IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-10 RITE-2

Min-Yuh Day *, Chun Tu, Shih-Jhen Huang, Hou Cheng Vong, Sih-Wei Wu

* myday@mail.tku.edu.tw

Department of Information Management, Tamkang University, New Taipei City, Taiwan

In this paper, we describe the IMTKU (**I**nformation **M**anagement at **T**am**K**ang **U**niversity) textual entailment system for recognizing inference in text at NTCIR-10 RITE-2 (Recognizing Inference in Text). We proposed a textual entailment system using a hybrid approach that integrate semantic features and machine learning techniques for recognizing inference in text at NTCIR-10 RITE-2 task. We submitted 3 official runs for BC, MC and RITE4QA subtask. In NTCIR-10 RITE-2 task, IMTKU team achieved 0.509 in the CT-MC subtask, 0.663 in the CT-BC subtask; 0.402 in the CS-MC subtask, 0.627 in the CS-BC subtask; In MRR index, 0.257 in the CT-RITE4QA subtask, 0.338 in the CS-RITE4QA subtask. IMTKU is ranked #1 in the CS-RITE4QA subtask of NTCIR-10 RITE-2 task.

System Architecture
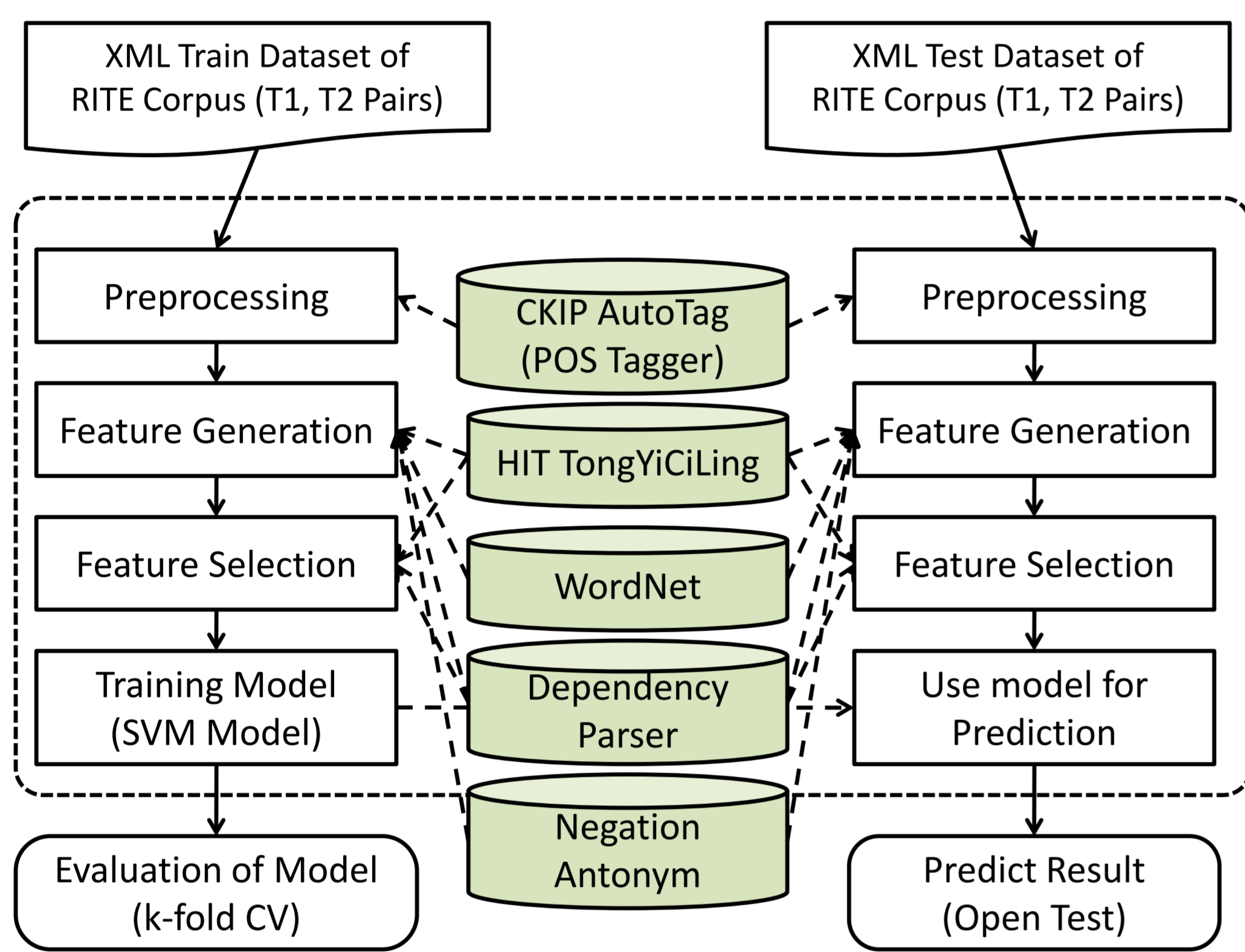
Performance

IMTKU **CT BC** Subtask Official Runs	Accuracy
RITE2-IMTKU-CT-BC-01	0.663
RITE2-IMTKU-CT-BC-02	0.515
RITE2-IMTKU-CT-BC-03	0.643

IMTKU **CT MC** Subtask Official Runs	Accuracy
RITE2-IMTKU-CT-MC-01	0.509
RITE2-IMTKU-CT-MC-02	0.366
RITE2-IMTKU-CT-MC-03	0.501

IMTKU **CS RITE4QA** Subtask Official Runs	Top1	MRR	Top5
RITE2-IMTKU-CS-RITE4QA-01	0.1067	0.1991	0.3867
RITE2-IMTKU-CS-RITE4QA-02	0.1467	0.2144	0.3600
RITE2-IMTKU-CS-RITE4QA-03	0.2800	0.3377	0.4267

IMTKU **CT RITE4QA** Subtask Official Runs	Top1	MRR	Top5
RITE2-IMTKU-CT-RITE4QA-01	0.1467	0.2258	0.3733
RITE2-IMTKU-CT-RITE4QA-02	0.1200	0.1984	0.3267
RITE2-IMTKU-CT-RITE4QA-03	0.1733	0.2603	0.4067
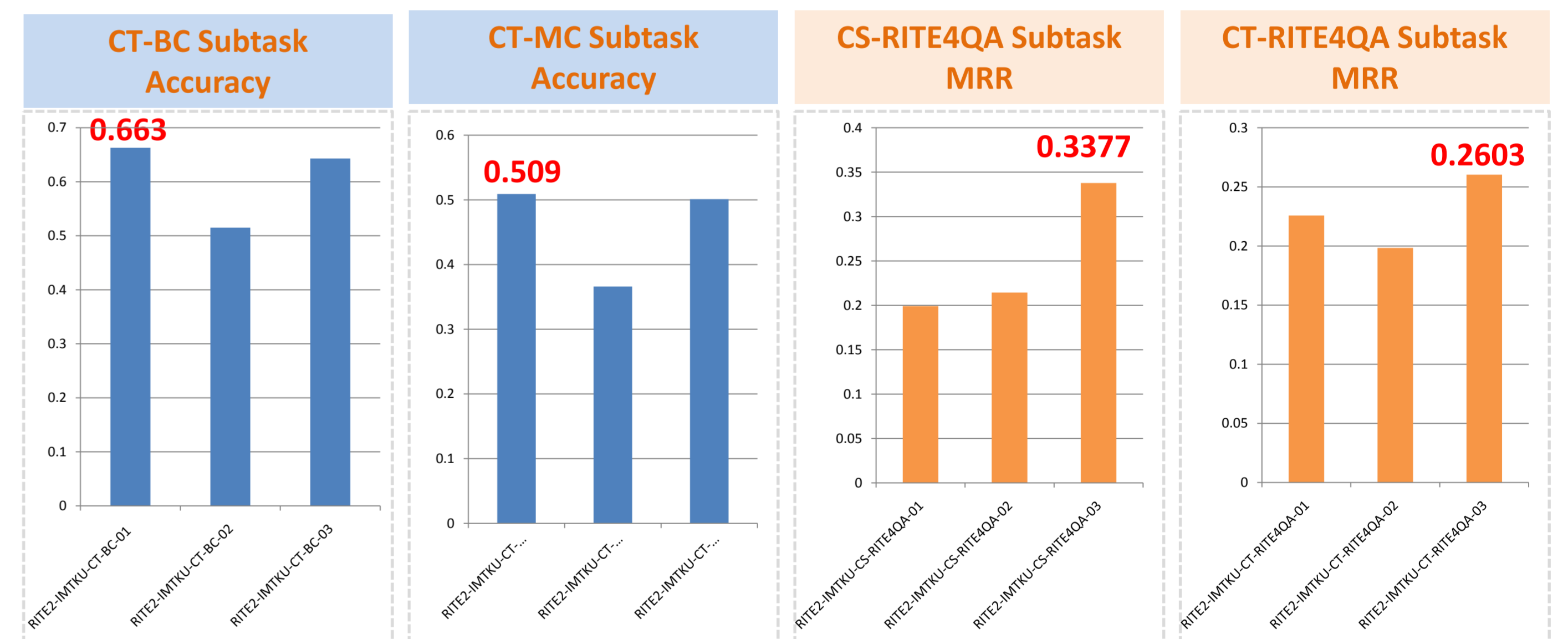
Methods for Official Runs

• RITE-2-IMTKU-CT-BC Subtask

RITE-2-IMTKU-CT-BC-01
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL), Stanford Parser
Method: Hybrid approach (Integrated Semantic features and Machine Learning Approach) for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity, Tree Edit Distance) in SVM.

RITE-2-IMTKU-CT-BC-02
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL), Stanford Parser
Method: Hybrid approach (Integrated Semantic features and Machine Learning Approach) for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

RITE-2-IMTKU-CT-BC-03
Tools: CKIP AutoTag, LibSVM
Resources: Stanford Parser
Method: Machine Learning Approach for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2.
Multiple Features used (Longest Common Substring, Word Length Ratio, Text Length, Similarity between t1 and t2, Tree Edit Distance) in SVM.

• RITE-2-IMTKU-CT-MC Subtask

RITE-2-IMTKU-CT-MC-01
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL)), Stanford Parser
Method: Hybrid approach (Integrated Knowledge Base and Machine Learning Approach) for NTCIR-10 RITE-2..
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2
Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity, Tree Edit Distance) in SVM.

RITE-2-IMTKU-CT-MC-02
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL)
Method: Hybrid approach (Integrated Knowledge Base and Machine Learning Approach) for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
Multiple Features used(Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

RITE-2-IMTKU-CT-MC-03
Tools: CKIP AutoTag, LibSVM
Resources: Stanford Parser
Method: Machine Learning Approach for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2.
Multiple Features used (Longest Common Substring, Word Length Ratio, Text Length, Similarity between t1 and t2, Tree Edit Distance) in SVM.

• RITE-2-IMTKU-CT-RITE4QA Subtask

RITE-2-IMTKU-CT-RITE4QA-01
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL)
Method: Hybrid approach (Integrated Knowledge Base and Machine Learning Approach) for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

RITE-2-IMTKU-CT-RITE4QA-02
Tools: CKIP AutoTag, LibSVM
Resources: HIT TongYiCiLing (HIT-TYCL)
Method: Machine Learning Approach for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Siilarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length) in SVM.

RITE-2-IMTKU-CT-RITE4QA-03
Tools: CKIP AutoTag, LibSVM
Resources: NONE
Method: Machine Learning Approach for NTCIR-9 RITE.
Feature Extraction from normalized t1 and t2.
Multiple Features used (Longest Common Substring, Text Length, Text Length Ratio, Antonym, Negation) in SVM.

• RITE-2-IMTKU-CS-RITE4QA Subtask

RITE-2-IMTKU-CS-RITE4QA-01
Tools: CKIP AutoTag, LibSVM
Resources: Bilingual Wordnet (SINICA BOW), HIT TongYiCiLing (HIT-TYCL)
Method: Hybrid approach (Integrated Knowledge Base and Machine Learning Approach) for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2.
Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

RITE-2-IMTKU-CS-RITE4QA-02
Tools: CKIP AutoTag, LibSVM
Resources: HIT TongYiCiLing (HIT-TYCL)
Method: Machine Learning Approach for NTCIR-10 RITE-2.
Feature Extraction from normalized t1 and t2.
Measure similarity match between t1 and t2.
Multiple Features used (Antonym, Negation, Word Based Siilarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length) in SVM.

RITE-2-IMTKU-CS-RITE4QA-03
Tools: CKIP AutoTag, LibSVM
Resources: NONE
Method: Machine Learning Approach for NTCIR-9 RITE.
Feature Extraction from normalized t1 and t2.
Multiple Features used (Longest Common Substring, Text Length, Text Length Ratio, Antonym, Negation) in SVM.

Discussion

- **Issues of Definition in RITE MC between NTCIR-9 and NTCIR-10:**
  - Definition of NTCIR-9 MC subtask : "A **5-way** labeling subtask to detect (**forward / reverse / bidirection**) entailment or no entailment (contradiction / independence) in a text pair."
  - Definition of NTCIR-10 MC subtask : "A **4-way** labeling subtask to detect (**forward / bidirection**) entailment or no entailment (contradiction / independence) in a text pair."
- Misused NTCIR-9 MC labels on NTCIR-10 MC test datasets where "Reverse" label should be excluded.

Cross Validation of Development and Test datasets of NTCIR-10 RITE-2 Task

Datasets	10 Fold CV Accuracy
RITE2_CT_dev_test_bc_g.txt (RITE2 BC Dev+Test Dataset: 1321+881 = 2202 pairs)	68.85%
RITE1_CT_r1000_dev_test_bc_g.txt (Random select **1000 pairs** from RITE1 BC Dev+Test Dataset)	73.83%
RITE1_CT_dev_test_bc_g.txt (RITE1 BC Dev+Test Dataset: 421+900 =**1321 pairs**)	72.29%
RITE1_CT_dev_bc_g.txt (gold standard) (RITE1 BC Development Dataset: **421 pairs**)	72.21%

DEMO: http://rite.im.tku.edu.tw

Information Management, Tamkang University

Tamkang University
淡江大學