RESEARCH STAY WEEK 7, Word embeddings

Daniel Cajas A01708637

CONTEXT

For a long time, word embedding was simply seen as one more step during NLP tasks. It isn't until fairly recently, arguably 2008, that it was split into its own task by the community. Since then there have been a lot of research into different ways of doing word embeddings and how they can be either directly used or its effects on downstream tasks.

There are 2 ways of evaluating word embeddings. Intrinsic, where the features (vectors) themselves are evaluated. And extrinsic, where different word embeddings are used with the same downstream tasks and they performances are compared. Both have significant issues. Intrinsic evaluation is complicated since there is no clear “right” solution as that would mean solving linguistics. On the other hand extrinsic evaluation is not generalizable and more expensive.

SEARCH METHODOLOGY

[2] Is a very comprehensive survey of word embedding evaluation methods. However i wanted to see know if there have been new advancements in the field. I used scopus and searched for word embedding evaluation articles published since 2022.

COMPARISON

https://docs.google.com/spreadsheets/d/19KFbfroBwfrYbnzowMjZ6c5nRp7nMZxvynRO4oW78KI/edit?usp=sharing

Title	Objective	Intrinsic/Extrinsic	Intrinsic Methods	Extrinsic methods / models	Takeaways
A survey of word embeddings evaluation methods	Survey of different evaluation methods and providing useful datasets accordingly.	Both	Conscious intrinsic evaluation - Word semantic similarity (Similarity) - Word analogy - Thematic Fit - Synonym detection - Outlier word detection Subconscious intrinsic evaluation - Semantic priming - Neural activation patterns - Eye movement data Thesaurus evaluation - Thesaurus vectors - Dictionary definition graph - Cross-match test - Semantic difference - Semantic networks Linguistic-driven methods - Photosemantic analysis - Bi-gram co-occurrence frequency	Noun phrase chunking, Named Entity Recognition, Sentiment Analysis, Shallow syntax parsing, Semantic role labeling, Negation Scope, POS tagging, Text classification, Metaphor detection, Paraphrase Detection, Textual entailment detection,	Systematized evaluation methods as well as highlighting relevant problems and giving usefull resources (datasets)
Evaluation methods for unsupervised word embeddings	Survey of different evaluation methods and showing the effectiveness of croudsourcing for evaluation.	Both	Absolute - Relatedness (Similarity) - Analogy - Categorization - Selectional precence (Subject vs object) Comparative - Outlier word with croudsourcing (Amazon Mechanical Turk)	Word chuncking, sentiment analysis	New evaluation method: Coherence Checking if a neighborhood of words is closeley related, instead of just similarity between pairs of words.
A Fistful of Vectors: A Tool for Intrinsic Evaluation of Word Embeddings	Creating n off the shelve solution for word embedding evaluation	Intrinsic	- Similarity - Analogy - Categorization - Outlier detection	N/A	Python package that runs a multitude of evaluations agains many datasets.

BIBLIOGRAFÍA

[1] Bakarov, A. (2018). A Survey of Word Embeddings Evaluation Methods. https://arxiv.org/abs/1801.09536

[2] Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015). Evaluation methods for unsupervised word embeddings. In L. Màrquez, C. Callison-Burch, & J. Su (Eds.), Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 298–307). Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1036

[3] Ascari, R., Giabelli, A., Malandri, L. et al. A Fistful of Vectors: A Tool for Intrinsic Evaluation of Word Embeddings. Cogn Comput 16, 949–963 (2024). https://doi.org/10.1007/s12559-023-10235-3