IASL RITE System at NTCIR-10

Cheng-Wei Shih
Institute of Information Science, Academia Sinica
National TsingHua University, Taiwan, R.O.C
dapi@iis.sinica.edu.tw

Chad Liu
Institute of Information Science, Academia Sinica, Taiwan, R.O.C
Chadliu23@iis.sinica.edu.tw

Cheng-Wei Lee
Institute of Information Science, Academia Sinica, Taiwan, R.O.C
aska@iis.sinica.edu.tw

Wen-Lian Hsu
Institute of Information Science, Academia Sinica, Taiwan, R.O.C
hsu@iis.sinica.edu.tw

ABSTRACT
At our second participation in NTCIR RITE, we developed a two-stage knowledge-based textual inference recognition system for both BC and MC subtasks in Chinese. Two main recognition systems, which are based on named entities, Chinese tokens, word dependency, and sentence length, were implemented to identify the entailment and contradiction between sentences. The evaluation result showed that our 2-stage system achieved 0.6714 and 0.4632 for traditional Chinese BC and MC subtasks respectively. It greatly surpassed our previous work in NTCIR-9 RITE. In the unofficial run of simplified Chinese, the accuracy of our system also reached 0.6045 in BC and 0.5094 in MC.

Team Name
IASL

Subtask
Traditional Chinese, Simplified Chinese (Unofficial)

Keywords
Knowledge-Based System, Entailment Recognition, Contradiction Detection.

1. INTRODUCTION
Since the First Recognizing Textual Entailment Challenge (RTE-1)[1] hold in 2005, text understanding and inference, which is believed one of the most challenging applications in natural language processing, has become a hot research topic. This year, NTCIR-10 RITE[2] continues the previous RITE task[3] that focused on the research, development, and evaluation of text understanding and inference system in Asian languages. Comparing with previous RITE task, the NTCIR-10 RITE similarly proposed classification of entailment directions in binary classes, multiple classes, and a special RITE4QA tracks with only one relevant change: the dataset no longer contain sentence pairs with "Reverse" entailment direction.

With the experience of developing a knowledge-based RITE system in NTCIR-9, we found that the data over-fitting is not only a serious problem for a statistical-based system. It is also very difficult to construct a knowledge-based RITE system for tackling complicated problem such as multiple-class inference direction recognition without fitting-data. Besides, after analyzing the result of our system in NTCIR-9 RITE, we also found that the ability of identifying contradiction statements still have a huge room to improve. In fact, the contradiction recognition had only reached 0.368 in F-Score, which is much lower than the overall system performance (0.501). These observations motivated us to view RITE problem from another perspective: can we build a robust text understanding and inference system that is capable of recognizing contradictive statements? For doing that we try to transform the original RITE multiple-class recognition task into three binary-classification problems: Does the text entail the hypothesis? Does the hypothesis entail the text? And does the text conflict the hypothesis? Table 1 shows the mapping of the original RITE entailment labels and the transformed three binary decisions.

Table 1. The Mapping of the original RITE entailment direction and the transformed three binary decisions.

T conflict H	T entail H	H entail T	RITE Result
true	true	true	Contradiction
true	true	false	Contradiction
true	false	true	Contradiction
true	false	false	Contradiction
false	true	true	Bidirectional
false	true	false	Forward
false	false	true	Reverse
false	false	false	Independent

According to Table 1, RITE multiple–class classification can be decomposed as two independent steps: Entailment recognition and contradiction detection. We believe it not only can reduce the cost of development and improve the robustness of the system; we can also pay much attention on investigating the characteristics of contradictory statements and the strategies for detecting them.

2. SYSTEM ARCHITECTURE
We based on our NTCIR-9 RITE work to reconstruct a 2-stage textual entailment system. After the preprocessing step, all the three binary decisions (Does the text conflict the hypothesis? Does the text entail the hypothesis? Does the hypothesis entail the text?) are determined via two recognition modules. Figure 1 shows the architecture of our 2-stage RITE system.

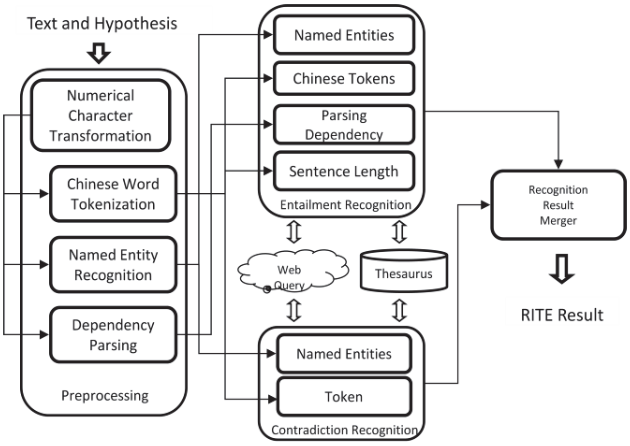

Figure 1. System Architecture

3. PREPROCESSING

The preprocessing which includes numerical character transformation, Chinese word tokenization, named entity recognition, and Chinese dependency parsing provides literal, syntactic, and semantic features for the following contradiction and entailment recognition modules. In this section, the details of these preprocessing steps are introduced.

3.1 Numerical Character Transformation

Fullwidth numerals, symbols, and Chinese numerals which appear in temporal, numerical, and range expressions may cause ambiguities and damage the accuracy of text analysis. By normalizing all these character variations, those ambiguities have a better chance to be resolved. For standardizing the character variations into digit, we collect some frequent Chinese character variations, and design a normalizing algorithm which is based on regular expression matching. For example, a sentence such as *"一九九九年十二月十日"* (December 10, 1999) will be converted into *"1999年12月10日"*. Terms with Chinese numerals such as *"一起"(together), "一共" (total)* will skip the normalization to avoid false transformation.

The other phenomenon worth to mention is the omitting in numerical/temporal expressions such as range and duration. For example, *"一九九九年十月至十二月" (October to December, 1999)* includes a start point and an end point of a period. But the year description of the end point is missing because the end point shares the year description with the start point. In text understanding, it may mislead the language process and cause false recognition. We design an algorithm to detect these description sharing and repair the omitted parts. Statements like *"一九九九年十月至十二月" (October to December, 1999)* will be restored to *"一九九九年十月至一九九九年十二月" (October, 1999 to December, 1999)*

3.2 Chinese Word Tokenization

Chinese word tokenization is a fundamental step of almost all Chinese language applications. By observing the presence of the Chinese tokens, the relation between text and hypothesis might be successfully discovered. In order to acquire tokenized sentences, we use CKIP Chinese segmentation tool [4] to segment all the input sentences and filter out stopwords and function words in the text.

3.3 Named Entity Recognition

It is widely believed that recognizing named entities (NE) is one of the most efficient features to determine the information richness of texts. In textual entailment, the occurrences of NE strongly indicate the inference direction and inference type between text and hypothesis.

Like our previous participation in NTCIR-9, we adopt a hybrid NE recognition system which applied in [5][6] for named entity recognition (NER). NEs such as person name, location, organization, temporal and numerical expressions, and certain kind of artifacts in the texts will be automatically labeled.

3.4 Chinese Dependency Parsing

We believe the use of grammatical dependency between words is beneficial to precisely retrieve useful information from a sentence. By investigating the structure of a parse tree, we can locate the grammatical-related word pairs and understand the syntactic and semantic roles between two connected words. We extract dependency word pairs from the parsing trees which are produced by CKIP Chinese parser [7]. Thereafter, the semantic role along dependency will be assigned on each participating word.

4. ENTAILMENT RECOGNITION

Four recognition modules, which include named entity-based, Chinese token-based, dependency word pair-based, and sentence length-based, are used independently to determine whether a sentence S from a sentence pair (S, S') entail S' or not. Almost all the modules follow the notion that S does not entail S' if S' contains information that cannot find in S. The entailment decisions generated by these recognition modules are integrated by voting. In this section, we describe the features and the notion of these four recognition modules.

4.1 Named Entity based Entailment Recognition

The main objective of this module is to compare named entities in S and S'. By addressing the unmentioned named entities in S', the entailment relation between (S, S') could be determined. First, we give the definition of unmentioned named entity:

Definition 1. Unmentioned Named Entity

Assume a named entity list $\mathrm{NE_S} \subset \mathrm{S}$ and a named entity list $\mathrm{NE_{S'}} \subset \mathrm{S'}$. If there exist a named entity n that $\mathrm{n} \notin \mathrm{NE_S}$ and $\mathrm{n} \in \mathrm{NE_{S'}}$, then n is an unmentioned named entity (UNE).

The rule of judging the entailment relation between S and S' is as follow:

$$\mathrm{Entailment(S, S')} = \begin{Bmatrix} \text{false, if there exist an UNE in (S, S')} \\ \text{true, otherwise} \end{Bmatrix}$$

For correctly identifying the UNEs, several issues need to be considered:

- NE synonym: NE synonym includes alias, abbreviations and person's titles. We not only use dictionaries but try to use web queries to deal with NE synonyms. The idea of using web query for synonym detection is based on the observation that named entities and their synonyms tend to appear closely. For example, *梵諦岡教廷(Vatican city/Roman Curia)*
- Temporal Expressions: a precise temporal expression can entail a less-precise temporal expression even though they are literally different. For example, *"1997 年 12 月"(December, 1997)* can entail *"1997 年" (1997)*
- The Unit of Numerical Expressions: some unit of numerical expression may literally different but conceptually equal. For example, *"10 公尺"(10 meters)* is the same as *"10 米" (10 meters)*

4.2 Chinese Token Based Entailment Recognition

Similar to the previous NE recognition module, Chinese token based entailment recognition try to locate exclusive tokens in (S, S'). The definition of exclusive tokens is as follow:

Definition 2. Exclusive Token

Assume a Chinese token list $T_S \subseteq S$ and another Chinese token list $T_{S'} \subseteq S'$. If there exist a Chinese token t that $t \notin T_S$ and $t \in T_{S'}$ then t is an exclusive token (ET).

The rule of judging if S entail S' is as follow:

$$\mathrm{Entailment}(S, S') = \begin{Bmatrix} \text{false, if there exist an ET in } (S, S') \\ \text{true, otherwise} \end{Bmatrix}$$

The most important issue in the exclusive token identification is the token similarity. With this matter, thesauri such as E-hownet [8], Chinese concept dictionary [9], Tongyichichilin [10], and Sinica Bow [11] is used in our system. We estimate the similarity of two tokens by checking the mutual distance on the taxnonmy of these thesauri. If the similarity of two tokens from S and S' respectively is greater than the predefined threshold, these two tokens will be regarded as synonyms.

4.3 Dependency Word pair Based Entailment Recognition

In this recognition module, dependency word pairs (DWP) which are composed of tokens will be extracted. Figure 2 shows an example of extracting dependency word pairs from the sentence *"廓爾喀族入侵尼泊爾" (Gurkhas invaded Nepal)*. In order to reduce noisy information and remove relatively unimportant word pairs which may contains stopwords and function words, only dependency word pairs with POS combinations of adjective-noun, noun-noun, noun-verb, verb-noun, and adverb-verb, are selected. After POS filtering, the entailment relation of (S, S') can be identified by checking the number of remaining dependency word pairs. The rule of determining entailment relation demonstrates as follow:

$$\mathrm{Entailment}(S, S') = \begin{Bmatrix} \text{false, if there exist an DWP that } DWP \in S' \text{ and } DWP \notin S \\ \text{true, otherwise} \end{Bmatrix}$$

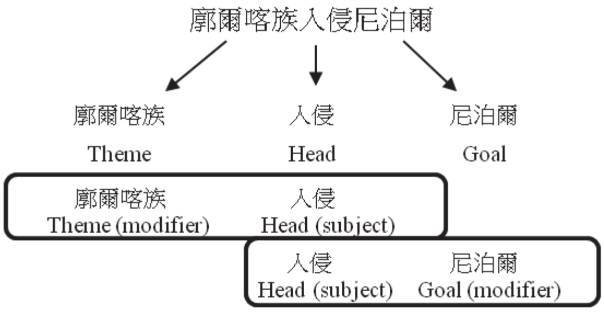

Figure 2. Example of extract dependency word pair from parsing tree

4.4 Sentence Length Based Entailment Recognition

Unlike the previous 3 modules, sentence length based entailment recognition follows a very simple idea that the length of sentence indicates the quantity of containing information. In NTCIR-9 RITE, it was a very useful feature to determine the direction of entailment. So we keep using this feature in this year's participation. Although sentence length can only report forward, reverse and bidirectional relations in NTCIR-9 RITE, it still can work fine in binary classification.

5. CONTRADICTION RECOGNITION

In contradiction recognition, we have two modules which are based on named entity and Chinese tokens. We will describe the detail of the modules in this section.

5.1 Named Entity based Contradiction Recognition

We predefined several circumstances which may indicate the occurrence of contradictive statements for detecting conflicted sentence pairs. These circumstances can be summarized as the following rule:

Assume a named entity list $NE_S \subset S$ and another named entity list $NE_{S'} \subset S'$. If there exist two named entities n_i and n_j that $n_i \neq n_j$, $n_i \in NE_S$, and $n_j \in NE_{S'}$. Then the sentence pair (S, S') are contradictive pair when all the following circumstances are matched.

1. The NE category of n_i and n_j are the same.
2. n_i and n_j refer to the same event.
3. n_i and n_j play the same semantic role in the event.

Like the entailment recognition, contradiction recognition need to consider NE synonym, temporal expressions, and the unit of numerical expressions to verify if $n_i \neq n_j$. Therefore, the analysis process applied in named entity based entailment recognition will be applied in this module as well.

The information of dependency word pairs provided by preprocessing are used to verify if n_i and n_j refer to the same event and play the same role. We extract the dependency word pairs from S and S' that n_i and n_j involved and observe the dependency. If these extracted word pairs contain the same predicate (verb or adjective), then these two word pairs could refer to the same event, and the semantic role of NE could be identified by the dependency. For instances, comparing the following two sentences:

"廓爾喀族入侵尼泊爾" (Gurkhas invaded Nepal).
"廓爾喀族入侵印度" (Gurkhas invaded India).

These two sentences will be considered as contradictive pairs for the reason that they own the same predicate but different NE for the argument.

5.2 Chinese Token Based Contradiction Recognition

Chinese token based contradiction recognition mainly relies on antonyms, negatives, incompatible statements, and unexchangeable token order in text.

- Antonyms and negatives: An antonyms and negative list from E-hownet**,** Chinese concept dictionary, and Sinica Bow are collected for contradiction detection. We also manually annotate some extra antonyms and negatives for improving the coverage of contradictive statements.
- Incompatible statement: A sentences may opposite to another due to the occurrence of incompatible statements such as manner, time period, and degree. We derive a list of incompatible terms via analyzing the lexicon definitions from E-hownet. When S and

S' contain incompatible statements, the pair (S, S') will be considered as contradictive, For example, if two tokens *早上 (morning)* and *下午(afternoon)* show in S and S' respectively, these two tokens are consider as incompatible.
-Unexchangeable token order: The location of a token may indicate the semantic role it plays in the sentence. If we exchange two tokens, it may cause the shift in sentence meaning. We found that most of the meaning-changing token exchanges are related to the relative location to the main verb of the sentence. Thereafter we set a rule that if a token exchange varies the order between the main verb and exchanging tokens, then this token exchange will cause the meaning conflict because the semantic role of the token may has been changed.

6. EVALUATION

Both BC and MC traditional Chinese test sets contain 881 unlabeled sentence pairs. Simplified Chinese dataset have 781 sentence pairs in MC and BC test data.

Our system was developed on the basis of using traditional Chinese knowledge sources and tools. When facing simplified Chinese sentences, we converted simplified Chinese sentence pairs into traditional Chinese pairs by simple character transforming, and use exactly the same system to process them. After our recognition process, the system output, which is multiple classification, is converted to BC labels for BC track.

We submitted two runs to compare the differences between our two participations of NTCIR RITE. Table 2 shows the official evaluation results of our works.

Table 2. Evaluation result comparison between our NTCIR-9 RITE (RITE 1) and NTCIR-10 RITE system (RITE 2) participating systems.

	CT-BC	CT-MC	CS-BC	CS-MC
RITE 1	0.5177	0.3304	0.5060	0.3495
RITE 2	0.6714	0.4632	0.6045	0.5094

The precision and recall of the 4 MC labels are shown in Table 3. The performance of the two runs in each track is shown in Table 3, 4, 5, and 6.

Table 3. Evaluation result comparison on multiple-class (MC) classification in traditional Chinese (CT)

	RITE 1 Precision/Recall	RITE 2 Precision/Recall
F	.4758/.8384	.5399/.8049
B	.4471/.2517	5306/.5166
C	.2243/.2105	.3625/.2544
I	.3153/.1215	.5273/.3021

Table 4. Evaluation result comparison on multiple-class (MC) classification in simplified Chinese (CS)

	RITE 1 Precision/Recall	RITE 2 Precision/Recall
F	.4705/.8051	.5751/.7329
B	.4835/.3034	.6134/.5034
C	.2600/.2453	.4021/.3679
I	.2845/.1304	.5000/.4190

Table 5. Evaluation result comparison on binary-class (BC) classification in traditional Chinese (CT)

	RITE 1 Precision/Recall	RITE 2 Precision/Recall
Y	.5684/.7808	.6864/.7495
N	.5291/.2935	.6648/.5920

Table 6. Evaluation result comparison on binary-class (BC) classification in simplified Chinese (CS)

	RITE 1 Precision/Recall	RITE 2 Precision/Recall
Y	.5558/.7441	.7098/.6190
N	.5000/.3008	.6682/.3983

The RITE4QA track, which is applied to measure the impact of RITE performance to question answering system, is conducted in both RITE campaigns. The evaluation result of RITE4QA of our system is shown in Table 7 and 8.

Table 7. Evaluation result of RITE4QA in traditional Chinese (CT)

	Worse Ranking Top1/MMR/Top5	Better Ranking Top1/MMR/Top5
RITE1	.1200/.2214/.3933	.3000/.3827/.5067
RITE2	.1067/.1601/2733	.3800/.4274/.4933

Table 8. Evaluation result of RITE4QA in simplified Chinese (CS)

	Worse Ranking Top1/MMR/Top5	Better Ranking Top1/MMR/Top5
RITE1	.1200/.2271/.4200	.3133/.3809/.4867
RITE2	.1267/.1880/.3200	.3800/.4271/.5000

7. CONCLUSION

The system we developed for NTCIR-10 RITE integrates two knowledge-based recognition modules relying on shallow linguistic features such as Chinese tokens, named entities, and dependencies. The evaluation results show that the system performance has been significantly improved comparing with our last RITE participation work. The two-stage architecture is also beneficial to the accuracy of recognition, especially on contradiction cases. However, the knowledge we used in the recognition modules are still very simple and surface. Relation such as independent statement still cannot find a promising approach to handle. Addressing deep understanding problem by probing those relative complicated cases may be our target for the future researches of textual entailment recognition.

ACKNOWLEDGMENTS
This research was supported in part by the National Science Council under grant NSC 101-3113-P-001-004.

REFERENCES

[1] Ido Dagan, and Oren Glickman,. Probabilistic textual entailment: Generic applied modeling of language variability. PASCAL workshop on Text Understanding and Mining, 2004.

[2] H. Shima, H. Kanayama, C.-W. Lee, C.-J. Lin, T. Mitamura, Y. Miyao, S. Shi, and K. Takeda. Overview of NTCIR-9 RITE: Recognizing Inference in TExt. In NTCIR-9 Proceedings, to appear, 2011.

[3] Cheng-Wei Shih, Cheng-Wei Lee, Ting-Hao Yang, and Wen-Lian Hsu (2011), "IASL RITE System at NTCIR-9," in Proceedings of NTCIR-9 Workshop, Tokyo, Japan (2011).

[4] Ma, Wei-Yun and Keh-Jiann Chen, 2003, "Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff", Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing, pp168-171.

[5] Cheng-Wei Lee, Cheng-Wei Shih, Min-Yuh Day, Tzong-Han Tsai, Tian-Jian Jiang, Chia-Wei Wu, Cheng-Lung Sung, Yu-Ren Chen, Shih-Hung Wu and Wen-Lian Hsu (2005), "ASQA: Academia Sinica Question Answering System for NTCIR-5 CLQA," in Proceedings of NTCIR-5 Workshop, Tokyo, Japan, 202-208, (2005).

[6] Cheng-Wei Lee, Min-Yuh Day, Cheng-Lung Sung, Yi-Hsun Lee, Tian-Jian Jiang, Chia-Wei Wu, Cheng-Wei Shih, Yu-Ren Chen, Wen-Lian Hsu, "Chinese-Chinese and English-Chinese Question Answering with ASQA at NTCIR-6 CLQA", in Proceedings of NTCIR-6 Workshop Meeting, Tokyo, Japan, May 15-18, 2007, pp. 175-181. (2007)

[7] Chinese Parser, http://ckip.iis.sinica.edu.tw/CKIP/parser.htm

[8] Chen Keh-Jiann, Shu-Ling Huang, Yueh-Yin Shih, Yi-Jun Chen, 2005, Extended-HowNet- A Representational Framework for Concepts, OntoLex 2005 - Ontologies and Lexical Resources IJCNLP-05 Workshop, Jeju Island, South Korea.

[9] Yu J.S. and Yu S.W. et al 2001 Introduction to Chinese Concept Dictionary, in International Conference on Chinese Computing (ICCC2001), pp361-367.

[10] 梅家驹,竺一鸣，高蕴琦等编.同义词词林.上海：上海辞书出版社，1983

[11] Chu-Ren Huang, Ru-Yng Chang, and Shiang-Bin Lee. 2004. Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO. In Proceedings of LREC2004, pages 1553- 1556, Lisbon, Portugal.