CITY UNIVERSITY OF HONG KONG

香港城市大學

Semantic Pattern for Question Answering System

基於語義模板的問答系統研究

Submitted to
Department of Computer Science
電腦科學系
in Partial Fulfillment of the Requirements
for the Degree of Philosophy as appropriate
哲學博士學位

by

HAO Tianyong
郝 天 永

July 2010
二零一零年七月

Abstract

With the dramatic development of the Internet and the emergence of Web 2.0, User-Interactive Question Answering (UIQA) systems have been developed and become very popular Web-based services. Unlike the traditional automatic Question Answering (QA) systems which obtain answers automatically, the User-Interactive QA systems serve as interactive platforms for users to help each other with human-provided answers, which overcome the shortcoming of poor quality of the automatic answers. Surface pattern is proved an effective way to retrieve answers automatically. However, surface pattern does not include semantic information and is therefore called “poor-knowledge approaches”. Hence, it cannot extract precise answers or relevant information without semantically analyzing questions and answers.

To solve this problem, we firstly propose a novel type of pattern called semantic pattern and give the formal definition. The architecture of UIQA system based on semantic pattern is also presented, which includes question structure analysis, pattern matching, pattern generation, pattern classification and answer extraction.

After that, to generate semantic pattern automatically and effectively, this thesis proposes a new automatic generation method of semantic patterns from free-text questions. This method uses structural processing and name entity recognition (NER) to obtain the main structure of a question. An entropy-based model is used to select suitable words from questions for generalization. WordNet is then applied in our algorithm to get the best semantic labels from our Tagger Ontology for such chosen words. An evaluation method is also proposed to estimate the suitability of the

generated patterns and is implemented in a real UIQA system. Experiments with 5500 questions show that 63.9% generated patterns are satisfactory on average.

Finally, this thesis presents one of the applications of semantic pattern as an example - an automatic method for building a semantic dictionary from existing semantic pattern based questions for question categorization. This dictionary consists of two main parts: Semantic Domain Terms (SDT), which is a domain specific term list, and Semantic Labeled Terms (SLT), which contains common terms tagged with semantic labels. We implement the semantic dictionary construction method on a set of 2509 questions with semantic patterns in our system. Experimental results show that the precision of question classification is improved by 7.5% on average after using the constructed semantic dictionary compared with the baseline method.

Table of Contents

Abstract....1
Table of Contents....3
Chapter 1 Introduction....1
1.1 Background and Motivation....1
1.2 Scope of Research....4
1.3 Organization of This Thesis....5
Chapter 2 Literature Review....7
2.1 Question Answering....7
2.2 Automatic Question Answering System....10
2.3 User-Interactive Question Answering System....16
Chapter 3 Semantic Pattern for User-Interactive Question Answering....24
3.1 Semantic Pattern as a New Conceptual Framework....24
3.2 Compared With Related Concepts....27
3.3 Formal Definition of Semantic Pattern....31
3.4 System Architecture Based on Semantic Pattern....33
3.4.1 Structure Analysis and Pattern Matching....34
3.4.2 Pattern Generation and Classification....36
3.4.2.1 Pattern Instantiation Level Metrics....36
3.4.2.2 Pattern Generation....39
3.4.3 Answer Extraction....41
Chapter 4 Automatic Generation of Semantic Pattern....44
4.1 Background and Motivation....44
4.2 Previous Research of Pattern Generation....46
4.3 The Automatic Generation Method....48
4.3.1 Structure Processing and Name Entity Recognition....48
4.3.2 Nouns/Verbs Selection for Generalization....50
4.3.3 Semantic Mapping and Tagging....51
4.4 Experiments....52
4.4.1 Implementation....52
4.4.2 Evaluation....53
4.4.3 Experimental Results....54
Chapter 5 Automatic Construction of Semantic Dictionary Based on Semantic Pattern....57
5.1. Motivation for Semantic Dictionary....57
5.2. The Architecture of Semantic Dictionary....59

5.2.1 SDT, SLT and LCMT 60
5.2.2 Tagger Ontology 61
5.3. Automatic Construction of the Semantic Dictionary 63
5.3.1 Building Semantic Domain Terms (SDT) 64
5.3.2 Building Semantic Labeled Terms (SLT) 65
5.4. Applying Semantic Dictionary for Question Classification 67
5.5. Experiments and Evaluation 70
Chapter 6 Conclusions and Future Work 73
6.1 Summary 73
6.2 Contributions 75
6.2 Future Work 76
Acknowledgments 78
Publications 79
References 81