Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10

Yotaro Watanabe, Yusuke Miyao, Junta Mizuno, Tomohide Shibata, Hiroshi Kanayama, Cheng-Wei Lee, Chuan-Jie Lin, Shuming Shi, Teruko Mitamura, Noriko Kando, Hideki Shima and Kohichi Takeda

1Tohoku University, Japan 2National Institute of Informatics, Japan 3Kyoto University, Japan 4IBM Research, Japan 5Academia Sinica, Taiwan 6National Taiwan Ocean University, Taiwan 7Microsoft Research Asia, P.R.China 8Carnegie Mellon University, U.S.A

INTRODUCTION
This poster introduces an overview of the RITE2 task in NTCIR-10. We evaluate systems that automatically recognize entailment, paraphrase, and contradiction between two texts written in Japanese, Simplified Chinese, or Traditional Chinese. The task consists of six subtasks: Binary classification of entailment (BC); Multi-class classification including paraphrase and contradiction (MC); Unit Test which provides a breakdown of linguistic phenomena (UnitTest); and three extrinsic application-oriented datasets: Exam BC, Exam Search and RITE4QA.

ADVANCES FROM NTCIR-9 RITE

DATASET SIZE

FORMAL RUN RESULTS

OUTCOME

- We constructed large-scale reusable evaluation datasets
- Baseline Tool is a Python development framework for implementing a RITE system based on machine learning
- Automatic evaluation output not only accuracy and macro-F1 but also confusion matrix and details of the results on the Unit Test data
- Best runs were able to outperform the simple machine learning-based baseline
- In the JA-BC subtask, the top system achieved over 80% in accuracy (in contrast to 58% in NTCIR-9 RITE)
- UnitTest dataset enabled detailed analysis of RTE systems
- Exam Search subtask is more realistic than the traditional settings

http://www.cl.ecei.tohoku.ac.jp/rite2/doku.php

Entrance Examination Subtask
Created based on actual entarance exams for university admission in Japan: National Center Test for University Admission. The subtask provides two types of data.

BC style (ExamBC)
The data is provided in the same form as the BC subtask. Systems are asked to recognize inference relations between t1 and t2. In this data, t1 is extracted from Wikipedia, while t2 is taken from university entrance exams.

Search style (ExamSearch)
In this data, t1 is not given. Systems are asked to retrieve texts that can be used as t1 from Wikipedia or textbooks, and answer whether t2 is entailed (inferred) from retrieved texts.

RITE4QA Subtask
Another application-oriented dataset created from the NTCIR-6 CLQA dataset (does an answer-candidate-bearing sentence entail a question in affirmative form?).

\[ t_1: \text{Yasunari Kawabata won the Nobel Prize in Literature for his novel “Snow Country”} \]
\[ t_2: \text{Yasunari Kawabata is the writer of “Snow Country”} \]