# Leveraging Smooth Manifolds for Lexical Semantic Change Detection across Corpora

Anmol Goel, Ponnurangam Kumaraguru

NeurIPS 2020 workshop on Differential Geometry meets Deep Learning

# Motivation

Lexical Semantic Change Detection (LSCD) is the task of identifying words that change meaning over time (diachrony) or across different domains (synchrony).

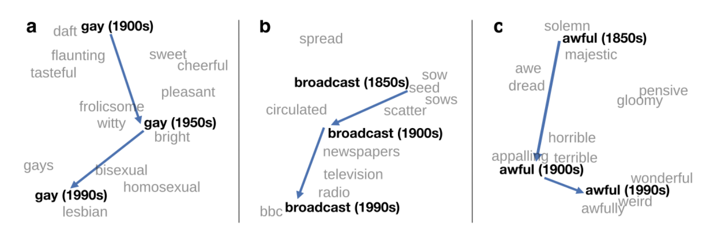


**Figure 1:** Changes in word meaning over time.

Image from: *`https://nlp.stanford.edu/projects/histwords/`*

# Motivation

- Word embeddings like word2vec or GloVe trained on different text corpus do not lie in the same embedding space.

# Motivation

- Word embeddings like word2vec or GloVe trained on different text corpus do not lie in the same embedding space.
- The different embedding spaces need to be aligned before comparing words via similarity measures like cosine similarity.

# Predominant Approach

Orthogonal Procrustes Alignment [2] learns an orthogonal matrix $R$ to project embedding matrix $X$ to $Y$.

$$R = \underset{Q^T Q = I}{\operatorname{argmin}} ||QX - Y||_2$$

# Proposed Approach

- We leverage the rich geometry of smooth Riemannian manifolds to learn the latent space.

# Proposed Approach

- We leverage the rich geometry of smooth Riemannian manifolds to learn the latent space.
- $$\underset{P,Q \in \mathcal{O}^d; B \succ 0}{\operatorname{argmin}} \; ||A^T P^T M Q B - Y||^2 + \lambda ||M||^2$$

# Proposed Approach

- We leverage the rich geometry of smooth Riemannian manifolds to learn the latent space.
- $$\underset{P,Q \in \mathcal{O}^d; B \succ 0}{\text{argmin}} \quad ||A^T P^T M Q B - Y||^2 + \lambda ||M||^2$$
- We learn corpus specific rotation matrices $P$ and $Q$ directly on the Stiefel Manifold.

# Proposed Approach

- We leverage the rich geometry of smooth Riemannian manifolds to learn the latent space.
-
$$\underset{P,Q \in \mathcal{O}^d; B \succ 0}{\text{argmin}} \; ||A^T P^T M Q B - Y||^2 + \lambda ||M||^2$$
- We learn corpus specific rotation matrices $P$ and $Q$ directly on the Stiefel Manifold.
- Additionally, we learn a common Mahalanobis metric $M$ to generalize cosine similarity by optimizing over the PSD Manifold.

# Proposed Approach

- We leverage the rich geometry of smooth Riemannian manifolds to learn the latent space.
- $$\underset{P,Q\in\mathcal{O}^d;B\succ 0}{\text{argmin}} \; ||A^T P^T M Q B - Y||^2 + \lambda ||M||^2$$
- We learn corpus specific rotation matrices $P$ and $Q$ directly on the Stiefel Manifold.
- Additionally, we learn a common Mahalanobis metric $M$ to generalize cosine similarity by optimizing over the PSD Manifold.
- Assuming that most words remain stable, we construct identity matrix Y which aligns each word to itself.

To quantitatively measure the amount of semantic change for a word $w_i$ from original embedding space $A$ and $B$, we use the following formulation involving cosine similarity [3].

$$LCS_{w_i} = cos(M^{\frac{1}{2}} Pa_i, M^{\frac{1}{2}} Qb_i)$$

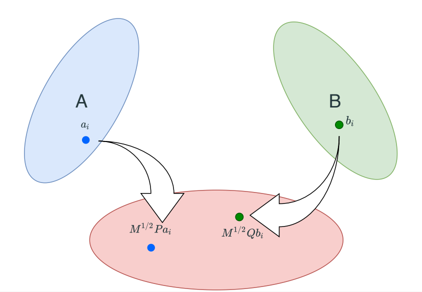


**Figure 2:** Projecting matrix $A$ and $B$ to a common latent space

# Manifold Optimization

- The proposed framework converts a constrained optimization problem in Euclidean space to an unconstrained optimization over smooth manifolds.

# Manifold Optimization

- The proposed framework converts a constrained optimization problem in Euclidean space to an unconstrained optimization over smooth manifolds.
- Structural bias of manifolds captures latent relationships between the two vector spaces.

# Results

DUREl consists of 22 German words annotated by human experts for semantic change across 2 centuries.

| Method | Measure | DURel |
|---|---|---|
| NN [1] | spearman | 0.59 |
| Proposed | spearman | **0.77** |

**Table 1:** Results on diachronic corpus DURel [4]

## Conclusion & Future Work

- We propose a manifold optimization based geometric framework for the task of LSC Detection.

# Conclusion & Future Work

- We propose a manifold optimization based geometric framework for the task of LSC Detection.
- Preliminary results suggest that posing the alignment as a classification problem leads to better results.

# Conclusion & Future Work

- We propose a manifold optimization based geometric framework for the task of LSC Detection.
- Preliminary results suggest that posing the alignment as a classification problem leads to better results.
- We hope to extend the approach by testing it on more standard datasets and use the aligned embeddings in downstream tasks like NER.

H. Gonen, G. Jawahar, D. Seddah, and Y. Goldberg.
**Simple, interpretable and stable method for detecting words with usage change across corpora.**
In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 538–555, 2020.

W. L. Hamilton, J. Leskovec, and D. Jurafsky.
**Diachronic word embeddings reveal statistical laws of semantic change.**
In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Aug. 2016.

P. Jawanpuria, A. Balgovind, A. Kunchukuttan, and B. Mishra.
**Learning multilingual word embeddings in latent metric space: A geometric approach.**
*Transactions of the Association for Computational Linguistics*, pages 107–120, 2019.

D. Schlechtweg, A. Hätty, M. del Tredici, and S. Schulte im Walde.
**A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains.**
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, 2019.