



The MIT Press

Journals

[Sign In / Register](#)
[Books](#)
[Journals](#)
[Digital](#)
[Resources](#)
[About](#)
[Contact](#)


Home | Computational Linguistics | List Article navigation
of Issues | Volume 40 , No. 3 | Improved
Estimation of Entropy for Evaluation of
Word Sense Induction



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

[Submission
Guidelines](#)
[Publication
Agreement](#)

Improved Estimation of Entropy for Evaluation of Word Sense Induction

[Linlin Li](#), [Ivan Titov](#) and
[Caroline Sporleder](#)

Posted Online September 05, 2014

https://doi.org/10.1162/COLI_a_00196

© 2014 Association for Computational Linguistics

Computational Linguistics
Volume 40 | Issue 3 | September 2014
p.671-685

 [Download Options](#) >


[Abstract](#) [Full Text](#) [Authors](#)

Information-theoretic measures are among the most standard techniques for evaluation of clustering methods including word sense induction (WSI) systems. Such measures rely on sample-based estimates of the entropy. However, the standard maximum likelihood estimates of the entropy are heavily biased with

Author Reprints

Reader Resources

Rights and Permissions
 Most Read
 Most Cited

More About Computational Linguistics 

Metrics 

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 articles are published under a [CC BY-NC-ND 4.0 license](#). For more information on allowed uses, please view the [CC license](#).

the bias dependent on, among other things, the number of clusters and the sample size. This makes the measures unreliable and unfair when the number of clusters produced by different systems vary and the sample size is not exceedingly large. This corresponds exactly to the setting of WSI evaluation where a ground-truth cluster sense number arguably does not exist and the standard evaluation scenarios use a small number of instances of each word to compute the score. We describe more accurate entropy estimators and analyze their performance both in simulations and on evaluation of WSI systems.

Forthcoming

Most Read




[See More](#)

- | | | |
|--|---|--|
| <p> Lexicon-Based Methods for Sentiment Analysis (13965 times)
 Maite Taboada et al.
 Computational Linguistics
 Volume: 37, Issue: 2, pp. 267-307</p> | <p> Computational Linguistics and Deep Learning (10500 times)
 Christopher D. Manning
 Computational Linguistics
 Volume: 41, Issue: 4, pp. 701-707</p> | <p> Near-Synonymy and Lexical Choice (3653 times)
 Philip Edmonds et al.
 Computational Linguistics
 Volume: 28, Issue: 2, pp. 105-144</p> |
|--|---|--|

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

Most Cited

[See More](#)

- | | | |
|--|---|--|
| <p> Lexicon-Based Methods for Sentiment Analysis (436 times)
 Maite Taboada et al.
 Computational Linguistics
 Volume: 37, Issue: 2, pp. 267-307</p> | <p> A Systematic Comparison of Various Statistical Alignment Models (174 times)
 Franz Josef Och et al.
 Computational Linguistics
 Volume: 29, Issue: 1, pp. 19-51</p> | <p> Opinion Word Expansion and Target Extraction through Double Propagation (147 times)
 Guang Qiu et al.
 Computational Linguistics
 Volume: 37, Issue: 1, pp. 9-27</p> |
|--|---|--|

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

Support OA at MITP

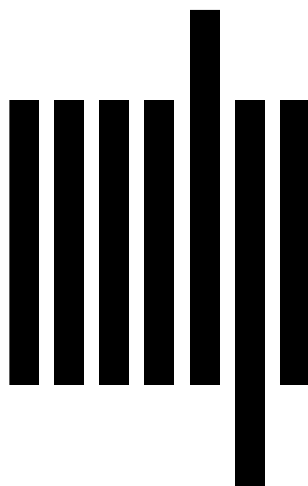
Download Options

Sign up for
Favorite  Alerts 

Download Citation  RSS TOC 

RSS Citation  Submit your article

Support OA at MITP 



Journals

Terms & Conditions

Privacy Statement

Contact Us

Books

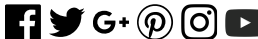
US

UK

Connect

One Rogers Street
Cambridge MA
02142-1209

Suite 2, 1 Duchess
Street London,
W1W 6AN, UK



© 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 Trademark Office.
[Site Help](#)