Extracting Concepts’ Relations and Users’ Preferences for Personalizing Query Disambiguation

Yan Chen, Georgia State University, USA
Yan-Qing Zhang, Georgia State University, USA

ABSTRACT

For most Web searching applications, queries are commonly ambiguous because words usually contain several meanings. Traditional Word Sense Disambiguation (WSD) methods use statistic models or ontology-based knowledge models to find the most appropriate sense for the ambiguous word. Since queries are usually short, the contexts of the queries may not always provide enough information for disambiguating queries. Thus, more than one interpretation may be found for one ambiguous query. In this paper, we propose a cluster-based WSD method, which finds out all appropriate interpretations for the query. Because some senses of one ambiguous word usually have very close semantic relations, we group those similar senses together for explaining the ambiguous word in one interpretation. If the cluster-based WSD method generates several contradictory interpretations for one ambiguous query, we extract users’ preferences from clickthrough data, and determine suitable concepts or concepts’ clusters that meet users’ interests for explaining the ambiguous query. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords:    Agglomerative Clustering; Ambiguous Queries; Information Content; Semantic Relation; Users’ Preferences

INTRODUCTION

Nowadays, Web search engines play a key role in retrieving information from the Internet to provide useful Web documents in response to users’ queries. The keywords-based search engines, like GOOGLE, YAHOO Search and MSN Live Search, explore documents by matching keywords in queries with words in documents. However, some keywords have more than one meaning, and such words may be related to different concepts in different contexts, so they are potentially ambiguous. Since current search engines simply search keywords separately and do not consider the contexts of queries, word sense ambiguity may result in searching errors.
for Web search applications. For example, if a user searches “drawing tables in a document” by MSN Live Search, five useless results related to the furniture table will be shown in the first result page. Therefore, an exact concept of a query may be determined by the contexts. Moreover, queries are usually short and contexts in queries do not always provide enough information for disambiguating queries. Under these circumstances, users’ preferences may be helpful for determining an appropriate concept for an ambiguous word. For an example, if a biologist searches “mouse”, we can speculate that the biologist is interested in Web pages related to a rodent “mouse” instead of a computer device “mouse.” Thus, both contexts of queries and users’ preferences are useful for disambiguating queries in Web search applications.

In fact, query disambiguation (QD) is a special application of Word Sense Disambiguation (WSD) problems. For most WSD problems, the set of possible meanings for a word is known ahead of time and stored in a lexical database. Then, the meaning for an ambiguous word is assigned depending on its contexts (Wilks, Slator, & Guthrie, 1996).

The traditional WSD methods seek one most related concept for an ambiguous word in a given context. However, ambiguous words usually contain some concepts that have very close semantic relations. From the Cambridge Dictionary, “bank” has four senses: (1) an organization where people and businesses can invest or borrow money, change it to foreign money, and so forth; (2) sloping raised land, especially along the sides of a river; (3) a pile or mass of things; or (4) a row of similar things. From above explanations, we may find that sense (3) and (4) have very similar meanings. Commonly, people are not aware of the senses’ small differences and use them interchangeably. Therefore, a cluster of similar senses for an ambiguous word may better describe user’s intents.

This article proposes a cluster-based WSD method. First, this article presents a new measure to evaluate the concepts’ similarities, which can better calculate the information content similarities than the Concept Distances proposed in the earlier research (Rada, Mili, Bicknell, & Blettner, 1989). Second, based on concepts’ similarities, this article presents an agglomerative clustering algorithm to find out all appropriate interpretations (sense clusters) for the query. Also, more than one similar sense may be grouped to explain a single ambiguous word in one interpretation, such as sense (3) and sense (4) can be both related to the ambiguous word “bank” in one interpretation. Third, by extracting users’ preferences from the clickthrough data, we may provide personalized interpretations that meet users’ interests.

The main contributions of the article are listed as follows:

1. Propose a new measure to evaluate the semantic relations between concepts in IS-A semantic tree.
2. Find out all appropriate interpretations for one query.
3. Find out all appropriate senses to explain one ambiguous word in one interpretation of the query.
4. Personalize QD for meeting users’ interests.

This article is organized as follows: First, we present a brief overview of the related work. Then, we describe our proposed method, followed by data collection and experiments section. Finally, we address some contributions and future related work.

RELATED WORK

Word Sense Disambiguation (WSD) is a task to determine one appropriate sense of a polysemous word within a specific context. Much effort has been put into solving the problem using a number of different approaches.

**Supervised learning.** This supervised machine learning method is one of the most frequently used approaches for WSD, which induces rules for disambiguating words from training data (manually sense-tagged context). Then, those rules can select correct senses of
Quantifying the Connectivity of a Semantic Warehouse and Understanding its Evolution over Time
[www.igi-global.com/article/quantifying-the-connectivity-of-a-semantic-warehouse-and-understanding-its-evolution-over-time/160172?camid=4v1a](www.igi-global.com/article/quantifying-the-connectivity-of-a-semantic-warehouse-and-understanding-its-evolution-over-time/160172?camid=4v1a)

Metamodels Construction Based on the Definition of Domain Ontologies
[www.igi-global.com/chapter/metamodels-construction-based-definition-domain/71855?camid=4v1a](www.igi-global.com/chapter/metamodels-construction-based-definition-domain/71855?camid=4v1a)