

# 不均衡最適輸送を用いた意味変化検出

岸野稜<sup>1</sup> 山際宏明<sup>1</sup> 永田亮<sup>2,5</sup> 横井祥<sup>3,4,5</sup> 下平英寿<sup>1,5</sup>

<sup>1</sup> 京都大学 <sup>2</sup> 甲南大学 <sup>3</sup> 国立国語研究所 <sup>4</sup> 東北大学 <sup>5</sup> 理化学研究所

kishino.ryo.32s@st.kyoto-u.ac.jp, hiroaki.yamagiwa@sys.i.kyoto-u.ac.jp,

nagata-acl2025@ml.hyogo-u.ac.jp, yokoi@ninja1.ac.jp, shimo@i.kyoto-u.ac.jp

## 概要

新旧のコーパスから得られる対象単語の文脈付き単語埋め込みの集合を比較することで、時代の経過に伴う単語の意味変化を検出できるようになった。ただし既存手法は、対象単語の意味が全体としてどのくらい変化したかを測るものであり、個々の使用事例に踏み込んだ分析には至っていない。本稿では、新旧コーパスにおける文脈付き単語埋め込みの集合の間に不均衡最適輸送を適用し、使用事例間のアライメントの過不足に注目することで、語義の消失・出現を自然に定式化できることを示す。特に、対象単語の使用事例ごとに、その事例における語義での単語使用頻度の変化を測る Sense Usage Shift (SUS) という量を導入する。SUS を用いることで、使用事例ごとの意味変化の定量化が可能であることを、評価実験を通して示す。

## 1 はじめに

時代とともに意味が変化した単語を特定し、具体的にどの語義が消失または出現したのかを同定することを意味変化検出と呼ぶ [2]。深層学習が自然言語処理に導入されて以後、新旧のコーパスから得られる単語埋め込みを用いた意味変化検出手法が注目を集めている。従来手法 [3, 4, 5] は、新旧コーパスから得られる対象単語の文脈付き埋め込みの集合間の差異を通して、単語の意味変化を検出する。

これらの手法は、対象単語の使用事例集合全体の変化に着目している。一方で、言語学的な分析では、個々の使用事例における詳細な情報も重要となる。例えば、対象単語のある語義での使用頻度が他の語義と比較してどの程度激しく変化したかを定量化できれば、対象単語の意味変化の様子や理由を見出すことができる。

そこで本稿では、新旧コーパスから得られる対象

本稿は [1] を改変したものである。

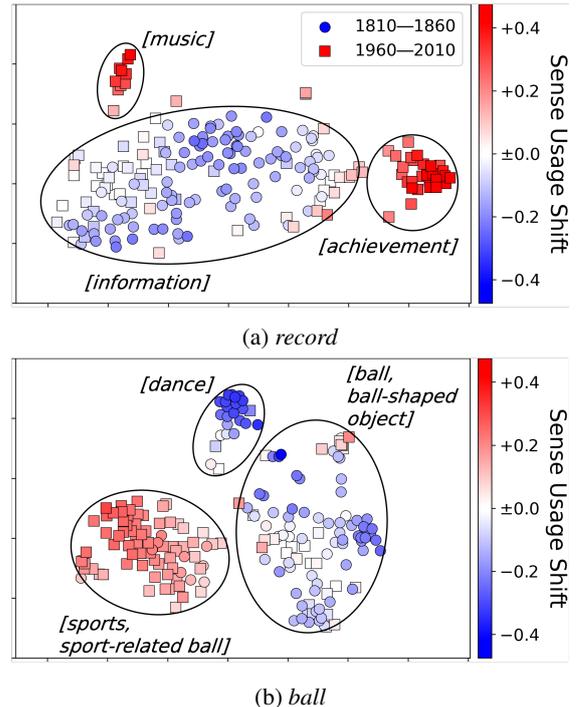


図 1: 新旧コーパスにおける各使用事例に関する対象単語の文脈付き埋め込みの t-SNE による可視化。各事例の色はその Sense Usage Shift (SUS) を表す。赤い事例の語義での単語使用頻度は他の語義での使用頻度に比べて増加し、青い事例の語義での使用頻度は減少していることを表す。詳細は 4.3 節を参照。

単語の文脈付き埋め込みの集合間に不均衡最適輸送 (UOT) を適用し、使用事例間のアライメントの過不足に着目することで、各使用事例の意味変化の様子を明らかにする。特に、意味変化の度合いを定量化するために Sense Usage Shift (SUS) という新しい指標を提案する。SUS は、各使用事例に対応する語義の相対的な使用頻度が、新旧のコーパス間でどの程度変化したかを表す量である。

図 1 において赤色で示した点群は、SUS 値の大きい、すなわち新コーパスで使用頻度が増加・出現した語義をもつ使用事例を表す。青色は、逆に消失の

表 1: 対象単語 *record* の使用事例のうち, SUS 値の上位と下位 3 件. SUS 値が大きい/小さい事例は, 新旧コーパス間において使用頻度が増加/減少した語義をもつ. 語義 ‘undefined’ については, 付録 A を参照されたい.

	年代	使用事例	語義	SUS
Top 3	1960–2010	... So did Sire <b>Records</b> ...	undefined ( <i>[music]</i> )	0.47
	1960–2010	... a team with the third-worst <b>record</b> ...	<i>[achievement]</i>	0.45
	1960–2010	... the AMCU single-season <b>record</b> ...	<i>[achievement]</i>	0.45
Bottom 3	1810–1860	... interpretations of the Mosaic <b>record</b> ...	<i>[information]</i>	-0.23
	1810–1860	... the <b>records</b> of a professed revelation...	<i>[information]</i>	-0.24
	1810–1860	... the <b>record</b> of whose wisdom is included in...	<i>[information]</i>	-0.25

傾向にある語義をもつ使用事例を表す. 表 1 には具体的な SUS 値の上位下位の使用事例を示した. 3.3 節で詳細に述べるが, 図 2 に例示するように, 通常最適輸送 (OT) を用いた従来手法 [4] では, 語義の出現・消失を適切にモデリングできない.

さらに, 実際の言語データを用いた評価実験で, SUS が事例ごとの意味変化を測る上で従来手法よりも有用であることを示す.

## 2 関連研究

### 2.1 意味変化検出

意味変化検出の主要な手法の 1 つに静的な単語埋め込みを用いる手法 [6, 7] がある. しかし, これらの手法は, 単語ごとの意味変化しか扱えず, 事例ごとの変化についての情報を得られない.

そこで, BERT [8] のような言語モデルから得られる, 新旧コーパスにおける対象単語の文脈付き単語埋め込みを利用する方法が提案されている. 新旧の埋め込みの集合をクラスタリングする手法 [3, 9] や, 埋め込みの集合を確率分布と見なして分布間の差異を測る方法 [5, 10] がある. 手法 [10] は von Mises-Fisher (vMF) 分布の対数密度比 (LDR) を用いて事例ごとの変化を捉える手法を提案した. しかし, 我々の知る限り, 事例ごとの変化についての定量評価をした先行研究は存在しない.

### 2.2 自然言語処理における最適輸送

OT は, 点群間の過不足のないアラインメントを通じて 2 つの確率分布間の距離を測る手法である. 一方で, UOT はアラインメントの過不足を許容する. OT, UOT は単語埋め込みに基づき 2 つの文章間の距離を計算するために広く用いられている [11, 12, 13].

[4, 14] は, OT を用いた意味変化検出手法を提案した. しかし, 我々の知る限り, UOT を意味変化検

出に利用した先行研究は存在しない.

## 3 背景：意味変化検出と最適輸送

### 3.1 問題設定

新旧コーパス間で対象単語  $w$  の意味がどのように変化したかを特定したい. 対象単語  $w$  を含む文脈を  $w$  の使用事例と呼ぶ. 簡単のため, 単に事例とも呼ぶ. 使用事例における  $w$  の語義を事例の語義と呼ぶ. 旧コーパスに属する対象単語  $w$  の  $m$  個の使用事例の集合を  $\{s_i^w\}_{i=1}^m$ , 新コーパスに属する  $n$  個の使用事例の集合を  $\{t_j^w\}_{j=1}^n$  とする. 事前学習済みの言語モデルを用いて, 各使用事例  $s_i^w, t_j^w$  に関する  $w$  の文脈付き埋め込み  $u_i^w, v_j^w \in \mathbb{R}^d$  を計算する. 以降, 上付き添字  $w$  は省略する.  $w$  の意味がコーパス間でどのように変化したかを特定するために, 分布  $\{u_i\}_{i=1}^m$  と  $\{v_j\}_{j=1}^n$  の差異を測る.

### 3.2 最適輸送

$n \in \mathbb{N}$  に対して,  $\mathbf{1}_n$  を全ての要素が 1 である  $n$  次元のベクトルとし,  $\mathbb{R}_+ = [0, \infty)$  とする. 文脈付き単語埋め込み  $u_i, v_j$  に対応する重みをそれぞれ  $a_i, b_j \in \mathbb{R}_+$  とし, これらが  $\sum_i a_i = \sum_j b_j = 1$  を満たすとする. このとき, 重み付きの埋め込みは確率分布とみなすことができる. また, 埋め込み  $u_i, v_j$  間の輸送コストを  $C_{ij} \in \mathbb{R}_+$ , 輸送量を  $T_{ij} \in \mathbb{R}_+$  とする. 総輸送コストを最小化する最適輸送 (OT) 問題は次のように定式化される:

$$\begin{aligned} \min_{T \in \mathbb{R}_+^{m \times n}} \sum_{i,j} T_{ij} C_{ij} \\ \text{s.t. } T \mathbf{1}_n = \mathbf{a}, \quad T^T \mathbf{1}_m = \mathbf{b}. \end{aligned} \quad (1)$$

ここで,  $\mathbf{a} = (a_1, \dots, a_m)^T, \mathbf{b} = (b_1, \dots, b_n)^T, T = (T_{ij})$  とおいた. 決定変数  $T$  は輸送行列と呼ばれ, 埋め込み集合間のアラインメントを与える. さらに, カップリング制約  $T \mathbf{1}_n = \mathbf{a}, T^T \mathbf{1}_m = \mathbf{b}$  は, 埋め込み集合間のアラインメントに過不足がないことを表す.

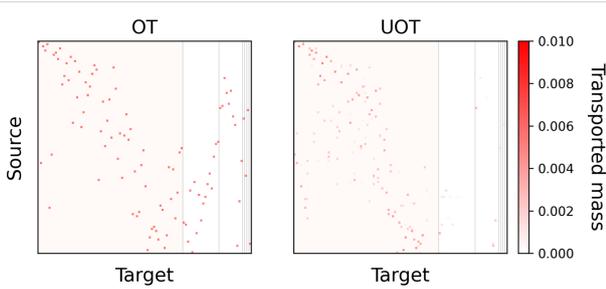


図 2: 対象単語 *record* について, OT と UOT の最適な輸送行列. 薄い赤色の箇所は, 新旧コーパス間で同じ語義をもつ事例間のアラインメントを表し, 白の箇所は異なる語義間のアラインメントを表す. (左) OT は異なる語義間での輸送を頻繁に行う. (右) UOT は同一語義間の輸送に制限される.

### 3.3 最適輸送の限界

[14] は最適な  $T$  を用いて, 総輸送コスト  $\sum_{i,j} T_{ij} C_{ij}$  の値を単語ごとの意味変化度として利用した. 通常の OT は埋め込み間の完全なアラインメントをとる. しかし, 対象単語  $w$  の意味が変化するとき, 新しい語義の出現や既存の語義の消失といった変化がコーパス間で生じる. OT の完全なアラインメントでは, このような大規模な意味変化を十分に捉えることができない場合がある. 図 2 の実際に計算された最適な輸送行列  $T$  の例からもわかるように, OT は異なる語義間での輸送を頻繁に行う.

## 4 提案手法

### 4.1 UOT による意味変化のモデル化

OT 問題 (1) のカップリング制約による完全なアラインメントは現実世界の意味変化を必ずしも反映するわけではない. そこで, カップリング制約を緩和した以下の問題を考える:

$$\min_{T \in \mathbb{R}_+^{m \times n}} \sum_{i,j} T_{ij} C_{ij} + \lambda_1 D_1(T \mathbf{1}_n, \mathbf{a}) + \lambda_2 D_2(T^T \mathbf{1}_m, \mathbf{b}). \quad (2)$$

ここで,  $D_1: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$  および  $D_2: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  はペナルティ関数であり,  $\lambda_1, \lambda_2$  はペナルティの大きさを制御するハイパーパラメータである. この問題 (2) は不均衡最適輸送 (UOT) 問題と呼ばれる. UOT は, 一定のコストを負担することでアラインメントの過不足を許容する.

いま, 式 (2) における  $T$  で表されるアラインメントの過不足に注目する. 図 2 の例より, UOT にお

ける輸送は同一の語義内に制限される傾向があり, アラインメントの過不足が各語義の使用頻度の変化を反映する.  $s_i$  について, 新コーパスに輸送される重みが元の重み  $a_i$  より少ない場合,  $s_i$  の語義をもつ事例が他の  $w$  の語義に比べて新コーパスでは相対的に少ないことを示す. つまり,  $s_i$  の語義での  $w$  の使用頻度が減少していることを意味する. 同様に,  $t_j$  について, 旧コーパスから受け取る重みが元の重み  $b_j$  より少ない場合,  $t_j$  の語義をもつ事例は旧コーパスでは相対的に少ないことを表す. つまり,  $t_j$  の語義での  $w$  の使用頻度が増加したことを意味する.

### 4.2 Sense usage shift

UOT によるアラインメントの過不足を通じて意味変化を定量的に測る. 各使用事例  $s_i$  または  $t_j$  の語義での単語使用頻度が他の語義での使用頻度と比較してどのように変化したかを測る指標として, Sense Usage Shift (SUS) を次のように定義する:

$$\text{SUS}(s_i) = -(a_i - \sum_j T_{ij}) / a_i, \quad (3)$$

$$\text{SUS}(t_j) = (b_j - \sum_i T_{ij}) / b_j. \quad (4)$$

つまり, SUS はアラインメントの過不足を元の重みで正規化した値である. 4.1 節より, ある事例の SUS 値が大きいほど, 新コーパスにおいて, その事例の語義で対象単語がより頻繁に使用されることを表す. 逆に, SUS 値が小さい場合は, その語義での使用頻度が減少したことを表す. 特に式 (3) において, SUS 値が正の場合に, その事例の語義での使用頻度が増加したことを表すために符号を反転した.

### 4.3 SUS による意味変化の解析

図 1 は, 対象単語 *record* と *ball* の新旧 100 個ずつの文脈付き埋め込みを t-SNE を用いて 2 次元に可視化したものである<sup>1)</sup>. データセットの詳細は 5 節を参照. この図において, クラスタ<sup>2)</sup>は, 対象単語の語義を表す. クラスタ内の使用事例を分析し, それらの語義を特定した. 図中のクラスタから, 使用頻度が増加または出現した語義をもつ使用事例の SUS 値は大きく, 使用頻度が減少または消失した語義をもつ使用事例の SUS 値は小さいことが確認できる.

1) t-SNE の設定では, scikit-learn の実装におけるデフォルト値である perplexity=30 を使用した.

2) 図 1 のクラスタは著者らが検討した. 全ての事例を厳密に確認したわけではないが, 各クラスタ内で語義の一貫性が保たれるよう努めた.

## 5 評価実験

### 5.1 準備

**データセット.** 実験には、英語用の Diachronic Word Usage Graph (DWUG) [15, 16] データセット ver. 3 を使用した。このデータセットには 46 個の対象単語が含まれており、各単語について、旧時代 (1810–1860) および新時代 (1960–2010) における約 100 個の使用事例が収録されている。DWUG の詳細については、付録 A を参照されたい。

**埋め込みの計算.** 文脈付き単語埋め込みを計算するために、XL-LEXEME [17] を使用した。

**UOT の設定.** 各使用事例に対して一様重みを設定した： $\mathbf{a} = \left(\frac{1}{m}, \dots, \frac{1}{m}\right)^T$ ,  $\mathbf{b} = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)^T$ 。事例間の輸送コストはコサイン距離を用いた： $C_{ij} = 1 - \cos(\mathbf{u}_i, \mathbf{v}_j)$ 。アラインメントの過不足のペナルティには L2 誤差を用いた： $D_1(\mathbf{T}\mathbf{1}_n, \mathbf{a}) = \|\mathbf{T}\mathbf{1}_n - \mathbf{a}\|_2^2$ ,  $D_2(\mathbf{T}^T\mathbf{1}_m, \mathbf{b}) = \|\mathbf{T}^T\mathbf{1}_m - \mathbf{b}\|_2^2$ 。UOT の先行研究 [18] に従い、 $\lambda_1 = \lambda_2 = \lambda$  と設定した。 $\lambda$  は全ての対象単語で一貫して等しい値を使用した。4.3 節における定性分析では、 $\lambda = 100$  に固定した。一方、5.2 節での定量評価では、検証データセットを用いて  $\lambda$  を調整した。詳細は付録 B を参照されたい。

### 5.2 実験：事例ごとの意味変化の定量化

本節では、SUS が使用事例ごとの意味変化<sup>3)</sup>の定量化において有効であるかを評価する。SUS を用いて、各語義での単語使用頻度の変化度を同定する。

対象単語の新旧コーパスにおける語義を  $k = 1, \dots, K$  とし、旧, 新コーパスでの語義  $k$  の使用頻度をそれぞれ  $X_k, Y_k$  とする。いま、 $X = (X_1, \dots, X_K)$ ,  $Y = (Y_1, \dots, Y_K)$  を語義頻度分布 (Sense Frequency Distribution; SFD) と呼ぶ。SFD  $X, Y$  に対して、語義  $k$  をもつ事例  $s_i$  の変化度を  $\tau(s_i; X, Y) = \log \frac{Y_k}{X_k}$  と定義する<sup>4)</sup>。同様に、事例  $t_j$  についても  $\tau(t_j; X, Y)$  を定める。

DWUG において、対象単語ごとに gold SFD  $X^*, Y^*$  が与えられている。対象単語の使用事例における gold 変化度を、 $\tau^*(\cdot) = \tau(\cdot; X^*, Y^*)$  として定義する。提案手法では SUS 値を用いて、 $\tau_{\text{SUS}}(\cdot) = \text{SUS}(\cdot)$  とする。ベースラインとして、WiDiD [9] を用いて埋

3) 厳密には対象単語の語義ごとに変化であるが、本稿では便宜的に「事例ごと」の意味変化と呼ぶ。

4)  $X_k = 0$  または  $Y_k = 0$  の場合、それぞれ全単語における変化度の最小値または最大値で補完する。

表 2: 事例ごとの意味変化を定量化する手法の性能.

手法	Spearman
$\tau_{\text{SUS}}$	<b>0.46</b>
$\tau_{\text{LDR}}$	0.40
$\tau_{\text{WiDiD}}$	0.31

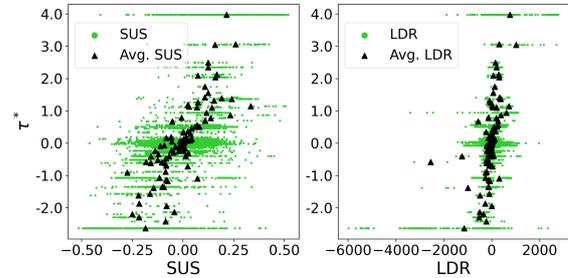


図 3: DWUG から計算された gold 変化度  $\tau^*$  と、各事例に対する SUS (左) および LDR (右) の値との関係。各  $\tau^*$  に対する横軸の値の平均値は  $\blacktriangle$  で示す。

め込みをクラスタリングし、SFD  $\hat{X}, \hat{Y}$  を推定する。そして、 $\tau^*$  の推定値を  $\tau_{\text{WiDiD}}(\cdot) = \tau(\cdot; \hat{X}, \hat{Y})$  とする。また、他のベースラインとして、 $\tau_{\text{LDR}}(\cdot)$  を新旧の推定された vMF 分布に関する事例の LDR とする。LDR の詳細については付録 C を参照されたい。

表 2 に、gold 変化度と各手法による変化度の Spearman 順位相関を示す。SUS が他の手法よりも各事例の語義での単語使用頻度の変化をより正確に予測している。

図 3 は、対象単語のすべての使用事例について、SUS および LDR の値を横軸に、それに対応する gold 変化度  $\tau^*$  を縦軸にプロットしたものである。同じ語義をもつ事例は同じ  $\tau^*$  を共有する。そこで、等しい  $\tau^*$  をもつ事例の SUS 値, LDR 値の平均値と gold 変化度との Spearman 順位相関を計算した。SUS は 0.85 の Spearman 順位相関係数であり、LDR は 0.74 であった。従って、SUS は各事例の語義での使用頻度の変化をより捉えられている。

## 6 結論

本稿では、新旧コーパスにおける対象単語の文脈付き埋め込み集合間に UOT を適用し、使用事例間のアラインメントの過不足を活用することを提案した。特に、各使用事例に対して SUS と呼ばれる新しい指標を定義し、その語義での単語使用頻度の変化を定量化した。そして、SUS の有効性を意味変化検出タスクにおける実験を通して定量的に評価した。

## 謝辞

本研究は、JSPS 科研費 22H05106, 23H03355, 23K24910, JST CREST JPMJCR21N3, JST SPRING JPMJSP2110, JST FORET JPMJFR2331 の助成を受けたものです。

## 参考文献

- [1] Ryo Kishino, Hiroaki Yamagiwa, Ryo Nagata, Sho Yokoi, and Hidetoshi Shimodaira. Quantifying lexical semantic shift via unbalanced optimal transport. *CoRR*, Vol. abs/2412.12569, , 2024.
- [2] Francesco Periti and Stefano Montanelli. Lexical semantic change through large language models: a survey. *ACM Comput. Surv.*, Vol. 56, No. 11, pp. 282:1–282:38, 2024.
- [3] Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. Analysing lexical semantic change with contextualised word representations. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020**, pp. 3960–3973. Association for Computational Linguistics, 2020.
- [4] Syrielle Montariol, Matej Martinc, and Lidia Pivovarova. Scalable and interpretable semantic change detection. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021**, pp. 4642–4652. Association for Computational Linguistics, 2021.
- [5] Taichi Aida and Danushka Bollegala. Unsupervised semantic variation prediction using the distribution of sibling embeddings. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 6868–6882, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [6] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In Katrin Erk and Noah A. Smith, editors, **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1489–1501, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [7] Taichi Aida, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi. A comprehensive analysis of PMI-based models for measuring semantic differences. In Kaibao Hu, Jong-Bok Kim, Chengqing Zong, and Emmanuele Chersoni, editors, **Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation**, pp. 21–31, Shanghai, China, 11 2021. Association for Computational Linguistics.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Tamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)**, pp. 4171–4186. Association for Computational Linguistics, 2019.
- [9] Francesco Periti, Alfio Ferrara, Stefano Montanelli, and Martin Ruskov. What is done is done: an incremental approach to semantic shift detection. In Nina Tahmasebi, Syrielle Montariol, Andrey Kutuzov, Simon Hengchen, Haim Dubossarsky, and Lars Borin, editors, **Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, LChange@ACL 2022, Dublin, Ireland, May 26-27, 2022**, pp. 33–43. Association for Computational Linguistics, 2022.
- [10] Ryo Nagata, Hiroya Takamura, Naoki Otani, and Yoshifumi Kawasaki. Variance matters: Detecting semantic differences without corpus/word alignment. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 15609–15622, Singapore, December 2023. Association for Computational Linguistics.
- [11] Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. From word embeddings to document distances. In Francis R. Bach and David M. Blei, editors, **Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015**, Vol. 37 of **JMLR Workshop and Conference Proceedings**, pp. 957–966. JMLR.org, 2015.
- [12] Sho Yokoi, Ryo Takahashi, Reina Akama, Jun Suzuki, and Kentaro Inui. Word rotator’s distance. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020**, pp. 2944–2960. Association for Computational Linguistics, 2020.
- [13] Yuki Arase, Han Bao, and Sho Yokoi. Unbalanced optimal transport for unbalanced word alignment. In Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023**, pp. 3966–3986. Association for Computational Linguistics, 2023.
- [14] Marko Pranjic, Kaja Dobrovoljc, Senja Pollak, and Matej Martinc. Semantic change detection for slovene language: a novel dataset and an approach based on optimal transport. *CoRR*, Vol. abs/2402.16596, , 2024.
- [15] Dominik Schlechtweg, Nina Tahmasebi, Simon Hengchen, Haim Dubossarsky, and Barbara McGillivray. DWUG: A large resource of diachronic word usage graphs in four languages. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 7079–7091, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [16] Dominik Schlechtweg, Pierluigi Cassotti, Bill Noble, David Alfter, Sabine Schulte im Walde, and Nina Tahmasebi. More dwugs: Extending and evaluating word usage graph datasets in multiple languages. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 16-16, 2024**, pp. 14379–14393. Association for Computational Linguistics, 2024.
- [17] Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. XL-LEXEME: WiC pretrained model for cross-lingual LEXical sEMantic changeE. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 1577–1585, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [18] Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. In Marc’Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, **Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual**, pp. 23270–23282, 2021.
- [19] Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann. Diachronic usage relatedness (durel): A framework for the annotation of lexical semantic change. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)**, pp. 169–174. Association for Computational Linguistics, 2018.
- [20] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, Vol. 22, No. 78, pp. 1–8, 2021.

表 3: 全対象単語を使用したときの 5.2 節のタスクの性能.

提案量	$\lambda$			
	1	10	100	1000
$\tau_{\text{SUS}}$	0.31	0.42	<b>0.50</b>	0.49

表 4: 5.2 節のタスクにおいて, 100 回の分割で, 検証データで最も多く選ばれた  $\lambda$  の値.

提案量	最も選ばれた $\lambda$	回数
$\tau_{\text{SUS}}$	100	73

## A DWUG の詳細

本付録では, DWUG [15, 16] についての詳細を示す. DWUG データセット<sup>5)</sup>では, 各対象単語の十分な数の使用事例ペアに, DUREL [19] と呼ばれる類似度スコアが人間により与えられている. DWUG は, これらの類似度を枝重み, 事例を頂点とするネットワークをクラスタリングすることで, 各使用事例の語義を同定し, 新旧の SFD を導出する. 本稿では, DWUG の語義を十分正確なものとして扱った.

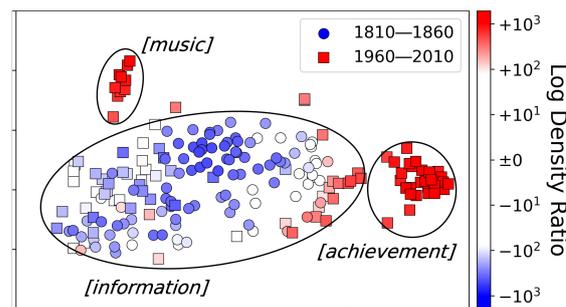
DWUG データセットの仕様上, 各対象単語の少数の使用事例には語義が割り当てられていない場合がある. 本稿では, このような語義を *undefined* と呼ぶ. しかし, 著者らが容易に語義を特定できる場合には, 欠損した語義を補完する. 例えば, 表 1 の *[music]* などである.

## B UOT の設定の詳細

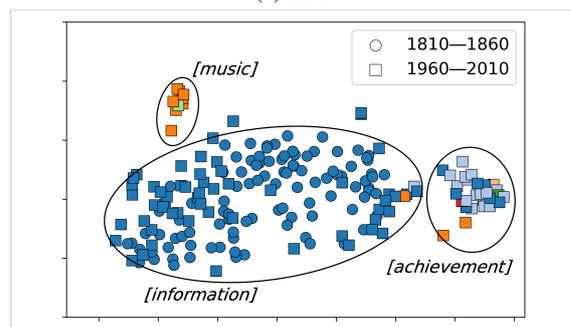
UOT の実装には Python Optimal Transport [20] が提供する MM アルゴリズムを使用した.

UOT (2) におけるハイパーパラメータ  $\lambda_1 = \lambda_2 = \lambda$  は, 5.2 節での性能評価の際に検証データ上でチューニングする必要がある.  $\lambda$  の探索範囲を決定するため, まず全ての対象単語を用いて性能評価を行う.  $\lambda$  を 1, 10, 100, 1000 の値で変化させた結果を表 3 に示す. これに基づき,  $\lambda \in \{10, 20, 50, 100, 200, 500, 1000\}$  を探索する. そして, 検証データ上でのチューニング時には, DWUG データセットに含まれる 46 個の対象単語を, 検証データとテストデータにランダムに 8:2 の比率で分割する. 検証データにおいて, 前述の探索範囲で最適な  $\lambda$  を決定する. WiDiD の場合は, *damping* と呼

5) <https://zenodo.org/records/14028531>



(a) LDR



(b) DWUG

図 4: 図 1a の再描画. 使用事例の色は (a) LDR 値, (b) DWUG における gold 語義を表す.

ばれるハイパーパラメータを  $\{0.5, 0.6, 0.7, 0.8, 0.9\}$  を探索する. その後, テストデータ上で性能評価を実施する. 一連のデータセットの分割を 100 回繰り返す. 各手法の最終的な性能は全ての分割におけるテスト性能の平均値として計算する. 表 4 に, 最も多く選ばれた  $\lambda$  の値を示す.

## C 対数密度比

本稿では, [10] の式 (8) から直接 LDR を計算した. 特に, 同論文の式 (5) を用いて旧, 新コーパスにおける vMF 分布の集中度パラメータの近似的な最尤推定量を計算し, *scipy* の *vonmises\_fisher* を用いて 2 つの vMF 分布の確率密度関数を計算した.

図 4a に, 図 1a を LDR を用いて色付けしたものを示す. 図 1a と図 4a の比較では, SUS と LDR のどちらも大まかには語義クラスタを適切に反映していることが分かる. ただし, DWUG の gold 語義で色を付けた図 4b から分かるように, 語義 *[information]* に対応するクラスタの最右部を LDR は異なるクラスタとしている.

一般に, 高次元空間での LDR の推定は, 密度比の大きな変動のために困難が多い. これに対して, SUS は UOT を介するノンパラメトリックな手法であり, 密度比推定を迂回していると考えられる.