

---

# 解决汉语自动分词的方法选择

曲维光

(南京师范大学数学与计算机科学学院 210097)

(南京师范大学文学院 210097)

E-mail: wgqu@pine.njnu.edu.cn

**摘要:** 本文分析了目前解决汉语自动分词问题的主流方法,如概率方法和机械匹配方法,指出它们存在的方法论的缺陷,虽然目前可以达到很高的识别率,但无法从根本上解决分词问题。通过对智能型方法的讨论,分析了基于专家系统方法和基于人工神经网络方法各自的优点和需要改进的问题以及解决思路,提出用文本挖掘的方法从大规模语料库中提取专家系统的规则,用概率方法提高处理速度,用神经网络,通过使用高层次语言知识来处理歧义切分,用机器学习方法来处理未登录词。今后的工作中将在该模型的指导下进行系统实现,以期达到预期的效果

**关键词:** 汉语自动分词,歧义切分,未登录词

**作者简介:** 曲维光(1964—),男,山东省牟平人,2002届博士生,讲师,主要研究领域为人工智能,计算语言学。

## The Selection of Methodology on Automatic Words Segmentation of Chinese

Qu Weiguang

(School of Mathematics and Computer, Nanjing Normal University, Nanjing, 210097)

(School of Literature, Nanjing Normal University, Nanjing, 210097)

E-mail: wgqu@pine.njnu.edu.cn

**Abstract:** Main trend methods currently used in automatic words segmentation, such as statistic method and mechanical matching method, are discussed. The shortcomings in methodology, which prevent from the ultimate solution, are showed in this paper. From the discussion of the methods of artificial intelligence and the analysis of expert system based and artificial neural networks based methods, the pro and con are presented. The solution scenarios are also draft, which will be suitable for the ultimate solution to words segmentation.

**Key words:** automatic Chinese word segmentation, ambiguity processing, unlisted words