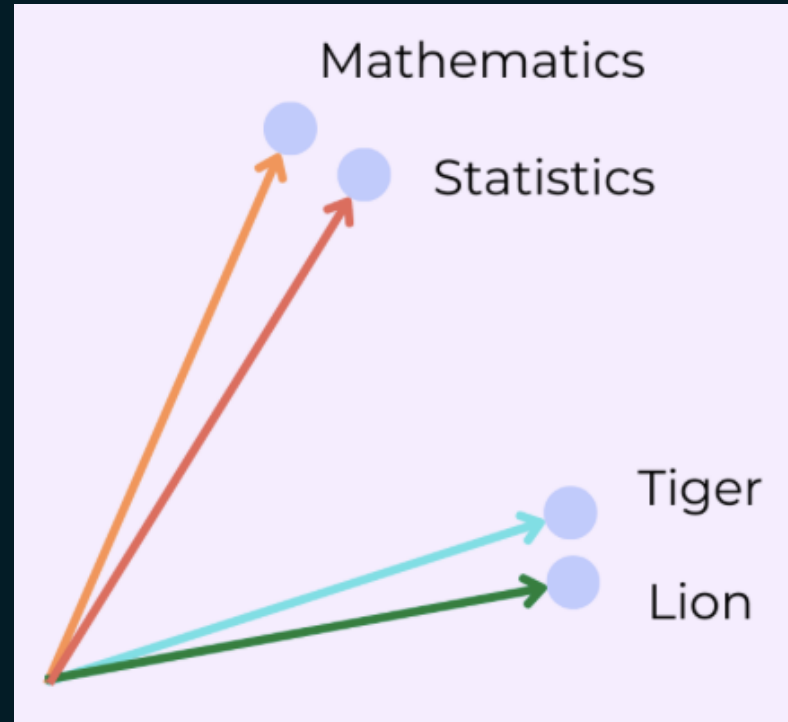


Word Embeddings in NLP Applications

Week 7 Mini Survey
Luis Alberto Portilla López

Word Embeddings Relevance and Usage

Vector representations of words that capture their meanings based on their context



Understanding the evaluation of word embeddings is crucial for improving their performance in real-world applications.

Key Evaluation Approaches

Intrinsic

Directly assesses embeddings based on linguistic tasks (e.g., word similarity, analogy tasks)

Extrinsic

Evaluates embeddings by measuring their impact on downstream tasks (e.g., text classification, sentiment analysis)

Challenges

No single evaluation method provides a complete picture; different tasks require different assessments

Specific Challenges

- **Subjectivity in human judgments**
- **Dataset biases and limitations**
- **Variability in performance across tasks and models**



These challenges highlight the need to use both intrinsic and extrinsic methods to fully understand embedding quality.

Search Methodology & Criteria



**CITATION CHAINING AND
FORWARD CITATION**



CRITERIA:



KEYWORD SEARCH



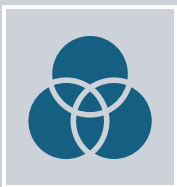
- Initial review of abstracts to assess relevance based on the title, publication venue, and year.



- Direct and indirect relevance to the paper being cross-referenced through the abstract.



- Consideration of the number of citations and field-weighted citation impact (fwci), a metric that measures the citation impact of a paper adjusted for disciplinary differences.



BOOLEAN SEARCH

Preliminary Terms



Key terms identified during the week:

- **Extrinsic Evaluation**
- **Intrinsic Evaluation**
- **Coherence**
- **Vector**

Document Comparison



“A Survey of Word Embeddings Evaluation Methods”



“Evaluation methods for unsupervised word embeddings”



“ Word Embeddings: A Survey”



“Word Embeddings for Sentiment Analysis: A Comprehensive Empirical Survey”

Characteristic	A Survey of Word Embeddings Evaluation Methods	Evaluation Methods for Unsupervised Word Embeddings	Evaluation Methods for Unsupervised Word Embeddings	Word Embeddings for Sentiment Analysis: A Comprehensive Survey
Primary Focus	Evaluation of intrinsic vs. extrinsic methods for word embeddings	Intrinsic and extrinsic evaluations of unsupervised embeddings	Intrinsic and extrinsic evaluation of word embeddings	Examines how word embeddings perform in sentiment analysis tasks
Objective	Systematize and classify word embeddings evaluation techniques	Evaluate and compare embeddings on relatedness, coherence, and downstream tasks	Explore strengths of word embeddings through different evaluation methods	Analyze the impact of training methods, corpus size, and thematic relevance on sentiment analysis
Application	Various NLP tasks including classification, NER, etc.	Relatedness, coherence evaluations, and tasks like chunking, sentiment analysis	Word similarity, analogy tasks, NLP performance	Sentiment analysis on tweets, song lyrics, movie and item reviews
Dataset	Multiple public datasets for intrinsic/extrinsic tasks	Wikipedia, Twitter, Google News, various combined corpora	Wikipedia, Common Crawl, other large text datasets	Large, thematically varied corpora (e.g., Twitter, Amazon reviews)

Characteristic	A Survey of Word Embeddings Evaluation Methods	Evaluation Methods for Unsupervised Word Embeddings	Evaluation Methods for Unsupervised Word Embeddings	Word Embeddings for Sentiment Analysis: A Comprehensive Survey
Evaluation Metrics	Correlation with human judgment, task performance scores	Precision, recall, accuracy on various linguistic tasks	Similarity scores, analogy accuracy, task-specific metrics	Sentiment polarity detection accuracy, effect of lexicon integration
Performance Insights	Varied results; no single evaluation method is universally reliable	Different embeddings excel in different tasks; task-specific tuning required	Mixed results; embeddings need to be contextually adapted	GloVe outperforms Word2Vec in sentiment analysis; lexicon-enhanced embeddings yield better results
Challenges Identified	Subjectivity in human-judged tasks, limited dataset diversity	Embedding performance varies widely across task types; coherence issues	Complex relationships between intrinsic and extrinsic metrics	Thematic relevance impacts performance; domain-specific embeddings often needed

Characteristic	A Survey of Word Embeddings Evaluation Methods	Evaluation Methods for Unsupervised Word Embeddings	Evaluation Methods for Unsupervised Word Embeddings	Word Embeddings for Sentiment Analysis: A Comprehensive Survey
Optimization Approach	Varied, often task-specific tuning and selection	Various spectral and optimization-based methods	Matrix factorization, negative sampling, language model adaptation	Incorporation of thematic, sentiment lexicons, and retraining on domain-specific data
Comparison with other models	Extensive comparison of intrinsic vs. extrinsic approaches	Comparative intrinsic evaluations using novel tasks	Comparisons mainly within intrinsic and extrinsic paradigms	Comparisons mainly within intrinsic and extrinsic paradigms

References

- Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015). Evaluation methods for unsupervised word embeddings. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 298-307. Lisbon, Portugal: Association for Computational Linguistics.
- Chiu, B., Korhonen, A., & Pyysalo, S. (2016). A Survey of Word Embeddings Evaluation Methods. *arXiv preprint arXiv:1608.02227*.
- Almeida, F., & Xexéo, G. (2019). Word Embeddings: A Survey. *arXiv preprint arXiv:1901.09069v2*.
- Çano, E., & Morisio, M. (2019). Word Embeddings for Sentiment Analysis: A Comprehensive Empirical Survey. Papers with Code. Retrieved from <https://paperswithcode.com/paper/word-embeddings-for-sentiment-analysis-a>