ICL Participation at NTCIR-9 RITE[i]

XU Xing, WANG Houfeng
Key Laboratory of Computational linguistics (Peking University), Ministry of Education, Beijing 100871
E-mail: xuxing@pku.edu.cn

Abstract

This paper describes ICL's participation at NTCIR-9 RITE. We chose BC & MC subtask. Textual entailment is a problem to predict whether an entailment holds for a given test-hypothesis pair. We built an inference model to solve this problem by means of using dependency syntax analysis (by LTP), lexical knowledge base (e.g. CCD), web information (e.g. Baidupedia) and probability method. We used AUC indicator to evaluate the ranking ability of our system.

Keywords: textual entailment; lexical entailment probability; dependency syntax analysis; Baidupedia; CCD (Chinese Wordnet)

1. Introduction

Textual Entailment is that a task to detect the entail relationship between a text pair, it can be applied to Q&A system, Intelligent search, etc. In NTCIR-9 RITE evaluating task, we need to know both the entail relationship (BC subtask) and the entail direction (MC subtask) for the given text pair. Texts are written in Japanese, Simplified Chinese, or Traditional Chinese, our team chose Simplified Chinese, BC&MC subtask.

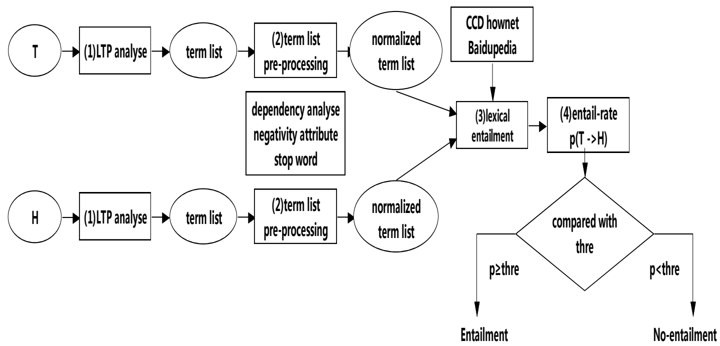

Figure1:System architecture

2. Pre-processing

LTP: LTP is a parser developed by Harbin Institute of Technology (HIT). The functions of LTP include Word Segmentation, POS tagging, Syntactic Analysis and Named Entity Recognition, etc. We get a word list after parsing a Chinese sentence. We merge the words in the same named entity to one word. For each word in the list, we have its lexical information (token, POS), syntactic information (father and children in the dependency syntax tree) and semantic information (named entity, semantic role).

More Lexical information:

(1) **Negativity Attribute**: If a word is

linked by a negative words (like “没有”, “不”) in the syntax tree, then its negativity attribute is true, else is false.

(2) **Stop Word**: Stop Word stands for the word doesn’t have real meaning, include punctuation, negative words and function word. All stop words will not participate in the calculation of the sentence entail-rate (probability). We use the Stop attribute to identify if a word is a stop word.

3. Main Module

When given a pair of two sentences t1 and t2, we must calculate the entail-rate of both $p(t1 \rightarrow t2)$ and $p(t2 \rightarrow t1)$. When we are calculating $p(t1 \rightarrow t2)$, we regard t1 as the text T, and t2 as the hypothesis H, the method to get $p(T \rightarrow H)$ is presented below, and $p(t2 \rightarrow t1)$ can be correspondingly got.

(1)word pair entail-rate $p(T_w \rightarrow H_w)$ ***:***

We choose a word T_w from the text and a word H_w from the hypothesis, we have the following rules to get the entail-rate $p(T_w \rightarrow H_w)$:

First, Direct Rule: $p_{identify}(T_w \rightarrow H_w)=1$ if H_w is the same as T_w or a part of T_w

Second, CCD(Chinese Wordnet) based Rule: $p_{\mathrm{CCD}}(T_w \rightarrow H_w)=1$ if H_w is in the synset, hypernym set, hyponym set, holonym set, meronym set or attribute set of T_w in CCD. And $p_{\mathrm{CCD}}(T_w \rightarrow H_w)$ = -1 if H_w is in the antonym set of T_w.

Third, Baidupedia based Rule: $p_{web}(T_w \rightarrow H_w)=1$ if H_w can be extracted from T_w’s Baidupedia card. Three are many ways to extracted related word pairs from the web, we can choose the words that are linked with the source-word by be-verb(in Chinese, like “是”, “表示”, “指”,etc.), the hyper-link text and the text it links to, or the word in the parenthesis(as supplement explain for the former word) and the former word.

For example, in the training data:

<pair id="24" label="F">

<t1>黄金周期间，国美、苏宁等企业创造了较好的销售业绩</t1>

<t2>黄金周期间，一些家电连锁企业创造了较好的销售业绩</t2>

</pair>

The word “国美” can’t be find in CCD and other knowledge base, but we can use its Baidupedia to find the entail relation between “家电”, “连锁”, the Baidupedia card of “国美” is:

国美电器（英语：GOME）是中国的一家连锁型家电销售企业，也是中国大陆最大的家电零售连锁企业，2009 年，国美电器入选中国世界纪录协会中国最大的家电零售连锁企业。成立于 1987 年 1 月 1 日。董事局主席张大中。在北京、太原、天津、上海、广州、深圳、青岛、长沙、香港等城市设立了 42 个分公司，及 1200 多家直营店面。

And we can find the word “家电” and “连锁” in the Baidupedia card and is linked with the be-verb “是”, so we can get

$$P_{web}(\text{国美} \rightarrow \text{家电})=1,\ P_{web}(\text{国美} \rightarrow \text{连锁})=1$$

.

We use the three rules in order and get the final entail-rate $p(T_w \rightarrow H_w)$.

(2)the entail-rate of one word $p(H_w)$ ***:***

First, get the word from text that entails H_w ***:***

H_w is a word in the hypothesis, all the words in the text is the candidate to entail this word, we use the word that have the maximal absolute value of entail-rate, named $T_{w-entail}$,

$$T_{w-entail} = \arg \operatorname*{maxabs}_{T_w \in T} p(T_w \to H_w)$$

Second, take into account the Negativity Attribute:

$$p(H_w) = P(T_{w-entail} \to H_w) * neg(T_{w-entail}) * neg(H_w)$$

Where $neg(w)$ is -1 is the Negativity Attribute is false and 1, otherwise.

(3)hypothesis entail-rate $p(H)$ ***:***

For each non-stop word in hypothesis, we get:

$$\begin{cases} polarity(H_w) = \operatorname{sgn}(p(H_w)) \\ simlarity(H_w) = |p(H_w)| \end{cases}$$

The we calculate the final entail-rate of $P(T \to H)$,

$$polarity(H) = \prod_{H_w \in H} polarity(H_w)$$

$$simlarity(H) = \frac{1}{n} \sum_{H_w \in H} simlarity(H_w),$$

n is the number of non-stop word in H

$$p(T \to H) = polarity(H) * similarity(H)$$

(4)from $P(T \to H)$ ***to the entail-type:***

Binary-class:

$$result = \begin{cases} p(t1 \to t2) \ge thre \parallel p(t2 \to t1) \ge thre & \text{Y} \\ otherwise & \text{N} \end{cases}$$

Multiple-class:

$$result = \begin{cases} p(t1 \to t2) \ge thre \,\&\&\, p(t2 \to t1) \ge thre & \text{B} \\ p(t1 \to t2) \ge thre \,\&\&\, p(t2 \to t1) < thre & \text{F} \\ p(t1 \to t2) < thre \,\&\&\, p(t2 \to t1) \ge thre & \text{R} \\ p(t1 \to t2) < -thre \,\&\&\, p(t2 \to t1) < -thre & \text{C} \\ otherwise & \text{I} \end{cases}$$

The threshold can be set manually. In the training data, precision get maximal value when thre=0.58(BC) and thre=0.72(MC), we use this value to label the test data.

4. Results

The distributions of our results for BC subtask is presented below:

Table 1: Results in NTCIR-9 RITE on BC subtask on test data

type	Precision	Recall	F-value
Yes	0.7106	0.9430	0.8105
No	0.7414	0.2986	0.4257
Total	0.7150		
AUC	0.6459		

AUC (Area Under ROC Curve) is a indicator to the ranking ability of a system, we can use it to charge our textual-entailment system because it is independent of the threshold. AUC can be calculated only for 2-way task.

The distributions of our results for MC subtask is presented below:

type	Precision	Recall	F-value
F	0.7759	0.4455	0.5660
R	0.6892	0.5604	0.6182
B	0.3351	0.8873	0.4865
C	0.3077	0.0541	0.0920
I	0.5811	0.6143	0.5972
Total	0.5061		

Table 2: Results in NTCIR-9 RITE on MC subtask on test data

As seen, the precision and recall is very low because the polarity formula:

$$polarity(H) = \prod_{H_w \in H} polarity(H_w) \quad \text{is}$$

too simple. And many contradict cases are that two words can't be true together, but they are not necessary to be antonyms. So, our system is hard to recognize contradict cases.

Ablation tests:

We do ablation tests to know if the impact of using web information:

system	BC	MC
Baseline	0.7150 AUC=0.6460	0.5061
Baseline+ Baidupedia	0.7076 AUC=0.6393	0.4914

Table 3: Results in NTCIR-9 RITE on MC subtask on test data

The system with Baidupedia has lower precision. Perhaps it is because we haven't use exact rules to extract the useful information from the Baidupedia cards, so it brings too many noise and can't improve the result.

5. Conclusion and Future work

The paper presents the architecture of the system used in NTCIR-9 RITE. We used CCD (Chinese Wordnet) to get lexical entailment, most words can be found in CCD but some Named Entities and OOV can't be found, so we introduced Baidupedia to improve the entailment of these words. But three are too many noises in the web, we will find more exact rules to extract the helpful information from the web in the future. Our system didn't work well in recognize contradict cases, we will improve the algorithm from lexical entailment to sentence entailment.

Reference

Hideki Shima, Hiroshi Kanayama, Cheng-Wei Lee, Chuan-Jie Lin, Teruko Mitamura, Yusuke Miyao, Shuming Shi, Koichi Takeda. *Overview of NTCIR-9 RITE: Recognizing Inference in Text* 2011

Glickman, Oren, Ido Dagan and Moshe Koppel. *A Probabilistic Lexical Approach to Textual Entailment[C].* In Proceedings of the Nineteenth Inter national Joint Conference on Artificial Intelligence. 2005:1682-1683.

Adrian Iftene. *TEXTUAL ENTAILMENT[D].* Iasi, Romania: "Al. I. Cuza" University, 2009.

A. Iftene and M.A. Moruz. 2009. *UAIC participation at rte5*.In TAC 2009 Notebook Papers.

Eyal Shnarch. *Lexical entailment and its extraction from wikipedia. Master thesis,* Computer Science Department, Bar-Ilan University, Israel, 2008.

[i] This project is supported by NSFC, Fund No: 90920011, 91024009