

ガウス過程による埋め込み点集合の時間遷移のモデル化

相田太一¹ 小町守² 小木曾智信³ 高村大也⁴ 持橋大地^{3,5}

¹ 東京都立大学 ² 一橋大学

³ 国立国語研究所 ⁴ 産業技術総合研究所 ⁵ 統計数理研究所

aida-taichi@ed.tmu.ac.jp mamoru.komachi@r.hit-u.ac.jp

togiso@ninja1.ac.jp takamura.hiroya@aist.go.jp daichi@ism.ac.jp

概要

単語の意味は時間とともに変化することがある。近年では、この時間的変動を捉えるため、埋め込み空間上で用例集合を分析する研究が数多く行われてきた。同様の時間的変動は、生態系の生息分布や社会学の犯罪発生分布など、他の多くの分野にも存在する。しかし、このような点集合の動的な変化は非常に複雑であり、解析が困難だという問題がある。本研究では、ガウス過程を用いて点集合を一つの複雑な分布として表現し、それを周波数空間の実ベクトルとしてコンパクトに表すことで、点集合の時間遷移を解析する手法を提案する。提案手法で単語の意味変化を分析し、応用例として社会学の空間データにも適用することで、その有用性を確認した。

1 はじめに

動的に変化する点集合の解析は、さまざまな分野において重要な課題である。例えば言語学では、特定の単語の意味が時間とともにどのように変化するか [1] について、用例から埋め込みを獲得し、その集合を分析する [2, 3] ことで、言語の変遷を理解する手助けとなる。また、生態系においても、植物・草食動物・肉食動物の相互作用により、各種の生息地は時間とともに変化する [4] が、これらの時間的遷移を解析する [5] ことで、生態系の動向を明らかにすることが期待できる。社会学の分野では、人口移動や犯罪の発生などの社会現象が、特定の地域にどのように発生するかを時系列で解析する [6, 7] ことで、社会現象の変動パターンを理解することができる。このように、空間的に分布した点集合の時間遷移を理解することは、さまざまな現象の背後にある構造や相互作用を明らかにするための基盤となる。

これまでの研究では、点集合の時間的変化を扱う手法として、ある時点での点集合（例：単語の

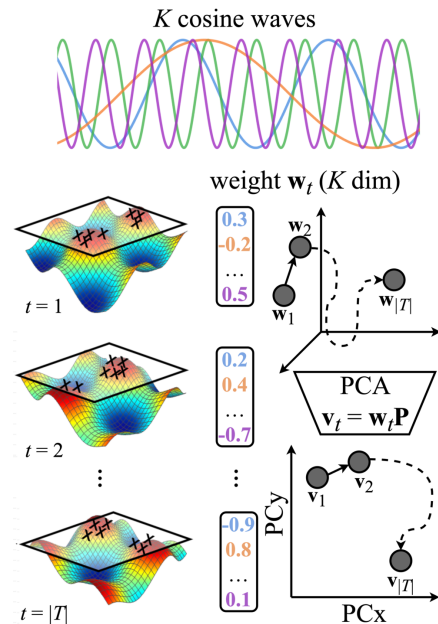


図 1: 概要図. K 個の余弦波をランダムに指定し、時期 $t \in T$ の点集合 ($\mathbb{R}^{N \times 2}$) について K 個の余弦波に対応する重みの推定を行う。これによって、任意の点集合の時間遷移 ($\mathbb{R}^{|T| \times (N \times 2)}$) を K 次元の重みの時間遷移 ($\mathbb{R}^{|T| \times K}$) として解釈することが可能となる。さらに、重みの集合に PCA を適用し、2つの主成分を選択して変換 ($\mathbb{R}^{|T| \times K} \rightarrow \mathbb{R}^{|T| \times 2}$) することで、点集合の時間遷移をコンパクトに表現できる。

用例埋め込み集合、ある生物の生息分布、犯罪の発生分布) を推定する手法が広く用いられてきた [5, 6, 7, 8, 9, 2]。しかし、これらの手法は特定の時点での各地の状態を表現するには有効であるものの、点集合全体の時間的な変動のメカニズムや異なる点集合間の関係性を理解するには限界がある。

本研究では、この問題に対処するために、ガウス過程を用いて各時期の点集合を一つの分布として表現し、時間による変動を解析するための新しい手法を提案する。通常のガウス過程はデータ点の情

報を全て保持することで複雑な分布を表現するが、分布の本質的な情報を抽出することは困難である。そこで本研究では、ガウス過程をコンパクトに表現する**乱択化フーリエ特徴** (Random Fourier Features, **RFF**) [10]を採用した。RFFは K 個の余弦波の組み合わせでデータ点の分布を近似するため、複雑な分布の本質的な情報を K 次元の周波数空間の実ベクトルとして表現することが可能となる。実験では、単語の意味変化を分析し、応用として社会学の犯罪発生分布にも適用することで、有用性を検証した。

2 RFF による点集合の遷移追跡

本研究では、対象とする空間 \mathcal{X} 上の点の密度を、次のガウス過程密度サンプラー [11]による複雑な確率密度として表現する。

$$p(\mathbf{x}) = \frac{\sigma(f(\mathbf{x}))}{Z}, \quad Z = \int_{\mathcal{X}} \sigma(f(\mathbf{x})) d\mathbf{x} \quad (1)$$

ここで $\sigma(x)$ はシグモイド関数 $\sigma(x) = 1/(1+e^{-x})$ で、 f はガウス過程 $GP(0, k(\mathbf{x}, \mathbf{x}'))$ からのサンプル、 k は \mathcal{X} 上の点 \mathbf{x} の類似度を表すカーネル関数である。 \mathcal{X} が2次元の場合の例を、[図2](#)に示した。

式(1)による点密度は正規化定数 Z が空間全体での積分を含むため、そのままでは計算できない。しかし、機械学習分野で提案された対照学習 [12]の方法を用いることで、式(1)の解である f は Z 自体も推定する未知変数の一つとして、次の式(2)を最大化する解として求めることができる。

$$\sum_{\mathbf{x} \in \mathcal{D}} \log p(\mathbf{x}) + \sum_{\mathbf{x}' \in \overline{\mathcal{D}}} \log(1-p(\mathbf{x}')) \quad (2)$$

ここで \mathcal{D} は観測された正例の点集合、 $\overline{\mathcal{D}}$ はランダムに生成された負例の点集合である。

このとき、式(1)でのガウス過程 f は通常は、正例と負例のデータ点における値 $f(\mathbf{x})$ の集合 $\mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^T$ としてノンパラメトリックに表現される。しかし、これには N 個のデータ点の間のカーネル行列を計算する必要があるため計算量が $O(N^3)$ となる上、**分布の大局的な情報を陽に含んでいない**という問題がある。

一方、ガウス過程のカーネル関数 $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$ は、この関数をフーリエ変換して積分表示し、積分をモンテカルロ積分で置き換えることで、特徴関数 $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_K(\mathbf{x}))^T$ をカーネルトリックを介さず、ガウスカーネルの場合は

$$\phi_k(\mathbf{x}) = \sqrt{2} \cos(\omega_k^T \mathbf{x} + b_k) \quad (3)$$

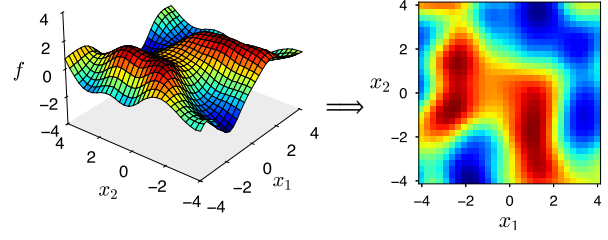


図2: 2次元のガウス過程密度サンプラー [11]の概要。ガウス過程から生成された関数 f を値が正になるように変換して正規化し確率密度にすることで、非常に複雑な密度関数を表現することができる。

$$\omega_k \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad b_k \sim \text{Unif}[0, 1]$$

と陽に近似できることが知られている(**乱択化フーリエ特徴 (RFF)** [10])。ここで、 $\omega_k \in \mathbb{R}^D$ は各特徴の周波数に対応するベクトル、 b_k は位相を表す。これは、ガウス過程からのサンプル f が、式(3)に従うランダムな周波数 ω_k をもつ **K 個の余弦波の重みつき和として表せる**ことを意味している。もともと、ガウス過程は線形回帰モデル $y = \phi(\mathbf{x})^T \mathbf{w}$ をカーネル化したものであること [13]を思い出すと、今や \mathbf{f} は、この特徴関数を用いた単なる線形回帰モデル

$$\underbrace{\begin{pmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}}_{\mathbf{f}} = \underbrace{\begin{pmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \vdots \\ \phi(\mathbf{x}_N) \end{pmatrix}}_{\Phi} \underbrace{\begin{pmatrix} w_1 \\ \vdots \\ w_K \end{pmatrix}}_{\mathbf{w}} \quad (4)$$

すなわち、 $\mathbf{f} = \Phi \mathbf{w}$ と表すことができる。ここで点集合が決まれば計画行列(特徴関数行列) Φ は定数であり、 K 次元($K \ll N$)の重みベクトル $\mathbf{w} \in \mathbb{R}^K$ によって、[図1](#)のように \mathbf{f} 、すなわち式(1)を通じて**複雑な点密度全体を表現**することが可能になる。

学習では特徴関数 $\phi_k(\cdot)$ をサンプリングして固定し、最適な回帰係数 \mathbf{w} の推定を行う。本研究では、こうして**点集合を一つの重みベクトルとして表現できる**RFFを複数の時期に適用することで、点集合の時間遷移を解析する手法を提案する。概要図を[図1](#)に示した。まず、全ての時期で統一して適用する特徴関数行列 Φ をランダムに指定する([図1上](#))。次に、各時期 $t \in \{1, 2, \dots, |T|\}$ の点集合に対してRFFを適用し、回帰係数 \mathbf{w}_t を学習する。学習にはランダムウォークMH法 [14]を用いた。全ての時期に対して同じ特徴関数を使用するため、各時期で得られる回帰係数 \mathbf{w}_t は時期間で直接比較することが可能となる([図1左下](#))。その後、得られた回

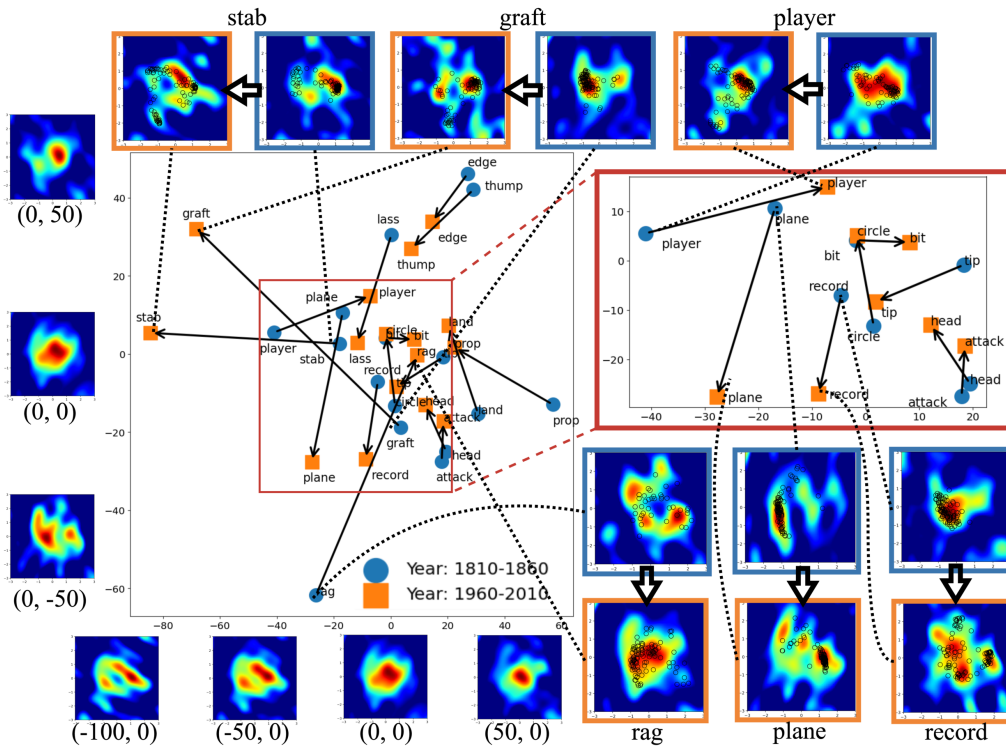


図 3: 言語データの分析結果 (横: 第 3 主成分, 縦: 第 5 主成分)。

帰係数の集合 $\{w_1, w_2, \dots, w_{|T|}\}$ に対する分析を行う。回帰係数は K 次元であり、直接解釈することが難しいため、PCA を用いて次元削減を行い、2つの主成分を選択して可視化を行うことで、点集合の時間遷移を視覚的に確認することができる (図 1 右下)。

3 分析例：単語の意味変化

まず、提案手法が単語の意味変化の種類を判別できるかを確認するため、以下の分析を行う。今回は任意の単語に関する用例埋め込み集合の時間遷移を対象とし、意味変化検出の代表的なベンチマークである SemEval-2020 Task 1 [15] の英語データセットを使用した。対象単語には評価用の単語リストの中で意味が変化した単語を使用し、用例から埋め込みの獲得にはベンチマークで最高性能を獲得した XLM-R_{large} [16] ベースのモデル [17] を使用した。得られるベクトルは 1,024 次元であるが、今回は PCA で 2 次元に圧縮して分析を行った。¹⁾ 分析では $K = 30$ 個の余弦波からなる特徴関数行列 Φ を用いて、二つの時期 $t_1 = 1810-1860$ 年, $t_2 = 1960-2010$ 年で回帰係数 w_t を推定した。

推定した回帰係数を PCA で可視化した結果

1) 提案手法は 3 次元以上にも適用可能であるが、2 次元は視覚的理解に最も適しており、低次元でも十分に意味の変化を表現できる [18] ため、今回は 2 次元を採用した。

を 図 3 に示す。²⁾ ヒートマップ中の白丸 (o) は事例 (対象単語の用例から獲得した埋め込み) を示す。

横軸 (第 3 主成分) 方向の変化 横軸方向に大きく移動した単語 stab と graft について見てみると、どちらも t_1 から t_2 にかけて負の方向に大きく移動しており、ヒートマップからは分布が広がっていることがわかる。単語 stab は「(物理的に) 刺す」という意味に「指摘する」という用法が追加され、単語 graft は「接ぎ木」という意味に対して「(医学的な) 移植」「賄賂」の用法が増えたことから、横軸の移動方向からは単語がより抽象的な概念で使用されるようになる様子を読み取ることができる。

縦軸 (第 5 主成分) 方向の変化 図 3 右側の拡大部分に位置する plane と record はどちらも t_1 から t_2 にかけて縦軸の負の方向に移動しており、ヒートマップを見ても集合が大きくなっていることが読み取れる。どちらも技術革新により意味が増えている単語 (「飛行機」, 「(音楽の) レコード」) であるが、今回の分析結果より点集合から推定した回帰係数の時間遷移も同様であることを確認できた。³⁾

2) 上位の主成分は今回の目的とは異なる時期の傾向が強く表れている (図 6) ため、対象から除外した。

3) 各軸で逆方向に移動した単語として、player と rag がある。どちらも古い時期での意味「俳優」, 「ぼろ着」が減って新たな意味「メディア再生機」, 「低俗な新聞」を獲得している。技術革新による変化 (player) は縦軸, 比喩的な拡張 (rag) は横

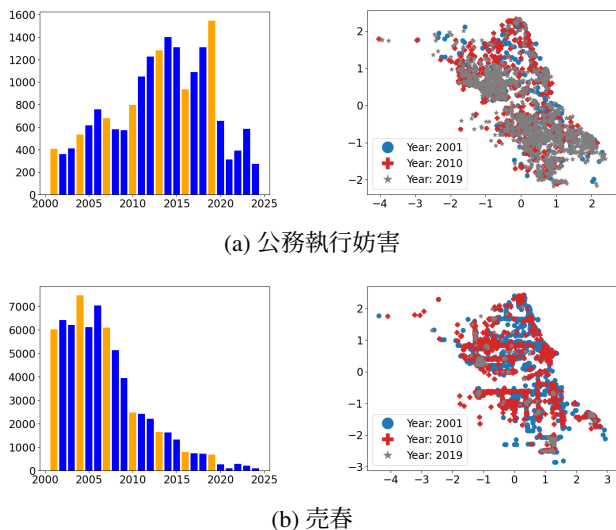


図4: シカゴの犯罪データの可視化。(左) 年別の発生件数。対象とした年を橙色で示す。(右) 対象年の初期(2001年)・中期(2010年)・後期(2019年)における発生箇所を描画した結果。

また、点集合の時間遷移を直接表現するために、各時期で推定した回帰係数の差ベクトル $w_{t_2} - w_{t_1}$ をPCAで可視化したところ、付録図7に示すように、点集合の時間遷移の種類を2次元上で分類できることが示された。この発見も今回の手法による新たな方向性であり、詳細な分析は今後の課題とする。

4 応用例：犯罪発生分布

次に、他分野における応用例として、提案手法を空間データである犯罪発生分布に適用して分析を行った。今回はシカゴの犯罪データ⁴⁾を使用した。このデータには事件ごとに犯罪カテゴリ⁵⁾と発生時刻、発生箇所(緯度・経度)などの情報が記録されている。今回は2001-2019年に発生件数が増加した「公務執行妨害」と「銃刀法違反」、減少した「売春」と「麻薬」を対象とした。年別の発生件数および発生箇所の情報を図4と図8に示す。分析では、2001年から2019年までの3年ごと、計7ヶ年を対象とし、各年から200件ずつデータを抽出した。その後、地理情報の中心化を施し、 $K=30$ の特徴関数行列 Φ を用いて各年の回帰係数 w_t の推定を行った。

軸であるが、図3ではplayerが横軸、ragが縦軸で移動していることがわかる。このように単調ではない意味の変化については、より多くの軸を用いた分析や単語ごとの適切な余弦波の設定が考えられるが、詳細は今後の課題とする。

4) <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/data>
 5) 犯罪カテゴリは36種類定義されており、一つの事件に一つの犯罪カテゴリが割り振られる。

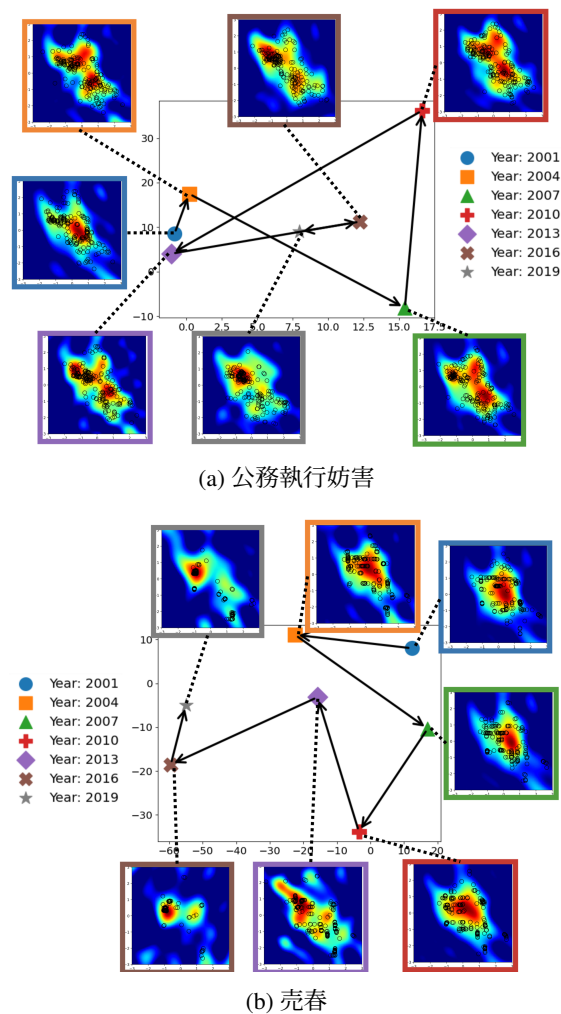


図5: 犯罪データの分析結果。

上位2つの主成分での公務執行妨害と売春に関する結果を図5に示す。どちらも横軸(第1主成分)の負の方向に遷移するが、増加傾向にある公務執行妨害は10付近で留まるのに対し、減少傾向にある売春では-50まで大きく移動していることがわかる。横軸が負の方向に移動するほど、ヒートマップでは特定の地点に集中するようになり、**犯罪が一箇所で集中的に発生する様子を読み取ることができる**。同様の現象は他2つのカテゴリでも確認された(図9)。これらの結果は、言語に限らず、任意のイベントについて発生箇所(点集合)の時間遷移を解釈する上での提案手法の有用性を示している。

5 おわりに

本研究では、点集合の時間遷移を解析するために、潜在的なガウス過程を利用した手法を提案し、有用性を確認した。今後はさらに多様なデータについて分析を行い、提案手法の有用性を検証したい。

謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2156 および JST さきがけ JPMJPR2366 の支援を受けたものです。

参考文献

- [1] Taichi Aida, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi. A comprehensive analysis of PMI-based models for measuring semantic differences. In **Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation**, pp. 21–31, Shanghai, China, 11 2021. Association for Computational Linguistics.
- [2] Taichi Aida and Danushka Bollegala. Unsupervised semantic variation prediction using the distribution of sibling embeddings. In **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 6868–6882, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [3] Ryo Nagata, Hiroya Takamura, Naoki Otani, and Yoshifumi Kawasaki. Variance matters: Detecting semantic differences without corpus/word alignment. In **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 15609–15622, Singapore, December 2023. Association for Computational Linguistics.
- [4] Eugene P Odum. Fundamentals of ecology. **Printing Company Ltd**, 1971.
- [5] Jamie M Kass, Bruno Vilela, Matthew E Aiello-Lammens, Robert Muscarella, Cory Merow, and Robert P Anderson. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. **Methods in Ecology and Evolution**, Vol. 9, No. 4, pp. 1151–1156, 2018.
- [6] Daisuke Murakami and Yoshiki Yamagata. Estimation of gridded population and gdp scenarios with spatially explicit statistical downscaling. **Sustainability**, Vol. 11, No. 7, 2019.
- [7] Daisuke Murakami, Mami Kajita, and Seiji Kajita. Scalable model selection for spatial additive mixed modeling: Application to crime analysis. **ISPRS International Journal of Geo-Information**, Vol. 9, No. 10, 2020.
- [8] David Weisburd, Laura A Wyckoff, Justin Ready, John E Eck, Joshua C Hinkle, and Frank Gajewski. Does crime just move around the corner? a controlled study of spatial displacement and diffusion of crime control benefits. **Criminology**, Vol. 44, No. 3, pp. 549–592, 2006.
- [9] Lee Hachadoorian, Stuart R. Gaffin, and Robert Engelman. **Projecting a Gridded Population of the World Using Ratio Methods of Trend Extrapolation**, pp. 13–25. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [10] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, **Advances in Neural Information Processing Systems**, Vol. 20. Curran Associates, Inc., 2007.
- [11] Iain Murray, David MacKay, and Ryan P. Adams. The Gaussian Process Density Sampler. In **NIPS 2008**, 2008.
- [12] Michael U. Gutmann and Aapo Hyvärinen. Noise-Contrastive Estimation of Unnormalized Statistical Models, with Applications to Natural Image Statistics. **Journal of Machine Learning Research**, Vol. 13, No. 11, pp. 307–361, 2012.
- [13] 持橋大地, 大羽成征. ガウス過程と機械学習. 機械学習プロフェッショナルシリーズ. 講談社, 2019.
- [14] C.M. Bishop, 元田, 栗田, 樋口, 松本, 村田 (監訳). パターン認識と機械学習: ベイズ理論による統計的予測」(上・下) (Pattern Recognition and Machine Learning). 丸善出版, 2012.
- [15] Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. SemEval-2020 task 1: Unsupervised lexical semantic change detection. In **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 1–23, Barcelona (online), December 2020. International Committee for Computational Linguistics.
- [16] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 8440–8451, Online, July 2020. Association for Computational Linguistics.
- [17] Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. XL-LEXEME: WiC pretrained model for cross-lingual LEXical sEMantic change. In **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] Taichi Aida and Danushka Bollegala. Investigating the contextualised word embedding dimensions specified for contextual and temporal semantic changes, 2024. arXiv 2407.02820. (Accepted to COLING2025).

A 単語の意味変化

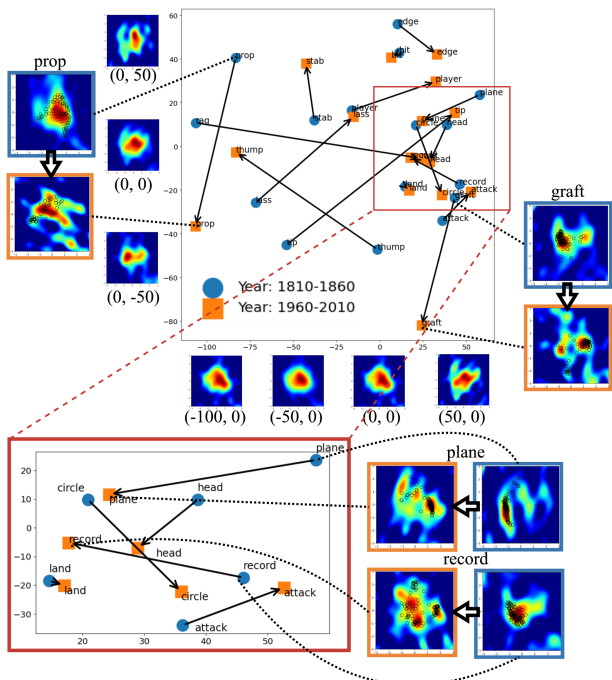


図 6: 言語データでの分析結果 (横: 第 1 主成分, 縦: 第 2 主成分). コーパスの時期の影響が強く出ているため, 分析から除外した.

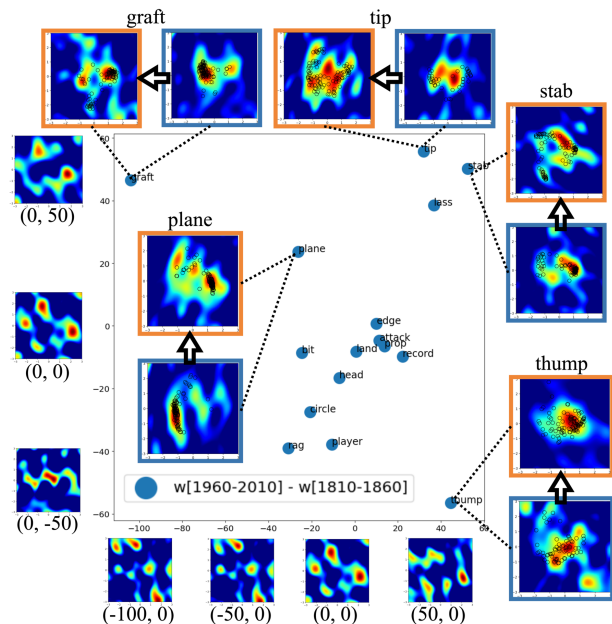


図 7: 回帰係数の差を可視化した結果. 横軸 (第 3 主成分) は意味の拡大・縮小を, 縦軸 (第 4 主成分) は意味の獲得・統合を示している.

B 犯罪発生分布

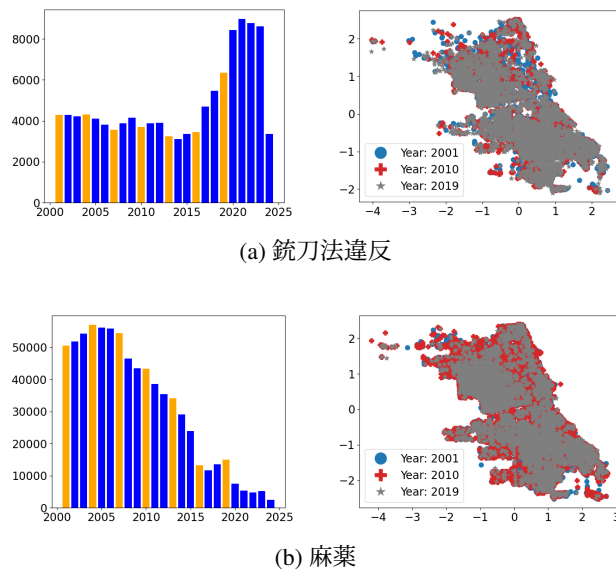


図 8: シカゴの犯罪データの可視化.

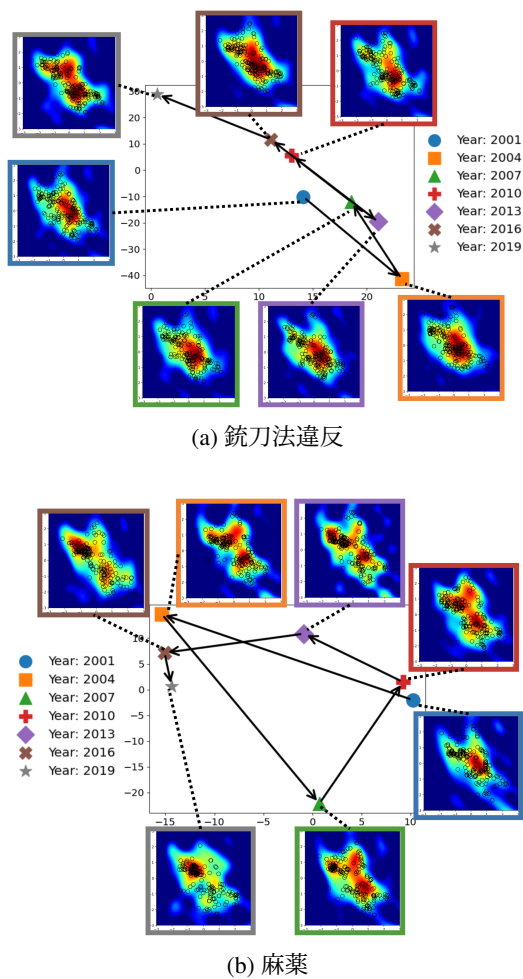


図 9: 犯罪データの分析結果.