Computer Speech and Language 74 (2022) 101337

Contents lists available at ScienceDirect

Computer Speech & Language

journal homepage: www.elsevier.com/locate/csl

A novel word sense disambiguation approach using WordNet knowledge graph

Mohannad AlMousa [a,*], Rachid Benlamri [b], Richard Khoury [c]

[a] Department of Software Engineering, Lakehead University, Thunder Bay, ON, P7B 5E1, Canada
[b] College of The North Atlantic Qatar, Qatar
[c] Department of Computer Science and Software Engineering, Université Laval, Québec, QC G1V 0A6, Canada

ARTICLE INFO

Keywords:
Semantic word sense disambiguation
Knowledge-based
Knowledge graph
WordNet

ABSTRACT

Various applications in computational linguistics and artificial intelligence rely on high-performing word sense disambiguation techniques to solve challenging tasks such as information retrieval, machine translation, question answering, and document clustering. While text comprehension is intuitive for humans, machines face tremendous challenges in processing and interpreting a human's natural language. This paper presents a novel knowledge-based word sense disambiguation algorithm, namely Sequential Contextual Similarity Matrix Multiplication (SCSMM). The SCSMM algorithm combines semantic similarity, heuristic knowledge, and document context to respectively exploit the merits of local sense-based context between consecutive terms, human knowledge about terms, and a document's main topic in disambiguating terms. Unlike other algorithms, the SCSMM algorithm guarantees the capture of the maximum sentence context while maintaining the terms' order within the sentence. The proposed algorithm outperformed all other algorithms when disambiguating nouns on the combined gold standard datasets, while demonstrating comparable results to current state-of-the-art word sense disambiguation systems when dealing with each dataset separately. Furthermore, the paper discusses the impact of granularity level, ambiguity rate, sentence size, and part of speech distribution on the performance of the proposed algorithm.

1. Introduction

Many Natural Language Processing (NLP) applications rely on Word Sense Disambiguation (WSD), either directly or indirectly. The list includes, but is not limited to Machine Translation (MT), Information Retrieval (IR), Question Answering (QA), Named Entity Recognition (NER), and text summarization. WSD is considered one of the oldest tasks of computational linguistics dating back to the 1940s. It started as a distinct task when machine translation was first developed. The first challenge that triggered WSD task is MT in the 1940s. Since then, researchers have been developing models and algorithms to improve the accuracy of this task using various approaches; supervised, semi-supervised, and knowledge-based systems. WSD is an essential task in many other applications, such as IR, information extraction, knowledge acquisition, and NLP. With the introduction of supervised machine learning in the 1990s, various supervised approaches attempted to solve the WSD task. More recent studies are exploring semi-supervised and unsupervised

* Corresponding author.
E-mail addresses: malmous@lakeheadu.ca (M. AlMousa), rachid.benlamri@cna-qatar.edu.qa (R. Benlamri), richard.khoury@ift.ulaval.ca (R. Khoury).
URL: http://www2.ift.ulaval.ca/~rikho/ (R. Khoury).

https://doi.org/10.1016/j.csl.2021.101337
Received 11 November 2020; Received in revised form 10 August 2021; Accepted 29 November 2021
Available online 1 January 2022