FudanNLP at RITE 2011: a Shallow Semantic Approach to Textual Entailment

Ling Cao, Xipeng Qiu and Xuanjing Huang
Media Computing & Web Intelligence Lab
Department of Computer Science and Engineering
Fudan University, Shanghai, China
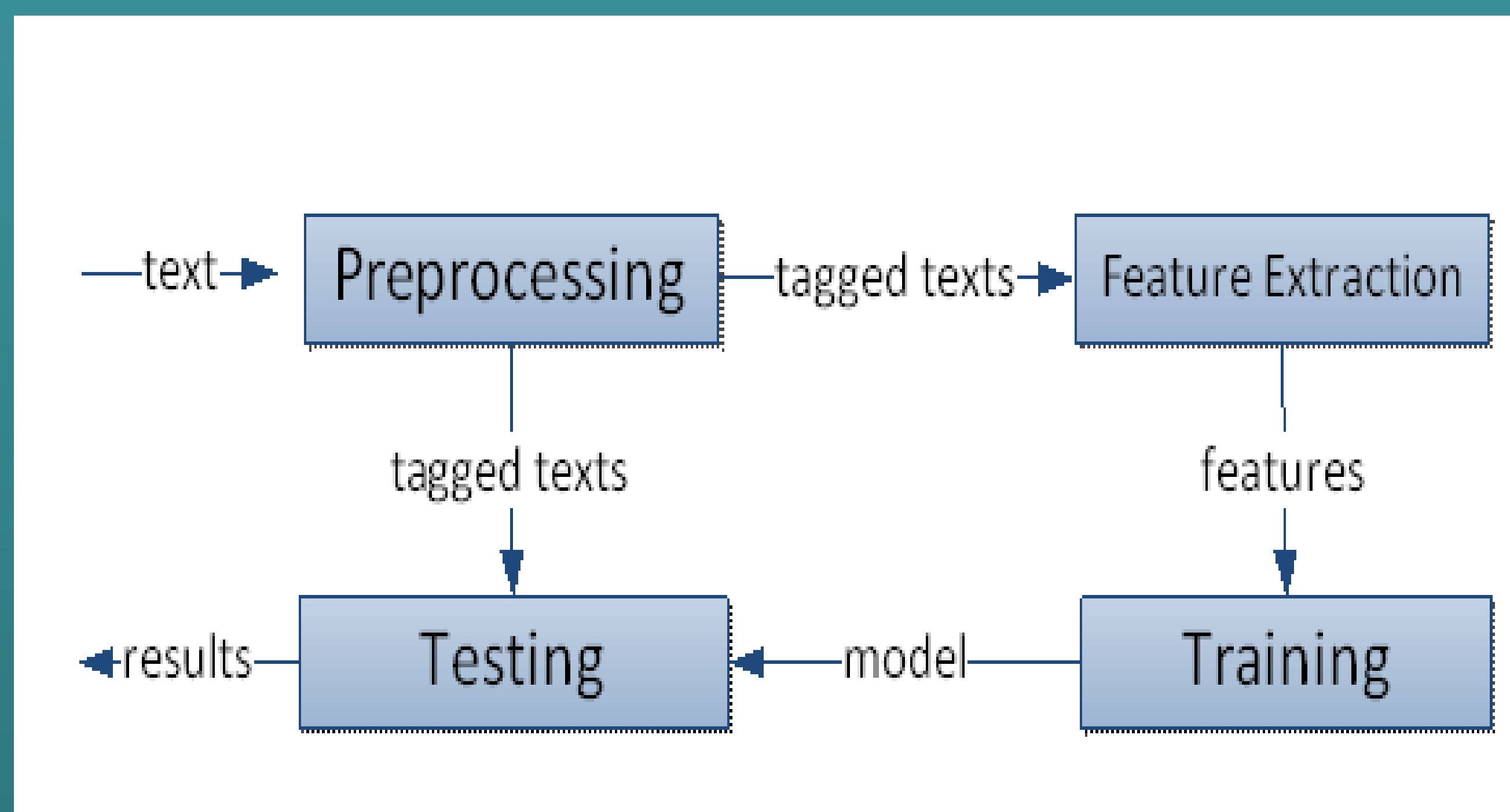11210240044,xpqiu,xjhuang@fudan.edu.cn

1. Task Description

The team of FudanNLP participated NTCIR-9 RITE 2011 Chinese simplified BC and MC subtasks. For two given text fragments:

1. BC subtask: Find whether one can be entailed by the other.
2. MC subtask: Find more detailed logic relations between texts, including forward entailment, backward entailment, bi-direction entailment, contradictory and independence.

2. System Overview

Our system is a proposal towards the resolution of the language viability problem. By using FudanNLP, our current approaches explore a method by finding common grounds between pairs with same semantic relation. With machine learning methods, system extract features from training sets on shallow semantic level.

3. Feature Selection

System select several features on shallow semantic level. The following features are constructed for use in the training of the model, and producing entailment predictions.

1. Named Entities:
   Based on the guideline of judging whether t1 entails t2 or not, t2 that introduce entities which are not mentioned by t1 indicates t2 cannot be entailed by t1, and vice versa.
2. Date & Time Information:
   If t1 and t2 happened in different periods of time, they can hardly entailed by each other.

3. Word Length & Word Overlap:
   Simplest but useful feature. If t1 is much longer than t2, it seems that t1 gives more information than t2, so t2 cannot entail t1. And if t1 and t2 are not similar enough, they may talk about different things.
4. Negation & Antonym
   The existence of negation words and antonyms always indicate contradictory between text pairs. System takes advantages of Hownet antonym dictionary to detect negation or antonym between texts

4. Experiments

System uses LibSVM for model training. We trained two different models with the only difference that feature Negation & Antonym was extracted by run 2 but not by run 1.

	Run 1	Run 2			Run 1	Run 2
BC subtask	0.746	0.76		F	0.624	0.634
MC subtask	0.58	0.585		R	0.824	0.747
				B	0.873	0.831
	Run 1	Run 2		C	0.189	0.351
Y	0.871	0.848		I	0.314	0.3
N	0.521	0.604				

7. Summary

The team of FudanNLP's system is built on machine learning framework with features selected on shallow semantic methods. System gets an accuracy as 76% on BC task and 58.5% on MC task. During experiments, it turns out that many problems cannot be solved without deep understanding of texts, future research of deep semantic understanding is required.

复 旦 大 学
Fudan University