KSU Team's System and Experience at the NTCIR-11 RITE-VAL Task

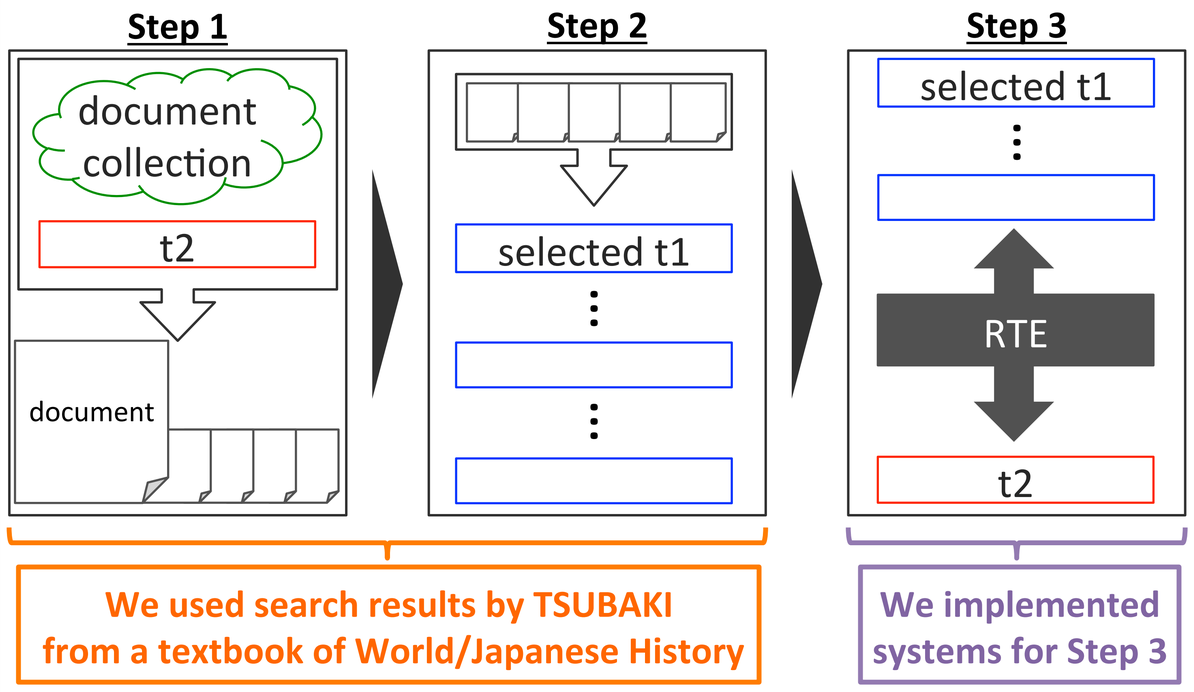
Tasuku Kimura, Hisashi Miyamori
Kyoto Sangyo University, Japan

Abstract

This paper describes the systems and results of the team KSU for RITE-VAL task in NTCIR-11. Three different systems were implemented for each of the two subtasks: Fact Validation and System Validation. In Fact Validation subtask, systems were designed respectively based on character overlap, existence of entailment result 'Y', and voting of entailment results. In System Validation subtask, systems were designed respectively using SVM, Random Forest, and Bagging, with features such as surface features, numerical expressions, location expressions, and named entities. Scores of the formal runs were 52.78% in macro F1 and 63.42% in accuracy with KSU-FV-02 in Fact Validation, and 66.96% in macro F1 and 79.84% in accuracy with KSU-SV-01 in System Validation.

Fact Validation

Fact Validation subtask is comprised of three steps. We implemented systems for Step 3. Systems were designed respectively based on character overlap, existence of entailment result 'Y', and voting of entailment results.



Our systems for FV are summarized in Table 1.

Table 1: Summary of our systems for FV

System	Initial RTE	Final RTE	Decision unit
KSU-JA-FV-01	character overlap ratio	threshold	sentence
KSU-JA-FV-02	base-MC	initial results for sentences	document
KSU-JA-FV-03	base-MC	voting of initial results for sentences	document

System Validation

In System Validation, systems were designed respectively using SVM(KSU-JA-SV-01), Random Forest(KSU-JA-SV-02), and Bagging(KSU-JA-SV-03), with features such as surface features, location expressions, named entities, and numerical expressions. Table 2 shows the features used in our systems.

Table 2: Features of for SV subtask

Category	Feature	Description	Data Type
Surface features	cos_sim_c	Cosine similarity of characters	Numeric
	cos_sim_w	Cosine similarity of content words	Numeric
	jc_coef_w	Jaccard coefficient of content words	Numeric
	lcs	LCS, the longest substrings common to t1 and t2,normalized by the length of t2	Numeric
Location	location	Whether location names in t2 are also referred to in t1	T/F
Named entities	ne_cos_sim	Cosine similarity of named entities	Numeric
	ne_diff	Whether a named entity exist in t2 which is not included in t1	T/F
	ne_n2subset	Whether all named entities in t2 are partially included in t1	T/F
Numerical expressions	numexp_diff	Whether one or more numerical expressions exist in t2 which do not match with those in t1	T/F
	numexp_exact	Whether all the numerical expressions in t2 are exactly included in t1	T/F
	numexp_n1subset	Whether all the numerical expressions in t1 are partially included in t2	T/F
	numexp_n2subset	Whether all the numerical expressions in t2 are partially included in t1	T/F

Formal Runs

The results of our systems for Fact Validation subtask and System Validation subtask were shown in Tables 3 and 4.

Table 3: Results of our runs for FV subtask

System	Macro F1	Accuracy
NUL-JA-FV-04(1st)	61.47	62.84
NUL-JA-FV-05(2nd)	59.94	61.67
NUL-JA-FV-01(3rd)	59.67	61.87
KSU-JA-FV-02	52.78	63.42
KSU-JA-FV-03	52.42	63.23
KSU-JA-FV-01	50.61	50.97

Table 4: Results of our runs for SV subtask

System	Macro F1	Accuracy
NUL-JA-SV-04(1st)	69.59	77.81
NUL-JA-SV-05(2nd)	68.94	77.96
NUL-JA-SV-01(3rd)	68.73	77.81
KSU-JA-SV-01	66.96	79.84
KSU-JA-SV-03	65.72	75.78
KSU-JA-SV-02	64.87	76.00

Unofficial Runs

After submitting the results of formal runs, errors were found in calculating some features used in SV subtask. Corrected results are shown below.

Table 5: Results of our formal and unofficial runs for SV subtask

Runs	System	Macro F1	Accuracy
formal runs (submitted)	KSU-JA-SV-01	66.96	79.84
	KSU-JA-SV-02	64.87	76.00
	KSU-JA-SV-03	65.72	75.78
unofficial runs (corrected)	KSU-JA-SV-01-C	66.01	79.48
	KSU-JA-SV-02-C	63.80	75.56
	KSU-JA-SV-03-C	67.18	76.50

Conclusion

Described the systems and results by KSU team

- In FV
  - Three systems were evaluated, each of which are based on character overlap ratios, existence of entailment result 'Y', and voting of entailment results
  - **Didn't achieve high recognition results**. Lots of work to do including features and classification methods
- In SV
  - Three systems were evaluated, each of which uses different classifiers, with surface features, numerical expressions, location and named entities features
  - **Achieved the fourth place in formal run**
  - Ablation analysis show that **surface features are still influential**
  - **Appropriate introduction of more semantic features is necessary** for further improvement

NTCIR-11 RITE-VAL
Poster ID - RITE06

The 11th NTCIR Conference, December 10-12, 2014, Tokyo, Japan

京都産業大学