Chapter 40
A Fuzzy Logic Based Synonym Resolution Approach for Automated Information Retrieval

Mamta Kathuria

YMCA University of Science and Technology, India

Chander Kumar Nagpal

YMCA University of Science and Technology, India

Neelam Duhan

YMCA University of Science and Technology, India

ABSTRACT

Precise semantic similarity measurement between words is vital from the viewpoint of many automated applications in the areas of word sense disambiguation, machine translation, information retrieval and data clustering, etc. Rapid growth of the automated resources and their diversified novel applications has further reinforced this requirement. However, accurate measurement of semantic similarity is a daunting task due to inherent ambiguities of the natural language, spread of web documents across various domains, localities and dialects. All these issues render to the inadequacy of the manually maintained semantic similarity resources (i.e. dictionaries). This article uses context sets of the words under consideration in multiple corpora to compute semantic similarity and provides credible and verifiable semantic similarity results directly usable for automated applications in the intelligent manner using fuzzy inference mechanism. It can also be used to strengthen the existing lexical resources by augmenting the context set and properly defined extent of semantic similarity.

INTRODUCTION

With the passage of time, human dependency has increased manifold on the automated resources. These resources depend upon the underlying inference mechanism and stored data for their decisions/actions. Unlike humans, these resources are devoid of their own intelligence and analytical capability. Therefore, it is necessary that these systems be embedded with the resources which are totally free from ambiguity so that their decisions/actions are reliable and credible. As and when an automated system has to deal with an inherently ambiguous environment such as Natural Language Processing(NLP), there are two options available to the designer to increase the reliability and credibility of the system:

1. To deal directly with the ambiguity at the inference level by designing the knowledge base with the help of tools like fuzzy logic which have the ability to handle the ambiguity to an extent and provide a coarse (not that precise) automated solution;
2. To make the underlying data less ambiguous and more suitable for automated applications.

In an automated system, which works in an inherently ambiguous environment, both the approaches may be required but an effort in the direction of second approach shall reduce the requirement of first approach making the system more precise.

One of the most frequently used automated systems in the present-day environment is *web search engine,* commonly referred to as search engine. It tries to provide a set of documents semantically related to the query submitted by the user in the natural language. For getting better results, the search needs to be widened by using terms semantically similar to input query. This leads to the requirement of unambiguous, efficient lexical resources providing precise semantic similarity information directly usable by the automated systems. The heterogeneity in web literature owing to its global spread over geographical boundaries, languages, dialects combined with other inherent ambiguities in the natural language makes this task quite difficult which is beyond the grasp of manually maintained dictionaries. To overcome this drawback, the current literature provides several empirical methods based upon web page count and text snippets count as discussed in literature survey presented in this paper. These methods basically fail to take into account the exact context of the word pair usage due to following reasons:

●   Occurrence of a pair of words within the same page does not ensure the similarity of their contexts;
●   Sometimes the size of the snippets is increased to ensure the larger number of instances which may result in the loss of context.

To ensure the better efficiency and credibility of these web based automated systems such as search engines, it is necessary that the underlying lexical resources be automation friendly and as less ambiguous as possible. An information retrieval system based upon the crude semantic similarity taken from online lexical sources like WordNet (2005), etc., without considering the context, leads to fetching of large number of undesired pages which otherwise may not be that useful. Therefore, there is a need to tune these lexical resources in such a manner that the derived synonyms are also provided with the following:

●   Their respective applicable context set;
●   Properly defined extent of semantic similarity.

Related Content

Hidden Markov Model Based Visemes Recognition, Part I: AdaBoost Approach
Say Wei Fooand Liang Donga (2009). *Visual Speech Recognition: Lip Segmentation and Mapping (pp. 326-355).*
www.irma-international.org/chapter/hidden-markov-model-based-visemes/31073

Agreement Technologies for Conflict Resolution
Vicente Julian, Victor Sanchez-Anguix, Stella Herasand Carlos Carrascosa (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications (pp. 464-484).*
www.irma-international.org/chapter/agreement-technologies-for-conflict-resolution/239951

Integrating Semantic Acquaintance for Sentiment Analysis
Neha Guptaand Rashmi Agrawal (2021). *Advanced Concepts, Methods, and Applications in Semantic Computing (pp. 93-112).*
www.irma-international.org/chapter/integrating-semantic-acquaintance-for-sentiment-analysis/271122

Translational Mismatches Involving Clitics (Illustrated from Serbian ~ Catalan Language Pair)
Jasmina Milieviand Àngels Catena (2015). *Modern Computational Models of Semantic Discovery in Natural Language (pp. 235-255).*
www.irma-international.org/chapter/translational-mismatches-involving-clitics-illustrated-from-serbian--catalan-language-pair/133881

Guru: Designing a Conversational Expert Intelligent Tutoring System
Andrew M. Olney, Natalie K. Personand Arthur C. Graesser (2012). *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches (pp. 156-171).*
www.irma-international.org/chapter/guru-designing-conversational-expert-intelligent/64586