

Title	参照解析と法令質問応答への適用
Author(s)	Tran, Thi Oanh
Citation	
Issue Date	2014-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/12109
Rights	
Description	Supervisor: 島津 明, 情報科学研究科, 博士

Reference Resolution and Its Application to Legal Question Answering

by

Tran Thi Oanh (1120008)

School of Information Science

Japan Advanced Institute of Science and Technology

March, 2014

Abstract

Natural languages are highly related by references within them. These references bring precious information: the sentences of a discourse could not be interpreted without knowing who or what entity is being talked about. Resolving resolution, therefore, is a very important task in natural language processing research. Of all reference phenomena, the coreference is the most popular phenomenon, and is attracting much research in reference resolution. In this dissertation, we will concentrate on this challenging task - *coreference resolution in general texts*. Moreover, we will also focus on *resolving references in a specific type of texts, i.e. legal texts*. The information on reference resolution not only helps people in understanding texts, but also supports other tasks such as question answering, text summarization, and machine translation. To illustrate one of these benefits, in this thesis, we will also investigate *an application of reference resolution to the task of question answering restricted to the legal domain*.

Most previous research proposed a pairwise approach to solve the task of coreference resolution. The drawback of this approach is that it can allow only one or two antecedent candidates to be considered simultaneously. So, it only determines how good a candidate is relative to the mention, but not how good a candidate is relative to all candidates. Our goal is to investigate another approach which can address this drawback. While coreference resolution in general texts attracts much attention among researchers, the task in legal texts has received very little attention so far. The main reasons are mostly the complex and long legal structures and sentences, specific terms, and especially the lack of language resources (i.e. annotated corpora) in this specific domain. Focusing on this interesting legal domain, this dissertation also aims at building a system which can automatically extract referents for references in real time. This is a new interesting task in the Legal Engineering research. Moreover, the goal of this dissertation also includes building an application of these reference resolvers to a useful question answering system restricted to the legal domain. Particularly, the following three problems are targeted in

this research:

- To realize coreference resolution in general texts, we present an empirical study on a listwise, which can address the drawback of the previous approach. This approach exploits a listwise learning-to-rank method which considers all antecedent candidates simultaneously, not only in the resolution phase but also in the training phase. Experimental results on the corpora of SemEval-2010 shared task 1 show that the proposed system yields a good performance in multiple languages when compared to previous participating systems as well as a baseline pairwise system using the ranking support vector machine as the learning algorithm. In comparison to the best participating system SUCRE, which uses the Decision Tree algorithm with best-first clustering strategy, the proposed system achieves comparative performance.
- For the task of reference resolution in legal texts, different from previous work that only considered the referent at the document targets, this work focuses on resolving references to the sub-document targets. Referents extracted are the smallest fragments of texts in documents, rather than the entire documents that contain the referenced texts. Based on the structures of references in legal texts, we propose a four-step framework to accomplish the task: mention detection, contextual information extraction, antecedent candidate generation, and antecedent determination. We also show how machine learning methods can be exploited in each step. The final system achieves 80.06% in the F1 score for detecting references, 85.61% accuracy for resolving them, and 67.02% in the F1 score on the end-to-end setting task on the Japanese National Pension Law corpus.
- This dissertation also presents a study aimed at exploiting reference information to build a question answering system restricted to the legal domain. Most previous research focuses on answering legal questions whose answers can be found in one document¹ without using reference information. However, there exist many legal questions, which require answers extracted from connections of more than one document. The connections between documents are represented by explicit or implicit references. To the best of our knowledge, this type of questions is not adequately considered in previous works. To cope with them, we propose a novel approach which allows to exploit the reference information between legal documents to find answers to these legal questions. This approach also uses the requisite-effectuation structures of legal sentences and some effective similarity measures based on legal terms to support finding correct answers without training data.

The contribution of this dissertation includes linguistic and computational aspects. Considering the linguistic viewpoint, our research helps in interpreting the sentences of any discourse. In the computational viewpoint, our research proposes effective solutions for linguistic problems using machine learning approaches.

¹The term ‘*documents*’ corresponds to articles, paragraphs, items, or sub-items according to the naming rules used in the legal domain.

Keywords: reference resolution, coreference resolution, legal texts, question answering, pairwise approach, listwise approach, learning-to-rank, logical structure, requisite-effectuation structures, mention detection, JNPL corpus.