlemagne attacked (personType).

QAScore

We define the following QAScore.

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

Where, Q is the question which hides target NE, q_i is ith word and d_k is kth word of the query results.
Moreover, $l(q_i,d_k)$ is the minimum distance which the word d_k match q_i against the synonym set in the query results. $\alpha_{q_i,d_k}, \gamma, \beta$ are hyper parameters.

Q2: Charlemagne attack (personType).
The score of Avars $QAScore(q,d_k)$ is
$QAScore(Q2,\ Avars) = \alpha_{charlemagne,Avars} exp\{-\gamma l(Charlemagne,\ Avars)^\beta\}$
Query results
''' Avars (personType) was attacked, ''', Charlemagne given imperial crown '''
22 Step
Nearest distance Avars from Charlemagne is 22 steps in the search results, then
$l(Charlemagne,\ Avars) = 22$

Cost/AveCost

We define the cost of the question Q which hide the word a_Q .
$Cost(Q) = \max_k\{Class(a_Q,d_k)QAScore(q,d_k)\} - QAScore(q,a_q)$
Where, $Class(a_Q,d_k)$ is the indicator function, which is 1 when a_Q has same class of d_k , otherwise 0.
Additionally, the total cost of all the virtual question $AveCost$ is defined by the mean of the scores of each virtual questions.

$$AveCost = \frac{1}{|Z|}\sum_{Q\in Z} Cost(Q)$$

We judge True, False by the AveCost.

Use abstract expression

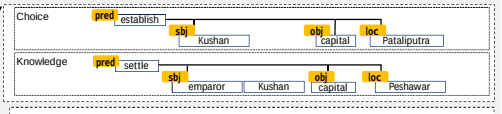
Hypothesis

A similar abstract expression of true sentence exists in the knowledge base, but false sentence's one does not.

TreeSimScore/WMScore/WEMScore

First, we convert choice sentences into pred-arg trees. Then we calculate similarity score against sentences in the knowledge base.

Choice : Kushan established capital in Pataliputra.
Knowledge : ··· emperor of Kushan settled capital in Peshawar ···



SimScore

We define SimScore(Tree Similarity Score) as below. Here, T_h is the choice sentence and T_t is the sentence in knowledge.

$$f_T(T_h,T_t) = f_m\left(T_h^{pred},T_t^{pred}\right) * \max_{r'\in R'} f_m\left(T_h^{r},T_t^{r'}\right) * \frac{1}{|R|}\sum_{r\in R} f_m(T_h^r,T_t^r)$$

T_h^r is set of role r's words, $f_m(A,B)$ is a function which returns 1 if one of word in A matches any of words in B.
$R' = \{sbj, obj\}$, $R = \{sbj, obj, time, loc, loc-to\}$

WMScore

We define WMScore (Word Match Score) as below.

$$f_W(T_h,T_t) = Boost(T_h,T_t) * \frac{1}{|W_h|}\sum_{w_h\in W_h}\max_{w_t\in W_t} f_w(w_h,w_t)$$

Here, W_h is words set in T_h, $f_w(w_h,w_t)$ is a function which returns 1.0 if w_h matches w_t, $Boost(T_h,T_t)$ is a function which returns 2.0 if SimScore exceeds 0.5, otherwise 1.0. This score considers not only match rate of words, but also SimScore.

WEMScore

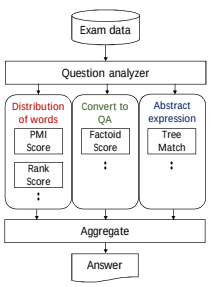
We define WEMScore (Word Exclude Match Score) as below.

$$f_{-W}(T_h,T_t) = \max_{w_h\in W_h, w_t'\in W_t'} f_e(w_h,w_t')$$

Here, W_h' is a word which does not match any of words in T_t, $f_e(w_h',w_t')$ is a function which returns 1.0 if w_h' holds exclusive relation against w_t'. We detect wrong word.

We judge True, False by the WMScore and WEMScore.

Aggregation



Formal Run Results : We got 1st place in phase 2, 3 !!

Phase	Exam	Run1	Run2	Run3
1	National Center Test (1999)	43	46	36
2	Benesse mock exam (2015 Jun/All/out of 175)	121	121	118
3	National Center Test (2011)	65	65	68
3	Benesse mock exam (2014 Sep/All/out of 125)	77	76	76

Discussion/Error Analysis

Main causes

Cause of error
Too frequent NE.
Fail to match synonym.
Fail to extract important words from the instruction.
Lack of recognizing causal relation.
NE is not contained in the dictionary.
Fail to recognize replacement of sbj and obj.
Fail to extract time expression.