CYUT Chinese Textual Entailment Recognition System for NTCIR-10 RITE-2

Shan-Shun Yang, Shih-Hung Wu
CSIE, Chaoyang University of Technology, Taiwan, R.O.C

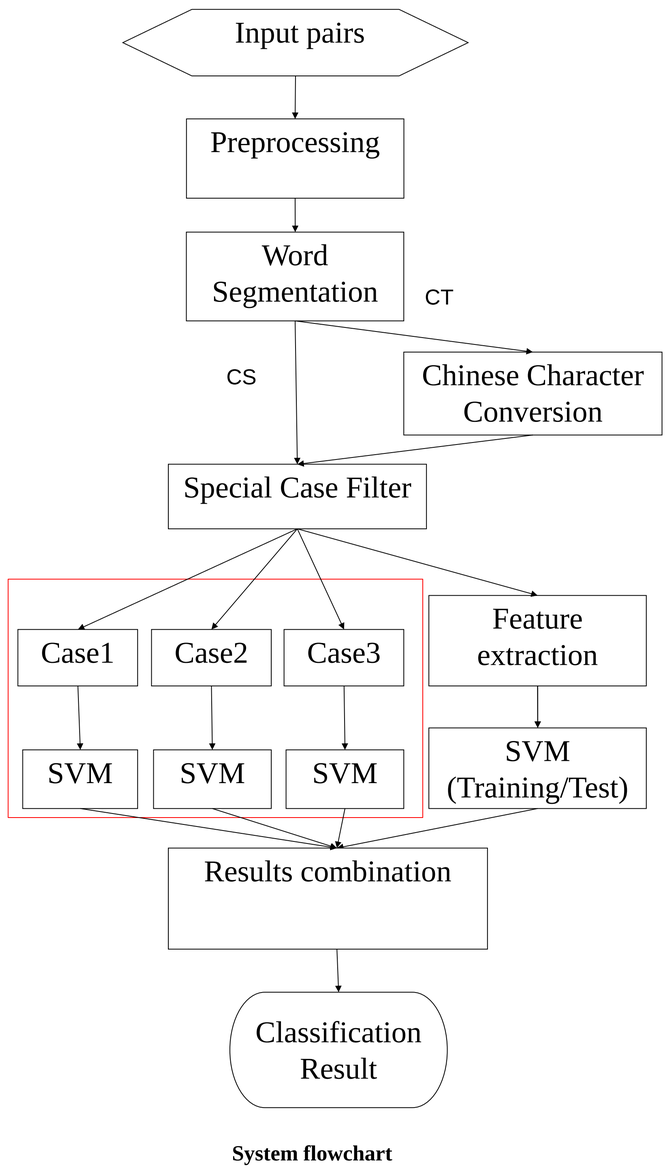
Liang-Pu Chen, Hung-Sheng Chiu,Ren-Dar Yang
Institute for Information Industry, Taipei, Taiwan, R.O.C

Abstract

Textual Entailment (TE) is a critical issue in natural language processing (NLP). In this paper we report our approach to the Chinese textual entailment and the system result on NTCIR-10 RITE-2 both simplified and traditional Chinese dataset. Our approach is based on more observation on training data and finding more types of linguistic features. The approach is a complement to the traditional machine learning approach, which treat every pair in a standard process. In the official runs, we tested three types of entailment features, i.e. the usage of negative words, time expression, and numbers. The experimental result is promising; we find this extensible approach can include more types.

System Architecture

The flowchart of our system is shown in Figure. The basic components are “preprocessing”, “word segmentation”, “Chinese character conversion”, ”special case filter”, “feature extraction”, ”sub-systems for special cases”, and “SVM” classifier



System flowchart

Features Used in Our System

No	Feature
1	unigram recall
2	unigram precision
3	unigram_F_measure
4	log_bleu_recall
5	log_bleu_precision
6	log_bleu_F_measure
7	difference in sentence length (character)
8	absolute difference in sentence length (character)
9	difference in sentence length (term)
10	absolute difference in sentence length (term)
11	Alignment score by GIZA++

Special cases in RITE-2 Chinese dataset

By observing the training corpus, we found that there are many special cases in the RITE-2 Chinese dataset. These cases are beyond the ability of our previous system. We decided to build a special case filter to pick them out and treated with special sub-systems

Case1. The only difference is a negative word

We found that there are many pairs with almost identical words; the only difference is one sentence contains a negative word, such that the entailment relation of the two sentences becomes contradiction. In some cases, the only difference is an antonym.

Case 2. Inconsistent Temporal Information

Time is important information in many sentences, either reports a particular day or a particular duration. Therefore, two sentences with inconsistent temporal information either are independent or contradict each other.

Case 3. Inconsistent Numbers

Numbers in sentences often describe the quantity of objects. Just as the previous case, inconsistent numbers might imply either independence or contradiction.

Examples of special cases

Case	Example
Case 1	T1:若望保祿二世是四百五十多年來第一位**非**義大利籍的教宗 T2:若望保祿二世是四百五十多年來第一位義大利籍的教宗
	T1: 小泉純一郎2001年贏得自民黨總裁選戰 T2:小泉純一郎2001年**未**贏得自民黨總裁選戰
	The two sentences contradict to each other.
Case 2	T1: 2003年 **4月22日**和平醫院爆發七位醫護人員集體感染 SARS, **4月24日**封院 T2 :2003年 **4月24日**和平醫院爆發七位醫護人員集體感染 SARS, **4月22 日**封院
	T1:電影《無極》在**12月12日**晚將舉行首映典禮 T2:電影《無極》在**12月11日**舉行首映典禮
	The inconsistent temporal information causes a contradiction.
Case 3	T1台、印雙邊貿易總額僅占台灣對外貿易總額**12%**，顯示雙方經貿合作還有很大的成長空間 T2台、印雙邊貿易總額僅占台灣對外貿易總額**1.2%**，顯示雙方經貿合作還有很大的成長空間
	T1:大陸「瞭望」雜誌的估算，東海油田蘊含石油**250億**噸 T2:大陸「瞭望」雜誌的估算，東海油田蘊含石油**250**噸
	The different number with the same context implies a contradiction.

The number of pairs in the three special cases

	negative word	Inconsistent Temporal Information	Inconsistent Numbers
Training Set	15(1.82%)	43(5.28%)	42(5.15%)
Test Set	42(5.37%)	60(7.68%)	83(10.62%)

In the future, we will build more sub-systems on the special cases, Special cases might include:

1.synonym

We find that synonym and near synonym in each domain are crucial for RTE. Our preprocessing module do help the performance, more information are needed to reduce the cost of computation in the later stage.

2.Background knowledge on facts

There are many cases, even human can be wrong if that person does not have enough background knowledge. For example, in news, the capital of a state is used as a pronoun of the state. Without the knowledge, it is not possible to answer correctly. The knowledge might be collect from various encyclopedia.

3.Syntactical tricks

In several cases, the only difference between t1 and t2 is the syntax. The system should knowledge which word is the subject and which term is object. A parser is needed in these cases.

4.Polysemy and named entity

In some cases, the terms are the same in the pairs, but the meaning is different. Since any term can be used as a name in different context, it is also important where the term is indicating a named entity or not.

5.Negative modifier

Adding one negative modifier into a sentence or replace one term with its anatomy will make the sentence become a contradiction. However anatomy is hard to list in various domains.

Experiment Result

CYUT-01:Extracted a set of 10 features from the training text to construct an SVM model, and use the SVM model to make decision.

CYUT-02:. Extracted a set of 11 features from the training text to construct an SVM model, and use the SVM model to make decision.

CYUT-03:Extracted a set of 11 features from the training text to construct an SVM model, and use the SVM model to make decision, An additional rule is used to pick up certain sentence pairs, which are opposite to each other.

Formal run results of systems in RITE 2 CS task

Participants	BC	MC
bcNLP	73.84	56.82
MIG	68.09	44.74
CYUT-03	67.86	40.37
CYUT-02	63.11	42.52
WHUTE	61.65	46.79
CYUT-01	61.17	40.37
IASL	60.45	50.94
IMTKUTE	54.28	23.89
Yuntech	53.52	40.89
WUST	5014	40.87
JUNLP	48.49	24.38
The paper	71.63	-

Formal run results of systems in RITE 2 CT task

Participants	BC	MC
IASL	67.14	46.32
MIG	67.07	45.15
IMTKUTE	65.99	32.36
WHUTE	65.55	45.50
Yuntech	62.31	40.14
KC99	57.67	42.16
CYUT-01	55.16	25.60
CYUT-02	52.64	26.26
CYUT-03	51.58	23.51
JUNLP	48.72	24.21
NTOUA	32.63	-
MCUIM	-	32.51
The paper	59.54	-

Additional Runs

The difference of the performances in CT and CS of our system is quite significant. We find that the character conversion is the major cause. We mentioned in special cases, a special processing unit might be design to some that special case.

Additional run results with different conversion systems

Character conversion system	BC	MC
Google Translate	53.52%	26.26%
CYUT system	59.54% (6.02%)	35.75% (9.49%)

The BC accuracy of special cases

System	negative word	Inconsistent Temporal Information	Inconsistent Numbers	Overall
Formal run	52.38%	58.33%	53.01%	67.86%
Special unit	71.42% (19.04%)	70% (11.67%)	72.28% (19.27%)	71.63% (3.77%)

Conclusions

Compared to the results of other teams , our system performs well in the CS-BC sub-task, where our best accuracy is 67.86%. However, the CT-BC sub-task, our best accuracy is much lower. In our additional run, we find that a better character conversion tool can help to boost the performance.

The treatment to the special cases did not show much improvement. We believe that special cases still need a separate process to deal with, but with more careful analysis. The additional run results show that special unit designed for special cases really help. If our analysis cover more special cases, the system can recognize textual entailment better.