

Executive Summary

Context-Free Named Entity Disambiguation with Wikipedia: A Computationally Efficient Method

Andrea Ricky

The semantic interpretation of unstructured text corpora is complicated by the intrinsic idiosyncrasies of natural languages. Homonymy and polysemy are language phenomena that refer to the coexistence of various possible meanings for a word or phrase in various contexts and informational backgrounds. Named Entity Disambiguation is the name of the branch of artificial intelligence (AI) and natural language processing (NLP) research concerned with this class of problems (NED). Due to the inherent difficulties and increased complexity of the task, deep neural networks are often used as "black boxes" to map input data to classification or approximation outputs. As a result, the data science lifecycle process frequently lacks an understanding of how these systems operate internally.

Our suggested techniques are based on rich Wikipedia semantic information and domain knowledge. Our domain knowledge base can be expanded with semantic data from any Wikipedia annotated text corpus and is not, however, restricted to Wikipedia. The second greatest commonness entity is mentioned as well, with the relative commonness measure effectively conveying the normalized difference of the top commonness object. Any end-to-end entity linking technique starts by extracting potential mentions from the unstructured text input that have semantic relevance. The highlight of this effort is the run-time performance. To reduce run-time complexity, the computational burden has been shifted to the development of data structures that can allow quick operations in memory.

Practically any contemporary commodity personal computer can be used to conduct the experiments for our evaluation. We used the Apache Spark framework 3.2.1 and a 40vCPU instance with 64 GB of memory to efficiently pre-process Wikipedia. Since the Wikification process strongly relies on knowledge, acquiring corpora that are semantically connected to Wikipedia entities can help us get even better results. To that end, the contribution of fuzzy matching for the growth of inter-wiki coverage and unsupervised link-prediction algorithms on the Wikipedia corpus may extend our knowledge base and increase the accuracy of our approaches.

The extremely effective and efficient methodologies can be used for a vast complexity reduction in conjunction with more complex, precision-oriented, and compute-intensive approaches in a layered architecture, for retrieving entity links using our relative-commonness-based methodology at the first stage, followed by a more complex, precision-oriented, and compute-intensive approach. This article's main objective has been to suggest and assess a new approach for lowering the present



computational hurdle for using a named entity disambiguation task. Our tests on well-known datasets show encouraging results, indicating that our approach is favourable for widespread use with huge data.

Source: [Information](#)

KEYWORDS

Named entity disambiguation; text annotation; context-free Wikification; word sense disambiguation; ontologies; Wikification; fast Wikification; artificial intelligence; machine learning

