ICRC_HITSZ:Leveraging Multiple Classifiers Voting for Textual Entailment Recognition

Yaoyun Zhang, Jun Xu, Chenlong Liu, Xiaolong Wang, Ruifeng Xu, Qingcai Chen, Xuan Wang, Yongshuai Hou and Buzhou Tang

Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, P.R. China

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

智能计算研究中心 Intelligent Computing Research Center

Introduction

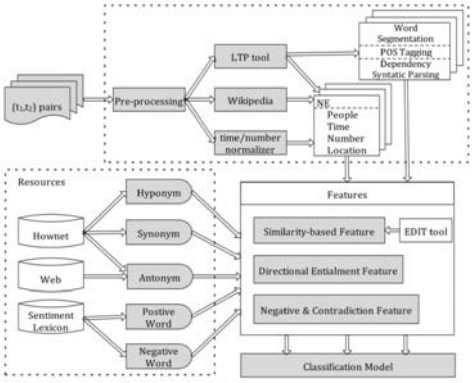
➢This paper presents the ICRC_HITSZ system in the NTCIR-9 RITE challenge. We participate in
- ❖the binary-class (BC) subtask,
- ❖the multi-class (MC) subtask,
- ❖the RITE4QA subtask

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

➢The textual entailment recognition models are built for the MC subtask, then the predicted labels are 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

System Description

➢**System Architecture**



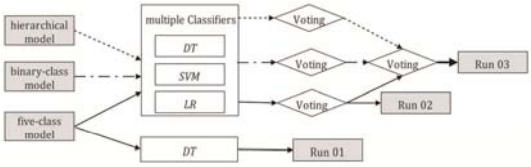
❖**Preprocessing Module** for sentence parsing and NE recognition.
- ◆Preprocessed each <t1, t2> pairs by the LTP tool
- ◆ Expand the terminology lexicon of LTP by Chinese Wikipedia
- ◆Number/time normalization

❖**Resource Pools** for lexicon relationship recognition
- ◆Synonym
- ◆Antonym
- ◆Hyponym
- ◆Polarity

❖**Features Extraction**

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.

❖**Classification Module**



◆***run01***: Using Decision Tree(*DT*) to build a five-class classifier

◆***run02***: Voting among three five-class classifiers: *DT*, Support Vector Machine(*SVM*) and Logistic Regression(*LR*)

◆***run03***: Voting among three classifiers using different class taxonomies and machine learning algorithms(*DT, SVM, LR*)

Evaluation

❖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.

❖Evaluation results of BC and MC subtask in CT

CT	BC	MC
run01	0.613	0.497
run02	0.597	

Errors are mainly caused by word segmentation and pos-tagging, because the LTP tool has the difficulty to process CT, especially CT NE recognition.

❖Evaluation results of RITE4QA subtask in CS and CT

CS&CT	TOP1	MRR5
*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 t_1 and t_2, the matched strings are dynamically added into the lexicon as proper nouns, which are proved to introduces more noises and affects the performance.

Conclusion

Different linguistic level features, multiple classifiers cascade voting and multiple problem representations are effective for RITE challenge. But there Is still much room left for further improvement, such as world knowledge inference , co-reference resolution etc.

References

[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.