Experiments for NTCIR-11 RITE-VAL Task at Shibaura Institute of Technology

Toru Sugimoto Toshiki Mizukoshi Ryosuke Masuda
Department of Information Science and Engineering, Shibaura Institute of Technology
3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan
sugimoto@shibaura-it.ac.jp

ABSTRACT

This paper reports the evaluation results of our textual entailment system at NTCIR-11 RITE-VAL task. We participated in the Japanese System Validation (SV) and Fact Validation (FV) subtasks. In our system, the meaning of a text is represented as a set of dependency triples consisting of two words and their relation. Comparing two sets of dependency triples with respect to conceptual similarity, we determine a correspondence between dependency triples, and finally identify textual entailment based on the overlap of dependency triples. This paper provides a description of our algorithm, the evaluation results and discussion on the results.

Team Name

SITLP

Subtasks

Fact Validation and System Validation (Japanese)

Keywords

textual entailment, dependency triple, conceptual similarity

1. INTRODUCTION

This paper reports the evaluation results of our textual entailment system at NTCIR-11 RITE-VAL task [1]. The meaning of a sentence can typically be considered as a set of *facts* expressed in it. In this view, textual entailment between two texts can be seen as a subsumption relation between sets of facts expressed in them. In a sentence, facts are often expressed by pairs of words, or *bunsetsu* segments in Japanese, one of which depends on the other. We introduce a notion of a *dependency triple* that consists of two dependent words and their relation type, which is often represented by functional expressions, e.g., postpositional particles, in the sentences. Therefore, sentence meaning is modeled as a set of dependency triples, and textual entailment as identifying how much of dependency triples in t_2 are subsumed in t_1. By taking this approach, we can deal nicely with word dependency as well as word transposition and adnominal clauses.

In order to determine to what extent a triple is subsumed in a text, we need to compare two triples. We define a conceptual similarity measure between two triples that can be calculated using EDR dictionary [2], Japanese WordNet [3] and Wikipedia redirect. With this measure, we determine a correspondence relation between two sets of triples, and finally identify textual entailment based on the overlap of triples.

2. ALGORITHM

2.1 Overview

Processing flow of our system is illustrated in Figure 1.

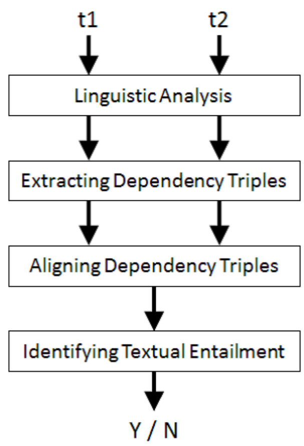

Figure 1: Processing Flow

When a text pair (t_1, t_2) is given, we first conduct linguistic analysis of each text. Then we extract dependency triples from each text, and obtain two sets of dependency triples corresponding to t_1 and t_2. Comparing these sets, we determine a correspondence between dependency triples, and identify whether t_1 entails t_2 or not based on the overlap of dependency triples.

2.2 Linguistic Analysis

First, we conduct the morphological analysis using MeCab [4] and the dependency analysis using CaboCha [5] for input Japanese texts t_1 and t_2. In these analyses, we also refer to Wikipedia entry names and a Japanese functional expression dictionary called Tsutsuji [6] to recognize proper nouns and functional expressions as whole words.

Then, referring to the EDR Japanese word dictionary, Japanese WordNet and Wikipedia redirect, we attach concept names (concept IDs, synsets and redirected entry names, respectively) to the content word of each bunsetsu segment in t_1 and t_2. We also attach functional expression IDs and semantic equivalence classes to functional expressions occurring in t_1 and t_2.

2.3 Extracting Dependency Triples

In this phase, a set of dependency triples is created from each text. A dependency triple (w_1, p, w_2) consists of two words w_1 and w_2, where w_1 depends on w_2, and the relation type p between w_1 and w_2. For each pair of dependent words contained in the dependency analysis result of a text, we create a dependency triple, in which the relation type is typically identified with a sequence of functional expressions or a postpositional particle of the dependent word. Exceptions are adnominal clauses and binding particles such as "*wa*" and "*mo*"; in these contexts, appropriate case particles are inferred using heuristic rules.

For various reasons, similar facts are not necessarily expressed by the same dependency structure in different texts. Therefore, we need to expand the set of dependency triples in order to enable more flexible matching and achieve higher recall. We apply the expansion rules shown in Table 1 to the triple set for t_1.

Table 1: Rules for Expanding a Triple Set

Condition	Added Triple
$(w_1, , w_2)$ where w_1 is a verb or an adjective and w_2 is a noun	$(w_2, , w_1)$
(w_1, p, w_2) where p is a coordinating particle such as "*to*" and "*ya*"	(w_2, p, w_1)
$(w_1, p_1, w_2), (w_2, p_2, w_3)$ where p_1 is a coordinating particle	(w_1, p_2, w_3)
$(w_1, p, w_3), (w_2, , w_3)$ where a w_2 is a verbal noun and w_3 is a verb "*suru*"	$(w_1, p, w_2{+}w_3)$

2.4 Aligning Dependency Triples

In order to identify how much of dependency triples in t_2 are subsumed in t_1, we conduct alignment between the triples in t_1 and the triples in t_2. Let T_1 and T_2 be the sets of triples in t_1 and t_2, respectively. The aim of this phase is to determine a triple $u \in T_1$ that *corresponds* to each $v \in T_2$ based on the conceptual similarity.

The similarity between dependency triples is the arithmetic mean of the similarity of each component of the triples. To calculate the similarity between triples (w_1, p, w_2) and (w'_1, p', w'_2), we calculate the similarity between w_1 and w'_1, the similarity between p and p' and the similarity between w_2 and w'_2, and figure out the sum of these three similarities divided by 3. The similarity between content words w_1 and w'_1 (and also between w_2 and w'_2) is calculated by using thesaurus, that is, EDR concept dictionary and Japanese WordNet. If all morphemes in w'_1 are contained in w_1 or w_1 is a descendant of w'_1 in the thesaurus, the similarity between w_1 and w'_1 is 1. In other cases, the similarity is calculated by using a variation of the path length based method [7], which considers the proportion between depths of concepts of both words and the most specific concept that subsumes them. For the similarity between functional expressions p and p', we use information in functional expression dictionary Tsutsuji. If p and p' has a common semantic equivalence class, the similarity is 1. Otherwise, the similarity is 0.

We determine a triple $f(v) \in T_1$ that corresponds to each $v \in T_2$ by the following procedure: for each triple $v \in T_2$, we calculate the similarity $sim(u, v)$ for each triple $u \in T_1$, and select as $f(v)$ a triple u that has the highest similarity toward v.

2.5 Identifying Textual Entailment

Based on the alignment explained in the previous section, we calculate a subsumption score as follows:

$$\text{subsumption score} = \frac{1}{|T_2|} \sum_{v \in T_2} sim(f(v), v)$$

where $sim(f(v), v)$ is the similarity between two triples $f(v)$ and v.

We conclude that t_1 entails t_2 if the calculated subsumption score is greater than a threshold value predetermined through a preliminary experiment using the training data.

2.6 Example

In this section, we illustrate our algorithm with the following simple example.

t1: きれいなツバメが低く飛んだ。
(A beautiful swallow flied low.)

t2: 鳥が低く飛んでいた。
(A bird was flying low.)

The dependency triples extracted from each text and the alignment result are shown in Figure 2.

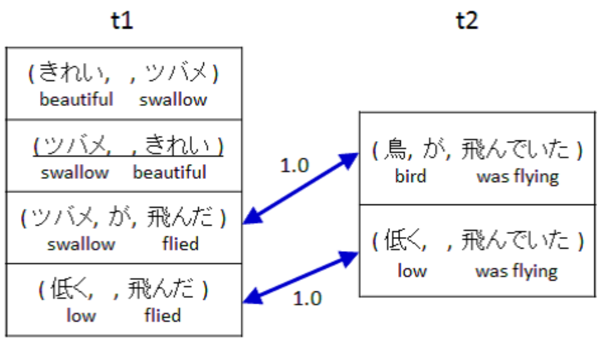

Figure 2: Dependency Triples and Alignment

The underlined triple in the figure is obtained by an expansion rule in Table 1. Since "*swallow*" is a descendant of "*bird*" in the thesaurus, the conceptual similarity between them is 1. The subsumption score between t_1 and t_2 is $(1 + 1)/2 = 1$, and we conclude that t_1 entails t_2.

2.7 Text Search Method

For the Fact Validation subtask, we implemented an algorithm to find texts that are likely to entail the given text t_2. We use an open source search engine Apache Solr [8] to retrieve candidate texts from textbooks. First, we extract a set of content words from the given text t_2. Next, we specify each content word as a query to Solr to retrieve several texts that contain as many content words as possible. Finally, for each retrieved text t_1, we examine whether t_1 entails t_2. If t_2 is entailed by some t_1, the result is Y, and otherwise the result is N.

3. EXPERIMENT

For the formal run, we submitted one run for the System Validation subtask and one run for the Fact Validation subtask. For the System Validation subtask, the accuracy of our system was 0.7223 and the Macro F1 is 0.5178. The confusion matrix is shown in Table 2.

Table 2: Result of the System Validation Subtask

SITLP-JA-SV-01		Answer Y	Answer N	Total
System	Y	49	93	142
	N	290	947	1237
Total		339	1040	1379

For the Fact Validation subtask, the accuracy of our system was 0.6089 and the Macro F1 is 0.4489. The confusion matrix is shown in Table 3.

Table 3: Result of the Fact Validation Subtask

SITLP-JA-FV-01		Answer Y	Answer N	Total
System	Y	18	11	29
	N	190	295	485
Total		208	306	514

4. DISCUSSION

The formal run result shows that the subsumption score calculated by our algorithm is inadequate to deal with various patterns of textual entailment occur in the formal run data. In particular, there are many cases in which our system failed to recognize entailment, that is, the subsumption score is low but the correct answer is Y. Analyzing such error cases, we found various reasons including paraphrase understanding, commonsense reasoning and numerical reasoning.

Another problem is known as non-monotonicity of linguistic contexts. Contexts such as negations and conditionals are said to be *downward-monotone*, and need to be dealt with carefully. For example, consider the following conditional statements.

t1: ツバメが低く飛ぶと、雨が降る。
(If a swallow flies low, it will rain.)

t2: 鳥が低く飛ぶと、豪雨が降る。
(If a bird flies low, it will rain heavily.)

The dependency triples extracted from each text and the alignment result are shown in Figure 3.

Although "*swallow*" is a descendant of "*bird*" in the thesaurus, since these words occurred in conditional clauses that are downward-monotone contexts, the entailment direction is reversed from "<" to ">", and as a result t_2 entails t_1.

To deal with this kind of linguistic phenomenon, MacCartney presented a resoning method using natural logic [9], and we applied this method to Japanese texts [10]. We are trying to adapt our method [10] to dependency triple based reasoning framework described in this paper, but detailed experiment is left for future work.

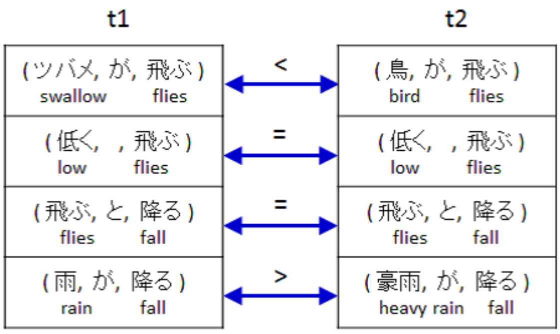

Figure 3: Dependency Triples and Alignment

5. CONCLUSION

In this paper, we reported the evaluation results of our textual entailment system at NTCIR-11 RITE-VAL task. We described our algorithm and the evaluation results, and discussed on the reasons for error and the future work.

6. REFERENCES

[1] Suguru Matsuyoshi, Yusuke Miyao, Tomohide Shibata, Chuan-Jie Lin, Cheng-Wei Shih, Yotaro Watanabe and Teruko Mitamura: Overview of the NTCIR-11 Recognizing Inference in TExt and Validation (RITE-VAL) Task, Proceedings of the 11th NTCIR Conference, 2014.

[2] EDR: EDR Electronic Dictionary Technical Guide, Japan Electronic Dictionary Research Institute, Ltd., 1993.

[3] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama and Kyoko Kanzaki: Development of Japanese WordNet, Proceedings of LREC 2008.

[4] Taku Kudo, Kaoru Yamamoto and Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of EMNLP 2004.

[5] Taku Kudo and Yuji Matsumoto: Japanese Dependency Analysis using Cascaded Chunking, Proceedings of 6th CoNLL, 2002.

[6] Suguru Matsuyoshi, Satoshi Sato and Takehito Utsuro: A Dictionary of Japanese Functional Expressions with Hierarchical Organization, Journal of Natural Language Processing, Vol. 14, No. 5, 2007.

[7] Zhibiao Wu and Martha Palmer: Verb Semantics and Lexical Selection, Proceedings of 32nd ACL, 1994.

[8] Solr: http://lucene.apache.org/solr/.

[9] Bill MacCartney and Christopher D. Manning: Natural Logic for Textual Inference, Proceedings of WTEP 2007.

[10] Ryosuke Masuda and Toru Sugimoto: Natural Logic for Japanese Textual Inference, Proceedings of the 26th Annual Conference of the Japanese Society for Artificial Intelligence, 2012.