A Textual Entailment System using Web based Machine Translation System

Partha Pakray*, Snehasis Neogi*, Sivaji Bandyopadhyay*, Alexander Gelbukh‡

*Jadavpur University, Kolkata-700032, India

‡Center for Computing Research, National Polytechnic Institute, Mexico City, Mexico

Abstract

The article presents the experiments carried out as part of the participation in Recognizing Inference in Text (NTCIR-9 RITE) for Japanese. NTCIR-9 RITE has four subtasks, Binary-class (BC) subtask, Multi-class (MC) subtask, Entrance Exam and NTCIR-9 RITE4QA. We have participated in all the four subtasks.

Task Name	Accuracy
BC Subtask	Run 1: 0.490
	Run 2: 0.500
	Run 3: 0.508
MC Subtask	Run: 0.175
Entrance Exam Subtask	Run: 0.5204
RITE4QA	Run: 0.5954

System Architecture 1 (Run 1): BC Task

The system accepts pairs of text snippets (text and hypothesis) at the input and gives a Boolean value at the output: Y if the t1 entails the t2 and N otherwise.

The architecture of the proposed system is described in Figure 1.

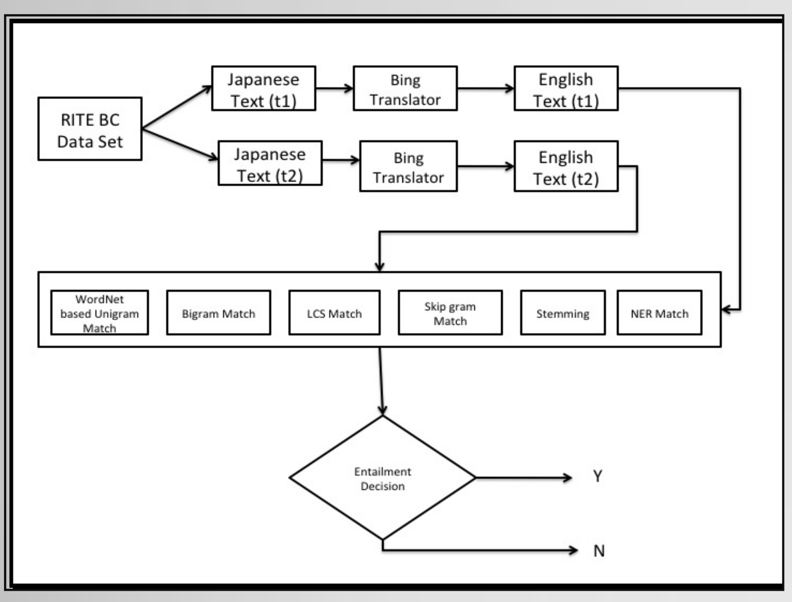

Figure 1: System Architecture BC Subtask (TE Recognition system using Web based MT System)

The system first identifies the (t1, t2) Japanese text segments from the NTCIR-9 RITE BC Subtask data. The Japanese (t1, t2) text segment is translated to English (t1, t2) text segment using the Japanese-English Bing translator machine translation (MT) system. The translated English text segments are then submitted to the textual entailment (TE) recognition system. The various components of the TE recognition system are:

- WordNet based Unigram Match
- Bigram Match
- Longest Common Subsequence
- Skip-grams
- Stemming
- Named Entity

For each component of the TE recognition system we determine a threshold value which is based on the experiments carried out on the NTCIR-9 RITE BC subtask development set. If the value of a particular TE module when applied on the NTCIR-9 RITE BC subtask test data is equal to or more than the predetermined threshold value, then the text-hypothesis pair is considered as an entailment. The pair is assigned the value of 1 meaning entailment (Y); otherwise, the pair is assigned a value of 0 meaning no entailment (N).

Textual Entailment Module (TE)

This TE module [1] is based on three types of matching

- WordNet based Unigram Match
- Bigram Match
- Skip-grams

System Architecture 2 (Run 2): BC Task

The system architecture 2 is based on N-Gram matching on the input Japanese text pair {t1, t2} to identify whether t1 entails (infers) a hypothesis t2 or not. The system first learns the N-Gram (unigram and bigram) word overlap on the development data. The learned system is then applied on the test data to classify whether the text entails the hypothesis or not.

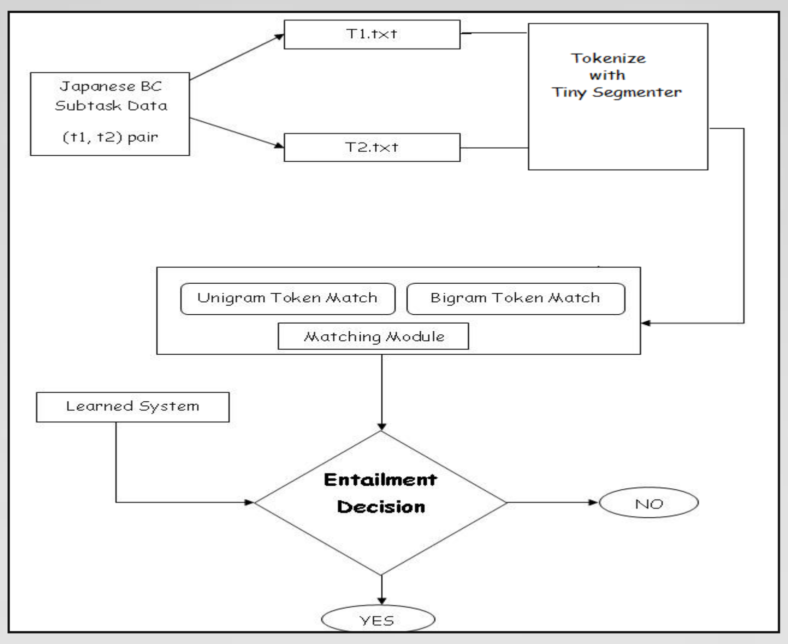

Figure 2. System Architecture Japanese BC subtask

In this architecture we calculate the percentage of unigram and bigram match experimented over the NTCIR-9 RITE BC subtask test data. These values are then compared with the predetermined threshold value carried out on the NTCIR-9 RITE BC subtask development data. If the N-gram matching percentage is more than or equal to the threshold value, entailment label 'YES' is assigned with the pair. Otherwise, entailment label "NO" is assigned which means no entailment.

System Architecture 3 (Run 3): BC Task

The system considers the output of the previous two systems (Run 1 and Run 2) as input. The entailment decision is taken based on voting on the decisions in Run 1 and Run 2.

ANDing Module to match labels

The system reads the labels of the system generated output files of Run 1 and Run 2 for a particular pair id. If the entailment decision of both the runs agree then this is output as the final entailment label. Otherwise, if they do not agree, the final entailment label will be "N" (NO).

System Architecture 4 (Run 1): MC Task

The RITE MC subtask development data is similar to that of the BC Subtask except the entailment label. Here multi – way entailment label is used to detect entailment or no entailment in the text – hypothesis pair. The 5-way labeling subtask detects (forward / reverse / bidirectional) entailment or no entailment (contradiction / independence) in a text pair.

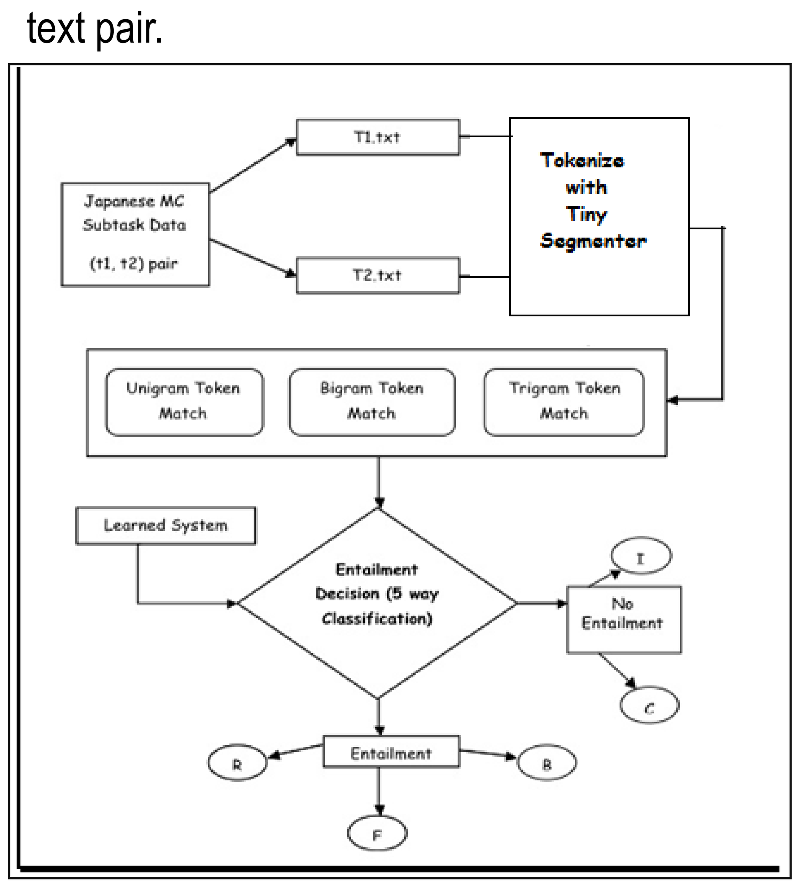

Figure 3: NTCIR-9 RITE MC subtask System Architecture

The NTCIR-9 RITE MC subtask test data is tokenized and the different N-Gram percentages (Unigram, bigram, trigram) are evaluated. These percentage values are then compared with the percentage values calculated over the development data. If a match is found for the percentage values for each text – hypothesis pair of the test data, the same entailment label {F,R,B,C,I} of the development data is used to classify the text – hypothesis of the test set. Otherwise it classifies the pair as I (Independence).

System Architecture 5: Entrance Exam Task

The Entrance Exam subtask is similar to that of BC subtask in terms of input and output. All the data are created from actual college-level entrance exams. In this subtask the entailment label is same as the BC subtask {Y, N}. The system for the entrance exam Subtask is similar to the system architecture 2 used for the BC subtask with some additions in the matching module.

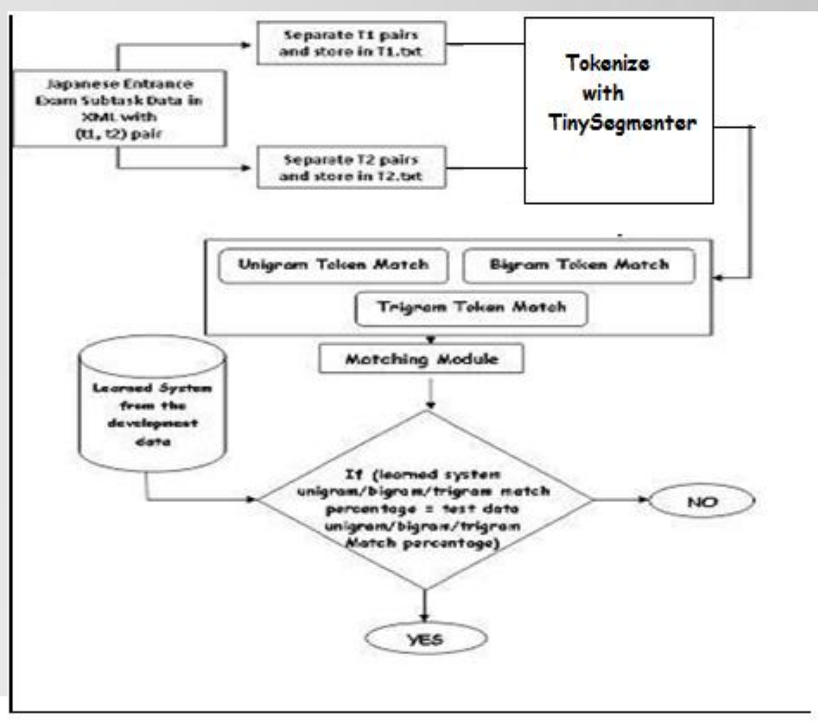

Figure 4. System Architecture for Entrance Exam and NTCIR-9 RITE4QA Subtask

The matching module stores different N-gram percentage values calculated over the NTCIR-9 RITE Entrance Exam subtask test data and compares these values with the previously learned values carried out on the development data. If the percentage values of the test data matches with the learned data, then the corresponding gold standard label is used to classify the test data.

System Architecture 6: RITE4QA

The RITE4QA Subtask is also similar to the BC subtask in terms of input and output, but it includes an embedded answer validation component in the Question Answering system. This way, the impact of NTCIR-9 RITE to an overall end-to-end application can be measured.

In this subtask, the same architecture and modules as in the NTCIR-9 RITE Entrance Exam Subtask are followed.

Conclusions

In this paper, textual entailment systems mainly based on the lexical similarity modules have been developed. For MC subtask, the system can learn using some supervised learning methods (e.g. SVM,CRF).

Acknowledgements

We acknowledge the support of the CONACYT Mexico – DST, India project "Answer Validation through Textual Entailment".

Authors	Email
Partha Pakray	parthapakray@gmail.com
Snehasis Neogi	snehasis1981@gmail.com
Sivaji Bandyopadhyay	sivaji_cse_ju@yahoo.com
Alexander Gelbukh	gelbukh@gelbukh.com