A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains

Dominik Schlechtweg  Anna Hätty  Marco Del Tredici  Sabine Schulte im Walde

Lexical Semantic Change Detection

Diachronic LSCD: detect sense-divergences for words over time in text
(1) 1796 Ein paar Donnerwetter nebst Regen trugen noch mehr zur Kühle bey.
(2) 1875 Potz Donnerwetter, bin aber ich g’loffen!

Synchronic LSCD: from general-language to domain-specific use
(3) general, ... um im Winter die Gleise von Schnee und Eis zu befreien.
(4) cooking Das Eiweiss zu Schnee schlagen und darunterheben.

→ we perform the first large-scale evaluation for LSC detection

Takeaway

Representation: SGNS performs best on average
▶ SGNS is more stable than expected

Alignment: OP alignment works
▶ SGNS should be mean centered before alignment

Measures: CD outperforms LND
▶ Dispersion measures have low performance

Best combination: SGNS+OP+CD

Models

Task, Corpora & Datasets

Ranking Task: Given two corpora C_a and C_b, rank all target words according to their degree of LSC between C_a and C_b as annotated by human judges.

Corpora:
▶ unsupervised
▶ distributional
▶ bag-of-words-based
▶ differ by
1. semantic representation type:
   ▶ semantic vector spaces
   ▶ topic distributions
2. alignment methods
3. LSCD measures

Datasets:
▶ DURel: rank of 22 target words annotated across time periods
  a: 1750–1799
  b: 1850–1899
▶ SURel: rank of 22 target words annotated across domains
  a: general-language
  b: domain-specific

Best Results

Table: Best results of five sampling, t=subsampling, Spearman m(h,l): mean, highest and lowest results.

Acknowledgments

We thank Heim Dubossarsky, Simon Hanchev, Andrei Karjus, Barbara McGillivray, Canett Oguz, Sascha Schlechtweg, Nina Tahmasebi and the three anonymous reviewers for their valuable comments. We further thank Michael Dorna and Brigitte Wang for their helpful advice. We also thank Lea Frermann for providing the code of SCAN and helping to set up the implementation.

References


Code for all models is available at:
github.com/Garrafao/LSCDetection