

# ZSWSL Text Entailment Recognizing System at NTCIR-9 RITE Task

Ranxu Su, Sheng Shang, Pan Wang, Haixu Liu, Yan Zheng

School of Computer Science, Beijing University of Posts and Telecommunications, 100876, Beijing, China  
suranxu.bupt@gmail.com

## Challenges in NTCIR RITE Task

Note: Large ratio of common words/syntactic structures doesn't guarantee the same semantic meaning. Traditional features based on lexical words, synonyms or Semantic Role Labeling met trouble when dealing with these non-entailment cases.

S1: 2004 年刘翔在雅典奥运会男子 **110 米栏**决赛上**首度**获得冠军。

- （S1: Xiang Liu won his **first** gold medal prize on men's **110 meter hurdles** at Athens Olympics in 2004）

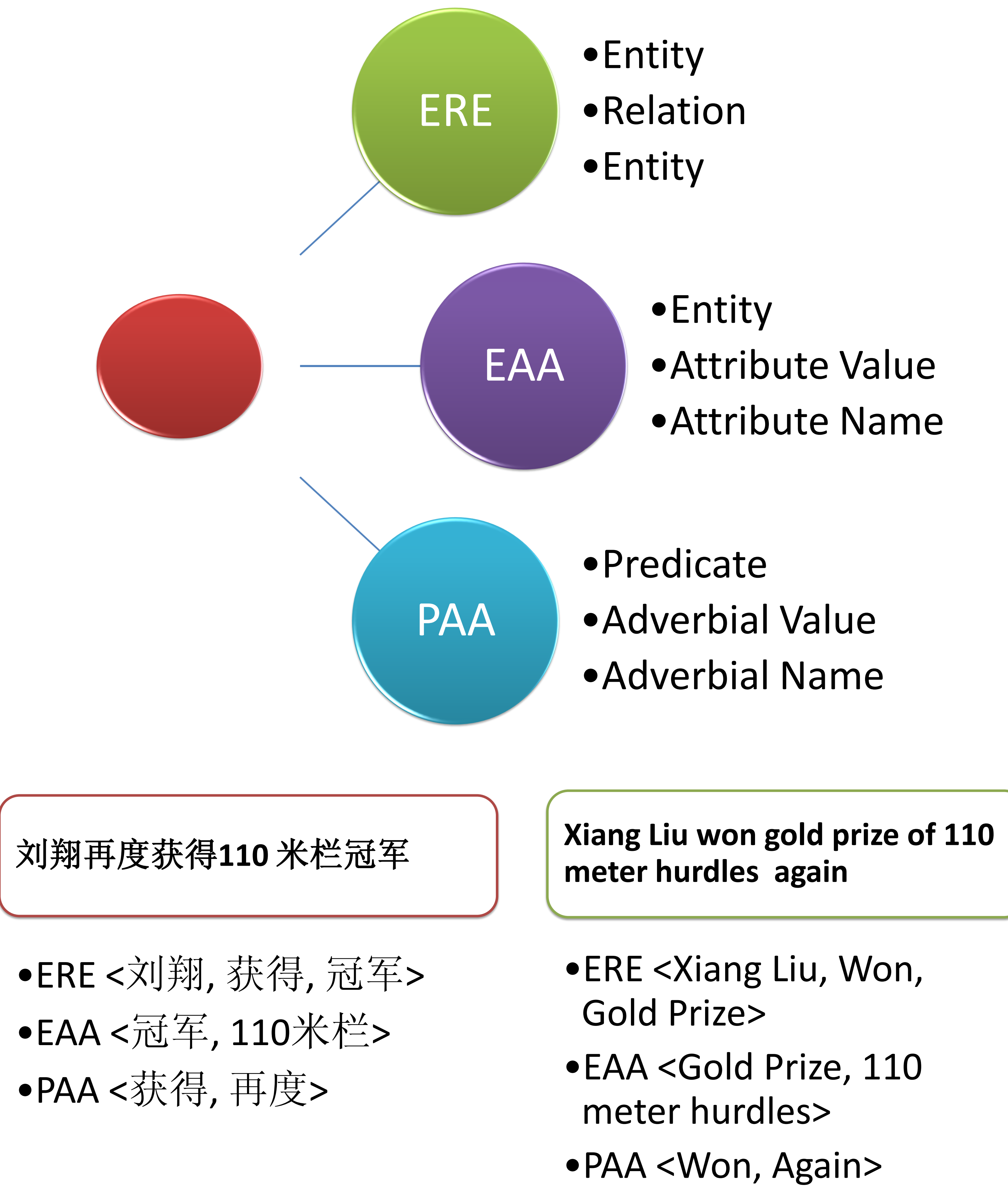
S2: 2004 年刘翔在雅典奥运会男子 **110 米栏**决赛上**再度**获得冠军。

- （S2: Xiang Liu won his **second** gold medal prize on men's **110 meter hurdles** at Athens Olympics in 2004）

S3: 2004 年刘翔在雅典奥运会男子**马拉松**决赛上**首度**获得冠军。

- （S3: Xiang Liu won his **first** gold medal prize on men's **Marathon Final** at Athens Olympics in 2004）

## Solution - Semantic Feature Extraction

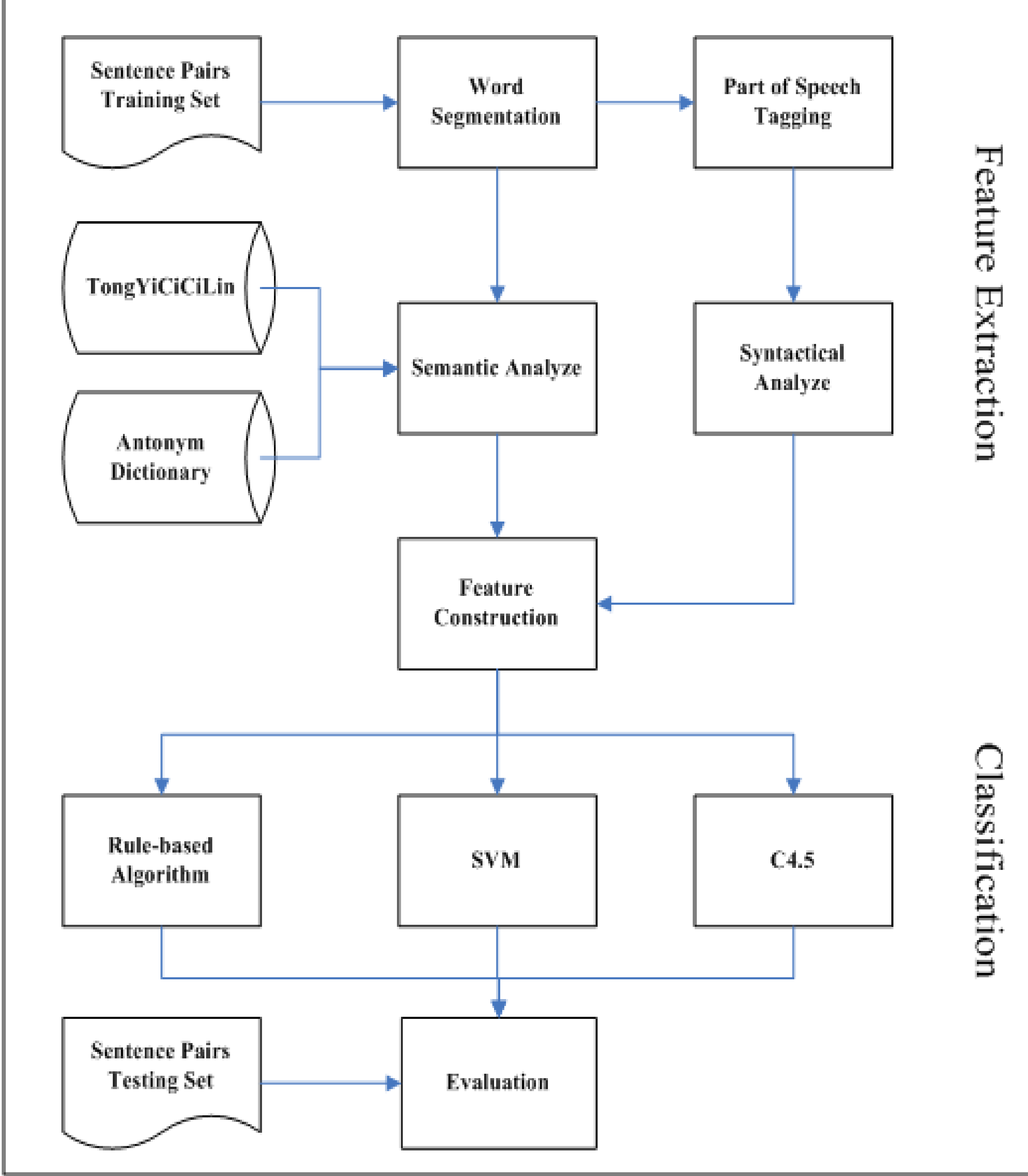


Sentences are compared by semantic structures, not lexical words, to generate semantic features.

## Machine Learning Algorithms Selection

| Models                                  | Precision |
|---|-----------|
| rule-based model                        | 46.1%     |
| SVM model                               | 46.0%     |
| C4.5 Decision Tree Model (perform best) | 55.9%     |

## System Architecture



## RITE Formal Run

| Sub-Task | Accuracy | Team Rank |
|----------|----------|-----------|
| BC       | 72.0%    | 7/12      |
| MC       | 61.9%    | 3/11      |

## Conclusion

We introduced a novel semantic feature extraction method. Three different models were applied and compared based on these features, where C4.5 outperform both rule-based algorithm and SVM. Evaluation result showed a good accuracy of 72.0% in BC sub task and 61.9% in MC subtask.