The MIT Press

Journals

Sign In / Register

Books Journals Digital Resources About Contact

Quarterly (March, June, September, December)

160pp. per issue

6 3/4 x 10

Founded: 1974

2018 Impact Factor: 1.319

2018 Google Scholar h5-index: 32

ISSN: 0891-2017

E-ISSN: 1530-9312

Journal Resources

Editorial Info

Abstracting and Indexing

Release Schedule

Advertising Info

Author Resources

Co-occurrence Retrieval: A Flexible Framework for Lexical Distributional Similarity

Julie Weeds and David Weir

Posted Online March 13, 2006

https://doi.org/10.1162/089120105775299122

© 2006 Association for Computational Linguistics

Computational Linguistics

Volume 31 | Issue 4 | December 2005

p.439-475

Download Options

Abstract Authors

Techniques that exploit knowledge of distributional similarity between words have been proposed in many areas of Natural Language Processing. For example, in language modeling, the sparse data problem can be alleviated by estimating the probabilities of unseen co-occurrences of events from the probabilities of seen co-occurrences of similar events. In other applications, distributional similarity is taken to be an

Submission Guidelines
Publication Agreement
Author Reprints

Reader Resources

Rights and Permissions
Most Read
Most Cited

More About Computational Linguistics

Metrics

67 Total citations

3 Recent citations

5.55 Field Citation Ratio

n/a Relative Citation Ratio

Open Access

Computational Linguistics Computational Linguistics is Open Access. All content is freely available in electronic format (Full text HTML, PDF, and PDF Plus) to readers across the globe. All

approximation to semantic similarity. However, due to the wide range of potential applications and the lack of a strict definition of the concept of distributional similarity, many methods of calculating distributional similarity have been proposed or adopted.

In this work, a flexible, parameterized framework for calculating distributional similarity is proposed. Within this framework, the problem of finding distributionally similar words is cast as one of co-occurrence retrieval (CR) for which precision and recall can be measured by analogy with the way they are measured in document retrieval. As will be shown, a number of popular existing measures of distributional similarity are simulated with parameter settings within the CR framework. In this article, the CR framework is then used to systematically investigate three fundamental questions concerning distributional similarity. First, is the relationship of lexical similarity necessarily symmetric, or are there advantages to be gained from considering it as an asymmetric relationship? Second, are some co-occurrences inherently more salient than others in the calculation of distributional similarity? Third, is it necessary to consider the difference in the extent to which each word occurs in each co-occurrence type?

Two application-based tasks are used for evaluation: automatic thesaurus generation and pseudo-disambiguation. It is possible to achieve significantly better results on both these tasks by varying the parameters within the CR framework rather than using other existing distributional similarity measures; it will also be shown that any single unparameterized measure is unlikely to be able to do better on both tasks. This is due to an inherent asymmetry in lexical substitutability and therefore also in lexical distributional similarity.

Forthcoming

Most Read

See More

Lexicon-Based Methods for Sentiment Analysis (14087 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307

Computational Linguistics and Deep Learning (10542 times)
Christopher D. Manning
Computational Linguistics
Volume: 41, Issue: 4, pp. 701-707

Near-Synonymy and Lexical Choice (3675 times)
Philip Edmonds et al.
Computational Linguistics
Volume: 28, Issue: 2, pp. 105-144

articles are published under a CC BY-NC-ND 4.0 license. For more information on allowed uses, please view the CC license.
Support OA at MITP

(Note that the Most Read numbers are based on the number of full text downloads over the last 12 months.)

Most Cited

See More

Lexicon-Based Methods for Sentiment Analysis (436 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307

A Systematic Comparison of Various Statistical Alignment Models (174 times)
Franz Josef Och et al.
Computational Linguistics
Volume: 29, Issue: 1, pp. 19-51

Opinion Word Expansion and Target Extraction through Double Propagation (147 times)
Guang Qiu et al.
Computational Linguistics
Volume: 37, Issue: 1, pp. 9-27

(Note that the Most Cited numbers are based on Crossref's Cited-by service and reflect citation information for the past 24 months.)

Download Options

Favorite
Sign up for Alerts
Download Citation
RSS TOC
RSS Citation
Submit your article
Support OA at MITP

© 2018 The MIT Press

Technology Partner: Atypon Systems, Inc.
CrossRef Member
COUNTER Member
The MIT Press colophon is registered in the U.S. Patent and