

Examining Existing Approaches Toward Semantic Disambiguation

Hyung Wook Choi – Drexel University, United States

Mat Kelly – Drexel University, United States

Abstract:

Inter-discipline collaboration allows scholars to adopt techniques or methodologies beyond their own specialization to their own research. However, due to increasing complexity among information fields, it may be difficult to understand the underlying semantic similarities among them. Many researchers have been trying to solve semantic disambiguation within or among fields. In this paper, we provide an overview of the methodologies used in prior research for semantic disambiguation from the Web of Science (WoS) and Library and Information Science Abstracts (LISA) databases from the last 10 years. This paper constitutes a preliminary step in identifying prior research in a domain we intend to explore further.

1.0 Introduction

Semantic disambiguation has been long discussed among researchers, especially in the natural language process (NLP) field. This problem has been subdivided into several aspects including Word Sense Disambiguation (WSD), Named-Entity Disambiguation (NED), and Semantic Textual Similarity (STS). Because one name entity can refer to multiple meanings, it is important to distinguish and connect the intended and correct entity. For instance, the term “polymorphism” can be interpreted differently depending on researchers' background knowledge or their specialties, e.g., two or more different morphs in a species in biology and the same interface for different data types in computer science.

A variety of studies have suggested methods and subsequently improved upon them to enhance the performance when using large-scale corpora. In this study, we examined contemporary articles and approaches from the last 10 years that studied issues related to semantic disambiguation and reviewed methods that they used. The purpose of this review is to understand contemporary approaches toward semantic disambiguation to inform our further research.

2.0 Methodology

We sought to identify existing literature on the topic of semantic disambiguation to inform our further research. To collect the dataset, we selected Web of Science (WoS) and Library and Information Science Abstracts (LISA) as the target databases and used the keyword "Semantic disambiguation" to search in July 2021. We additionally confined the search conditions with the additional restrictions of being written in English, published between 2012 and 2021, and having a publication type of either scholarly journal or article. The intention of this limitation was to perform a preliminary scan of recent work to inform our work and facilitate novelty rather than reiterate the redundancy of research.

The aforementioned search resulted in 263 articles from LISA and 461 from WoS. Next, we excluded 33 redundant articles then read through the resulting 691 articles' abstracts to decide whether each article is related to either contextualized disambiguation or Word Semantic Disambiguation (WSD). This resulted in the final of 142 articles for the final reviewing dataset for this study.

We reviewed the selected articles to identify the method or specific applied

algorithms. After the examination, we excluded additional 28 survey articles and cross-lingual articles.

3.0 Results

We identified methods that are used in 113 articles that are related to semantic disambiguation topics. We then extracted the keywords representative of the methods used from the papers. Next, we manually classified 113 articles based on frequently used method types. Articles that use methods that do not fall into types from the Figure 1 are omitted. WSD algorithm is the most applied algorithm within the set of 16 articles. The methods identified as "annotation", "kernel", and "embedding" are also used in 5 articles. One notable aspect we identified is that 36 studies adopted combined approaches to resolving semantic disambiguation problems, which aligns with the previous study that choosing a single approach in semantic disambiguation is not encouraged.

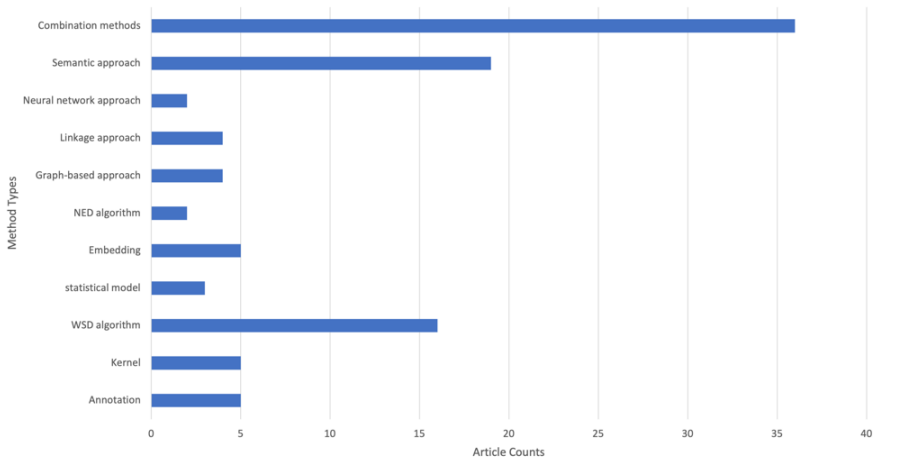


Figure 1. Research Methods Used in Articles and the Counts of Articles

4.0 Discussion

In the last decade, the methods to solve semantic disambiguation have been diverse, but adjusting WSD algorithms is still a major method that researchers choose. On the other hand, because semantic disambiguation is a problem that needs to be solved among different fields, the studies have been conducted from diverse perspectives, which resulted in embracing diverse methods. For instance, Abualhaja and Zimmermann (2016) used bee colony optimization to solve word sense disambiguation.

This study is a part of the process of resolving semantic disambiguation and discontinuity among different fields. This study was aimed to see different methods that are used to resolve semantic disambiguation from computer science and library and information science perspectives. The future work will include identifying and selecting specific modules or algorithms that can be modified or adjusted for better performances.

References

Abualhaja, Sallam and Karl-Heinz Zimmermann. 2016. "D-Bees: A novel method inspired by bee colony optimization for solving word sense disambiguation." *Swarm and evolutionary computation* 27: 188-195.