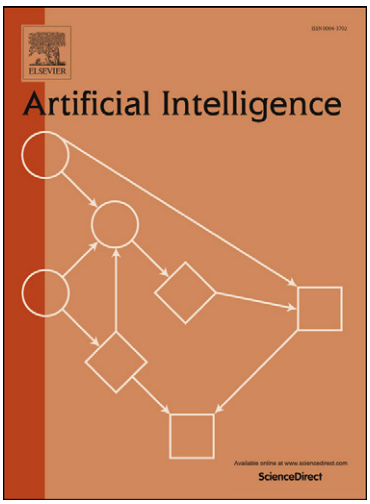
Accepted Manuscript



Feature assembly method for extracting relations in Chinese

Yanping Chen, Qinghua Zheng, Ping Chen

PII: S0004-3702(15)00104-6
DOI: http://dx.doi.org/10.1016/j.artint.2015.07.003
Reference: ARTINT 2867

To appear in: *Artificial Intelligence*

Received date: 5 February 2015
Revised date: 5 July 2015
Accepted date: 7 July 2015

Please cite this article in press as: Y. Chen et al., Feature assembly method for extracting relations in Chinese, *Artificial Intelligence* (2015), http://dx.doi.org/10.1016/j.artint.2015.07.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Feature Assembly Method for Extracting Relations in Chinese

Yanping Chen[1,*], Qinghua Zheng[1], Ping Chen[2]

Abstract

The goal of relation extraction is to detect relations between two entities in free text. In a sentence, a relation instance usually comprises a small number of words, which yields a sparse feature representation. To make better use of limited information in a relation instance, parsing trees and combined features are employed widely to capture the local dependencies of relation instances. However, the performance of parsing tree-based systems is often degraded by chunking or parsing errors. Combined features are used widely, but few studies have addressed how features can be combined to achieve optimal performance. Thus, in this study, we propose a feature assembly method for relation extraction. Six types of candidate features (head noun, POS tag, n-gram, omni-word, etc.) are employed as atomic features and six constraint conditions (singleton, position, syntax, etc.) are used to combine these features in different settings. Depending on the utilization of candidate features, different constraint conditions can be explored to achieve the optimal extraction performance. Our method is effective for capturing local dependencies and it reduces the errors caused by inaccurate parsing. We tested the proposed method using the ACE 2005 Chinese and English corpora, and it achieved state-of-the-art performance, where it was significantly superior to existing methods.

Keywords:

[*]No.28, Xianning West Road, Xi'an, Shaanxi, 710049, P.R. China; Tel: +86-29-82665191-814

Email addresses: ypench@gmail.com (Yanping Chen), qhzheng@mail.xjtu.edu.cn (Qinghua Zheng), ping.chen@umb.edu (Ping Chen)

[1]MOEKLINNS Lab, Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

[2]Department of Computer Science, University of Massachusetts Boston, Boston, USA

Download English Version:

https://daneshyari.com/en/article/6853203

Download Persian Version:

https://daneshyari.com/article/6853203

Daneshyari.com