ICRC_HITSZ at RITE: Leveraging Multiple Classifiers Voting for Textual Entailment Recognition

Yaoyun Zhang, Jun Xu, Chenlong Liu, Xiaolong Wang et al.

Key Laboratory of Network Oriented Intelligent Computation
Department of Computer Science and Technology
Shenzhen Graduate School, Harbin Institute of Technology

2011-12-08

Outline

- Introduction
- System Description
  - Preprocessing Module
  - Resource Pools
  - Feature Sets
  - Classification Module
- Evaluation and Discussion
- Summary

Introduction

- ICRC_HITSZ system participate in:
  - the binary-class (BC) subtask
  - the multi-class (MC) subtask
  - the RITE4QA subtask

on both simplified Chinese (CS) and traditional Chinese (CT) sides.

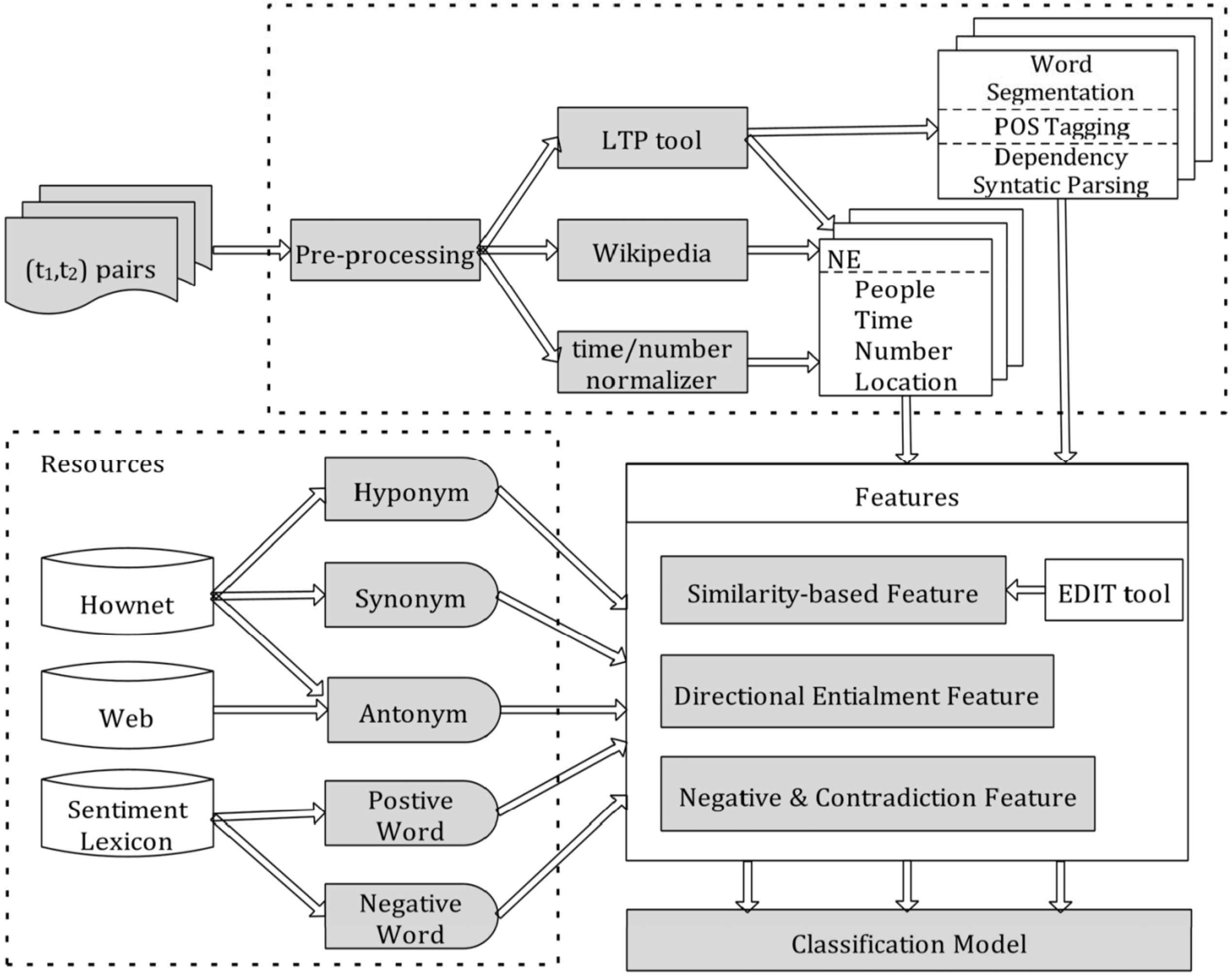
Introduction

- We build textual entailment recognition models for the MC subtask, the predicted labels are then mapped into Y/N classes for the BC and RITE4QA subtasks.
  - Extract different linguistic level features
  - Represent the problem with three different classification strategies
  - Build the recognition model with a cascade voting approach

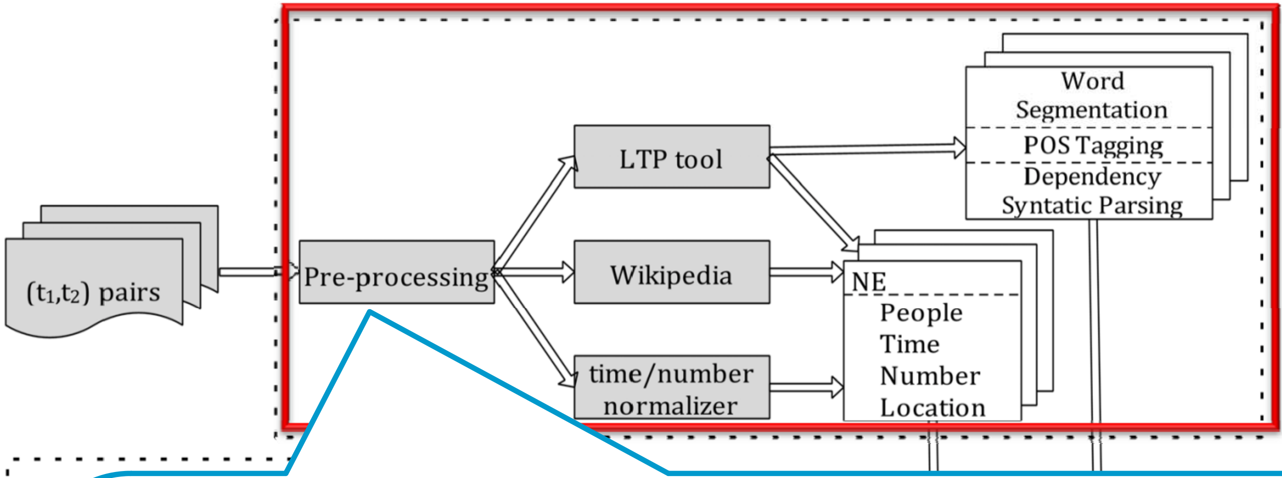
Outline

- Introduction
- System Description
  - Preprocessing Module
  - Resource Pools
  - Feature Sets
  - Classification Module
- Evaluation and Discussion
- Summary

System Architecture



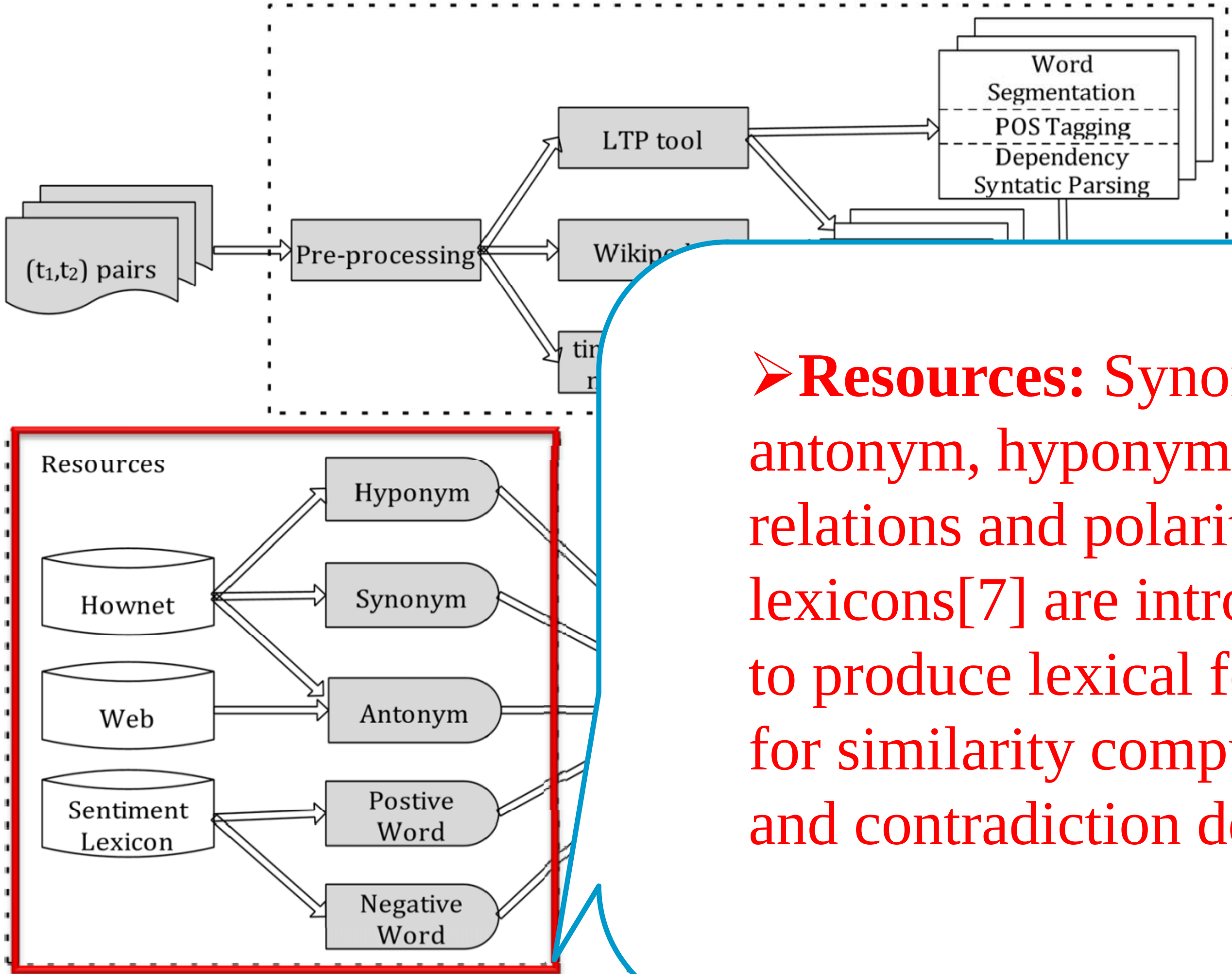
System Architecture



Pre-processing Module: sentence parsing and NE recognition etc.

- ❖Each <t1, t2> pair is first preprocessed by the LTP tool[3], for word segmentation, POS tagging, dependency syntactic parsing and named entity (NE) recognition.
- ❖Extract the page titles of Chinese Wikipedia to expand the NE lexicon of LTP.
- ❖A number/time normalization module is deployed to unify various formations of numbers/times.

System Architecture



➢**Resources:** Synonym, antonym, hyponym relations and polarity lexicons[7] are introduced to produce lexical features for similarity computation and contradiction detection.

1:http://www.keenage.comhttp://download.wikimedia.org/zhwiki
2:http://fyc.5156edu.com/

System Architecture

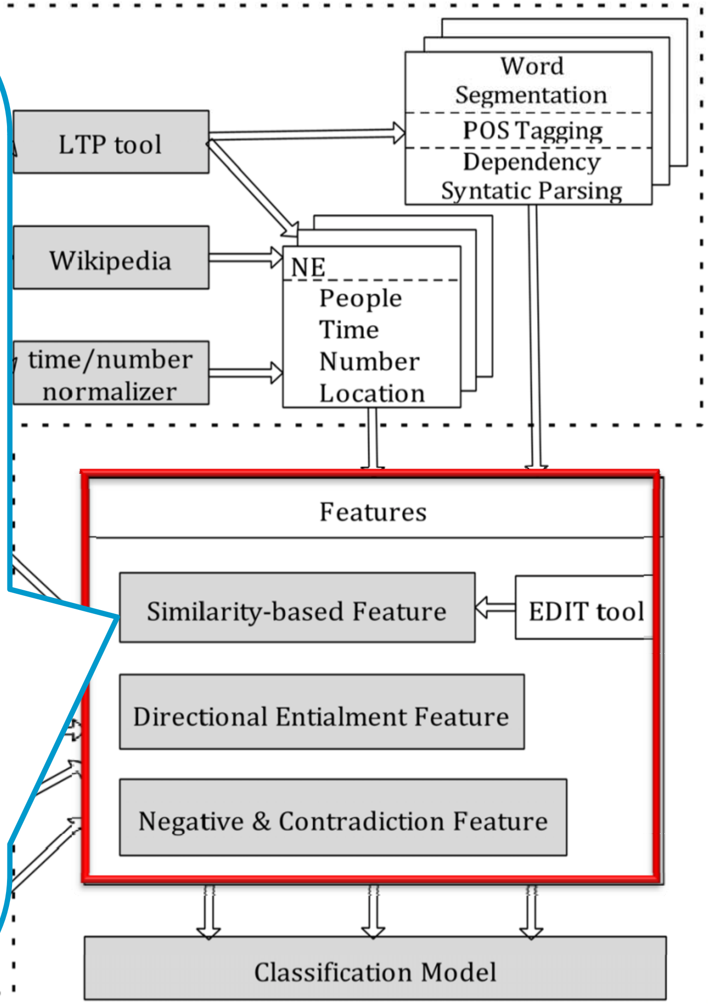
Similarity-based Features:

The open source package EDIT[4] is employed to generate similarity based features.

- ➢Edit distance
- ➢Word overlap
- ➢Cosine similarity etc

Three types of presentations of *t1 and t2* as input[5]:

- ➢*Tokens*
- ➢*POS tagging*
- ➢Subject-Verb-Object structures



System Architecture

Directional Entailment Features: indicating the entailment direction between *t1 and t2.*

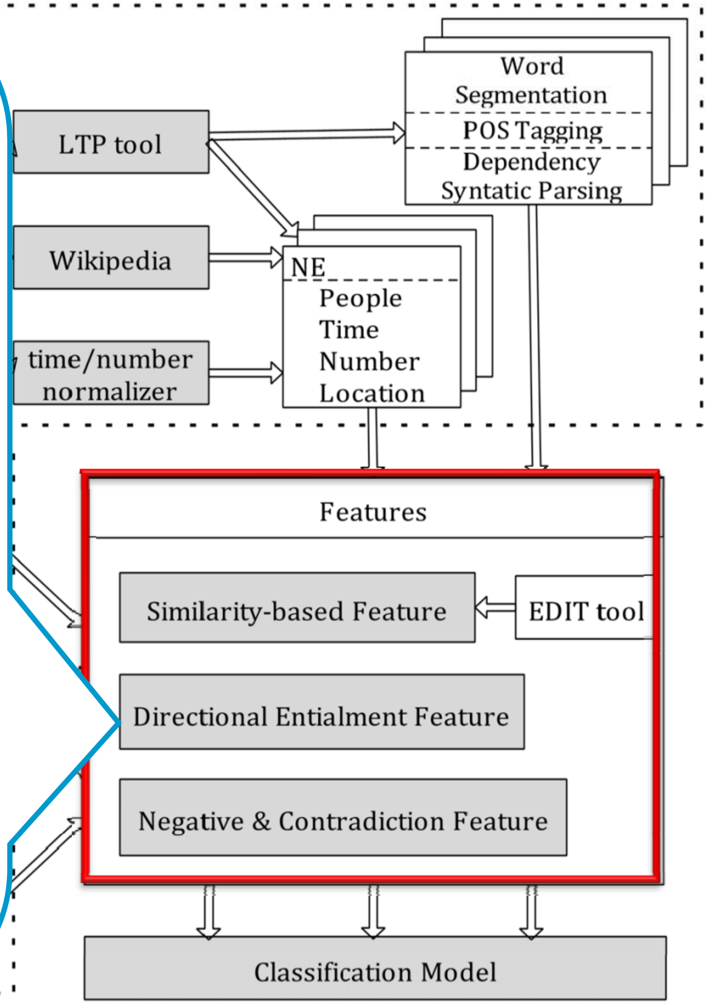
➢the proportions of equal linguistic units in *t1* or *t2*;

➢whether a specific type of NE only appears in *t1* or *t2*;

➢ entailment between *t1* and *t2* in different linguistic granularities, ranging from word to syntactic structures;

➢ whether there is any equivalent definitional (mostly is-a) relation between contents of *t1* and *t2*.

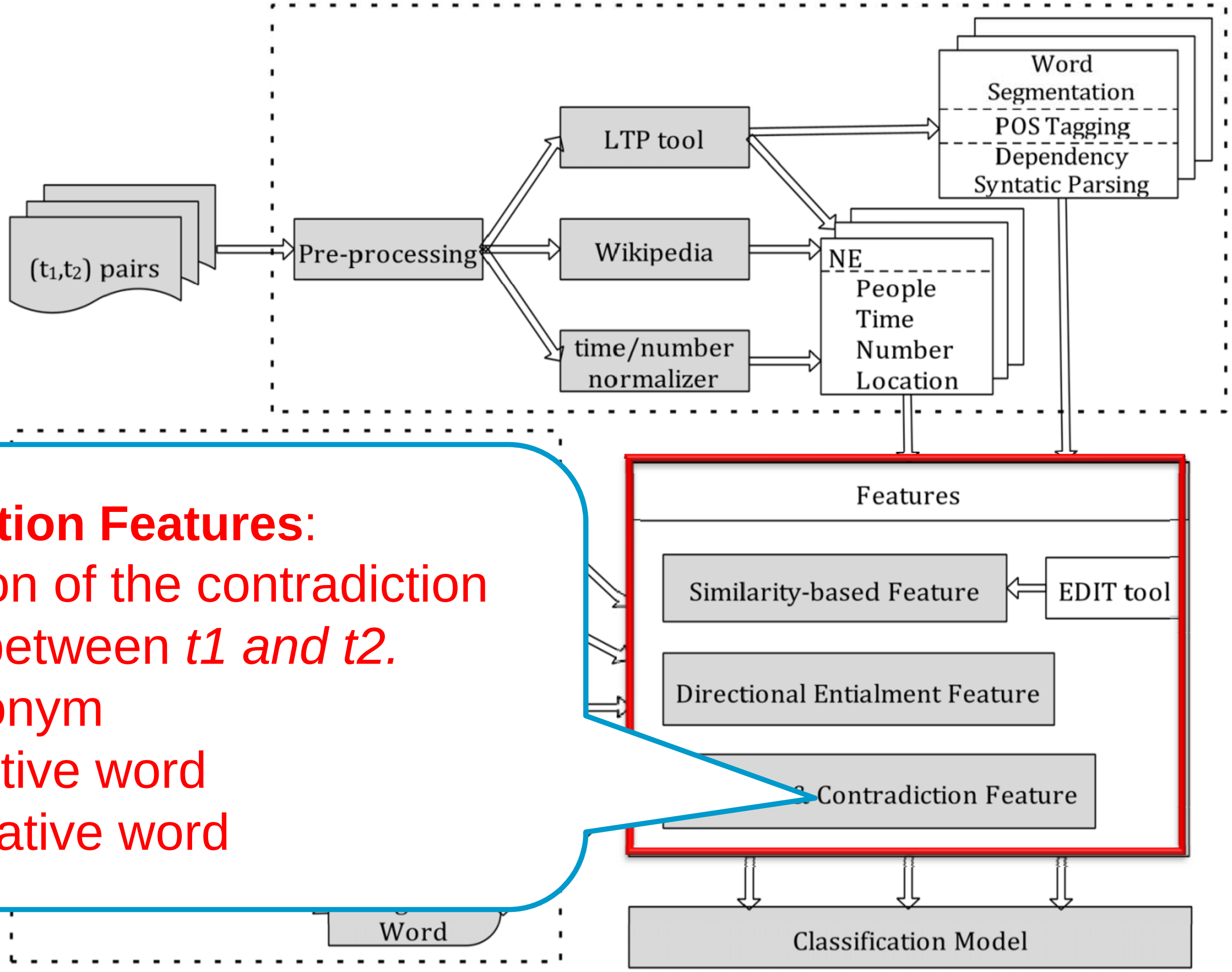
- B A->A is B



Directional Entailment Features in Detail

Type1: Number Proportion of Equal Linguistic Unit in T /H
Sentence length in terms of word numbers
Number of equal NEs
Number of equal content words
Number of equal nouns
Number of equal numbers
Number of equal times
Number of equal locations
Numbers of equal Sub_Verb_Obj Structures
Type2: NE's Existence in T or H
Numbers exist in T/H, but not in H/T
Times exist in T/H, but not in H/T
Locations exist in T/H, but not in H/T
Type3:Entailment of Different Linguistic Granularity
Words in T/H are hyponyms of words in H/T
If two words, A and B, are the same, whether A/B's modifier is the substring of the other'.
Whether a number from T/H can be entailed by a number from H/T
Whether a time from T/H can be entailed by a time from H/T
Whether a location from T/H can be entailed by a location from H/T
Whether a person from T/H can be entailed by a person from H/T
Whether a Sub_V_Obj Structure from T/H can be entailed by a Sub_V_Obj Structure from H/T
Whether a Sub_V_Obj Structure which has dependency relations with a NE from T/H can be entailed by a Sub_V_Obj Structure which has dependency relations with the same NE from H/T
Type4:Definitional Feature
In T/H, A is the attribute modifier of B, and H/T can be considered as representing a B is_a A relation. A and B can be a word or a phrase. The is_a relations are recognized by matching several simple syntactic patterns.
Both T and H can be considered as representing an A is_a B relation.

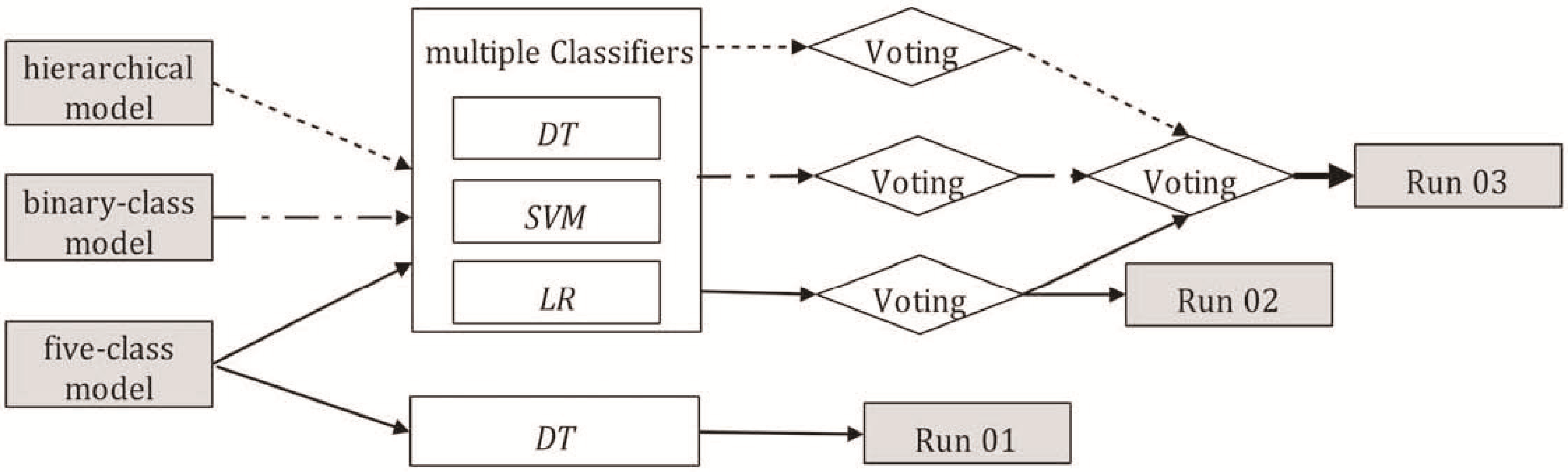
System Architecture



Contradiction Features:
Recognition of the contradiction relations between *t1 and t2.*

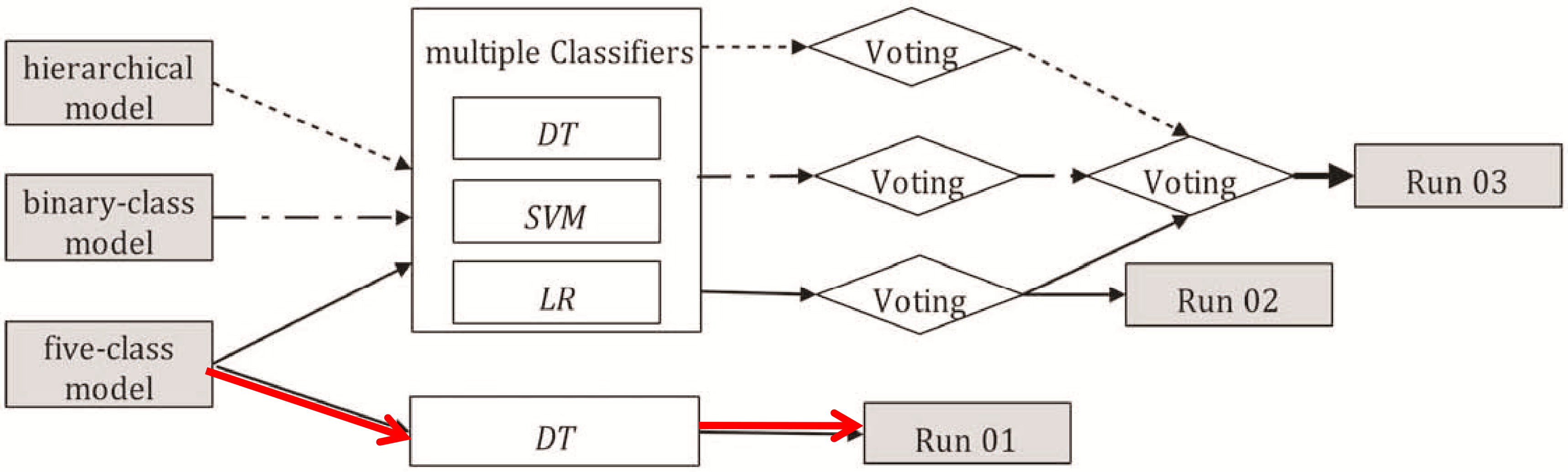
- Antonym
- Positive word
- Negative word

Classification Module



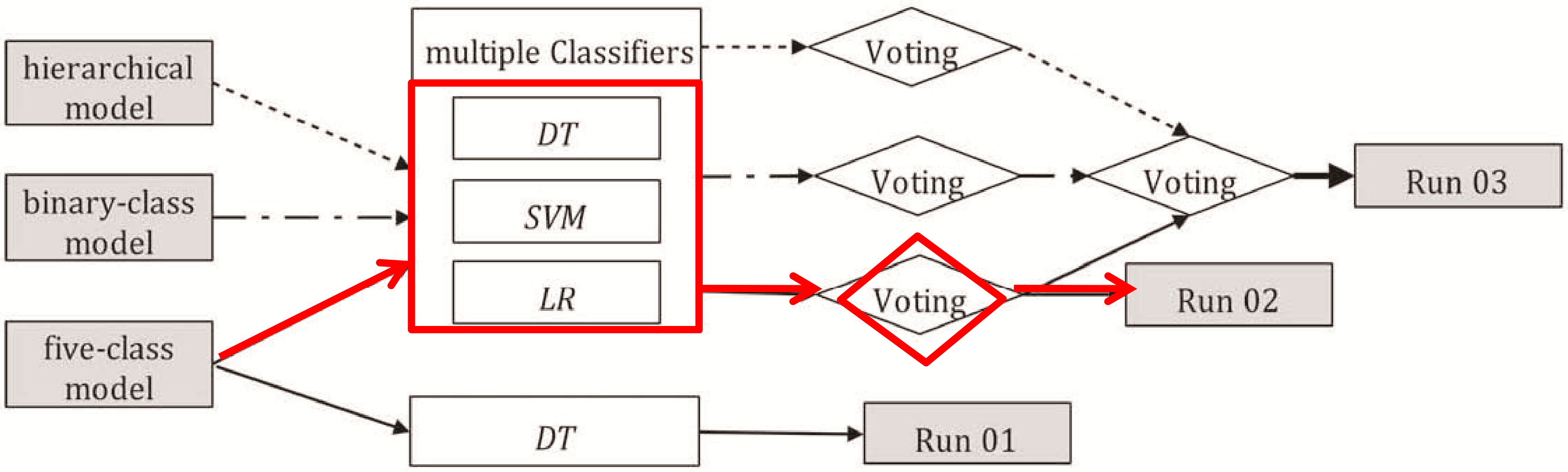
Assumption: classifiers built from different classification strategies are complementary to each other, so are the different machine learning methods.

Classification Module



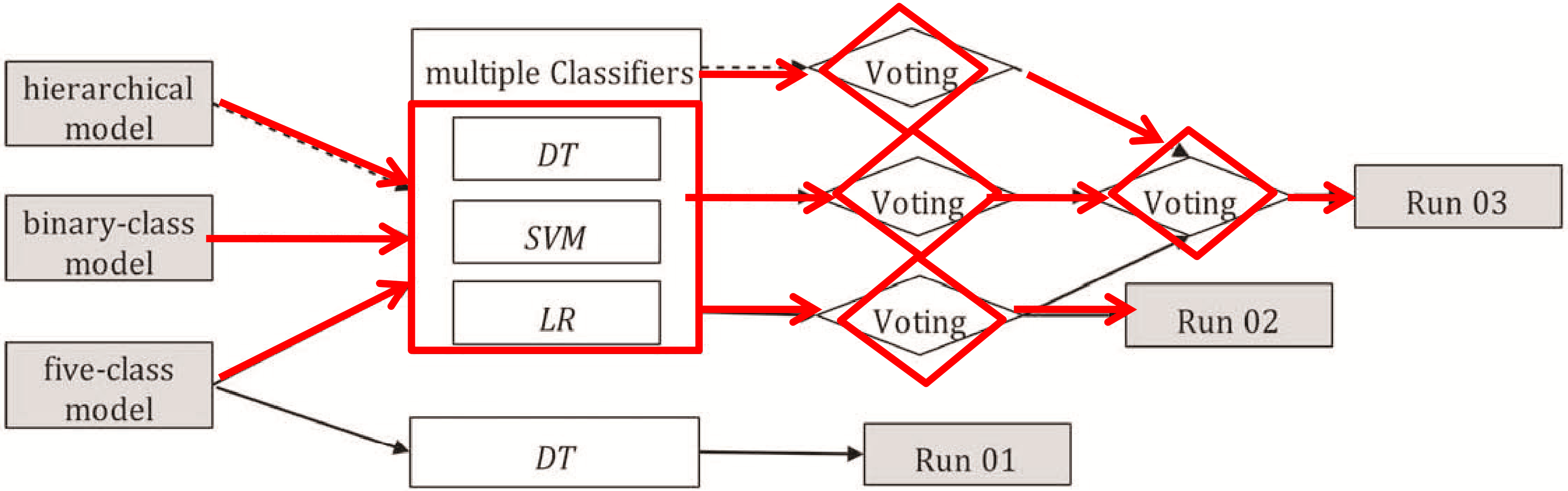
➢Run01: A five-class classifier, which is built by using Decision Tree.

Classification Module



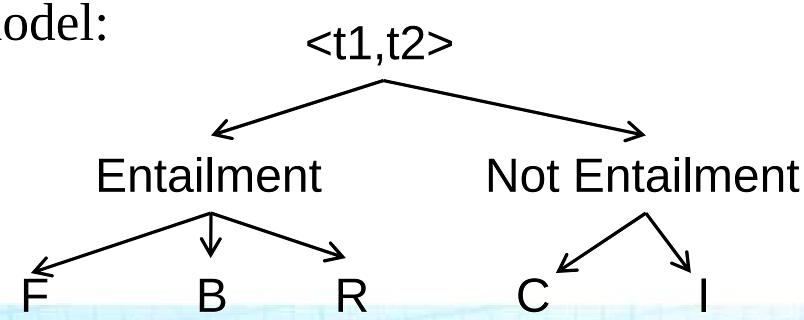
➢Run02: Voting among three five-class classifiers built from Decision Tree, SVM and Logistic Regression.

Classification Module



➢Run03: Cascade Voting.

- Binary-class Model: five binary-class classifiers for each label.
- Hierarchical model:



Outline

- Introduction
- System Description
  - Preprocessing Module
  - Resource Pools
  - Feature Sets
  - Classification Module
- Evaluation and Discussion
- Summary

Experimental Results

❖ Evaluation results of BC and MC subtask in CS

CS	BC	MC
run01	0.708	0.575
run02	0.757	0.624
run03	0.776	0.641

run02 enhances the accuracy for 6.92% and 8.52% from run01, while run03 further enhances the accuracy for 2.51% and 2.72% from run02. The accuracy enhancement of employing the voting of different problem representations is not as high as the voting of different ML methods.

Experimental Results

❖ Evaluation results of BC and MC subtask in CT

CT	BC	MC
run01	0.613	0.497
run02	0.597	

Run01_CT_BC: run01
Run02_CT_BC: run02
Run01_CT_MC: run02

The model developed for CS is applied directly to CT. Errors are mainly caused in the word segmentation and pos-tagging stages, because the LTP tool has the difficulty to process CT, especially CT NE recognition.

Eg.车诺比尔病毒(CT)=切尔诺贝利病毒(CS)
(Chernobyl, #716 in CT BC test set)

Experimental Results

❖ Evaluation results of RITE4QA subtask in CS and CT

CS&CT	TOP1	MRR
run01*	0.2479	0.3520
run02*	0.2234	0.2705
run03*	0.2262	0.3398

run01* : the same as run02, with N-class biased voting strategy
run02*: recognition model using SVM
run03*: the same as run02, with N-class biased voting strategy and dynamically updated lexicon

N-class biased voting: if one model outputs ‘N’, the final class is ‘N’.
Dynamic lexicon updating: a maximum-common-string matching is conducted between *t1* and *t2*, the matched strings are dynamically added into the lexicon as proper nouns, which are proved to introduces more noises and affects the performance.

Outline

- Introduction
- System Description
  - Preprocessing Module
  - Resource Pools
  - Feature Sets
  - Classification Module
- Evaluation and Discussion
- Summary

Summary

- Different linguistic level features, the cascade voting combining multiple classifiers and multiple problem representations are effective for RITE challenge. But there is still much room left for further improvement
- World knowledge Inference

  Eg. a mother should be female (#21 in CS test set)
- Co-reference Resolution
- Number/time Relation Inference

  Eg. the relation between a moment and a time interval
- Cause-Result Relation Inference

  Eg. A leads to B-> B is related to A (#13 of CS test set)

Reference

- [1] I. Androutsopoulos and P. Malakasiotis. A survey of paraphrasing and textual entailment methods. Journal of Artificial Intelligence Research., 38(1):135–187, May 2010.
- [2] L. Bentivogli, I. Dagan, H. T. Dang, D. Giampiccolo, and B. Magnini. The fifth pascal recognizing textual entailment challenge. In Proceedings of TAC'2009, 2009.
- [3] W. Che, Z. Li, and T. Liu. Ltp: A chinese language technology platform. In COLING (Demos), pages 13–16, 2010.
- [4] M. Kouylekov and M. Negri. An open-source package for recognizing textual entailment. In ACL (System Demonstrations), pages 42–47, 2010.
- [5] P. Malakasiotis and I. Androutsopoulos. Learning textual entailment using svms and string similarity measures. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, RTE’07, pages 42–47, 2007.
- [6] H. Shima, H. Kanayama, C.-W. Lee, C.-J. Lin, T. Mitamura, S. S. Y. Miyao, and K. Takeda. Overview of ntcir-9 rite: Recognizing inference in text. In NTCIR-9 Proceedings, to appear, 2011.
- [7] R. Xu and C. Kit. Incorporating feature-based and similarity-based opinion mining - ctl in ntcir-8 moat. In Proceedings of NTCIR-8 Workshop Meeting, pages 13–16, 2010.

Thanks for your attention

Q&A