

An Empirical Study on Pronoun Resolution in Chinese*

Wang Houfeng and Mei Zheng

Department of Computer Science and Technology
School of Electronic Engineering and Computer Science
Peking University, Beijing, 100871, China
{wanghf,kevinmei}@pku.edu.cn

Abstract. In this paper, we discuss how to identify three important features by our empirical observation — gender and number features of antecedent as well as grammatical role of personal pronoun, which have no overt mark in Chinese. Only a tagger with extended POS set and some special word-lists are used. Finally, We describe an implemented prototypical system to resolve personal pronouns. Evaluation shows that the result is satisfactory.

1 Introduction

Anaphora resolution is a very important task for NLP applications like Text summarization, Information Extraction, Multilingual Information Retrieval and so forth. In the past 10 years, researchers have made great advance in this field. Various approaches, such as linguistic knowledge based methods [5], Machine Learning methods [2][6] and robust methods with less language knowledge or even without parsing [1][3], were proposed. However, there are a number of issues remaining unsolved in anaphora resolution [4] and the precision and the recall are still low.

Unlike in some Euro-languages, the grammatical role of personal pronoun in Chinese, which is commonly thought as an important factor, has no overt mark; in addition, the gender and number features of potential antecedents are unclear. In this paper, we present an inexpensive approach to solving these problems and resolving personal pronoun based on our empirical observation.

Our approach makes use of a tagger with an extended set of POS, plus simple heuristic rules, and operates on basic agreements and preferences to resolve anaphora. The evaluation shows that our approach reaches a satisfactory result.

2 Approach

Personal pronoun resolution relies on the constraints between pronouns and antecedents. Among them, the grammatical roles, gender and number of pronouns and

* This work is funded by National Natural Science Foundation of Chinese (No. 60173005)