教科書にない自然言語処理

持 橋　大 地

1. はじめに

2022 年末に ChatGPT[*1)] が登場して以来, **自然言語処理**にもとづく人工知能 (AI) が急速に脚光を浴びるようになりました. ChatGPT は背後に, General Purpose Transformer と名付けられた超大規模なテキストを学習した大規模言語モデル (以下 LLM) を持っており, 対話形式でさまざまな質問に答えてくれます. そのため, まるで LLM があれば自然言語処理が何でもできるかのように思っている向きもあるようです.[*2)] 実際, LLM の学習アルゴリズムと使い方を解説した教科書は世に溢れており, 出版は止まるところを知りません.

ただし, よく見てみるとそれらの本は, Transformer の仕組みや文埋め込みの学習法などの技術については書いてあっても, それで何ができるかについては, 単語のタグ付け, 文書分類, 感情分析 (といってもほとんどは正負に分けるだけ) といった, 十年一日の簡単な話ばかりなのが現状です.

実際に社会で必要となる自然言語処理はもっと多様で, 解き方を自分で発見する必要があります. このとき, LLM を単に使っても解けないか, 非常に非効率なことも多いでしょう. たとえば, ある新製品に関して消費者のアンケートを 10 万件集めたとします. この 10 万件のテキストを果たして LLM に読み込ませることができるかは別にして, もしできたとしても, 「内容をまとめて下さい」と LLM に指示して出力された結果が正しいか, 漏れがないかをどうやって検証したらいいのでしょうか. また, 製品に対して批判的な意見のうち, どれが根拠のある重要な批判で, どれが軽い文句なのかを区別するには, どうしたらよいのでしょうか. 仮に LLM にこれらを質問して答えてくれたとしても, それは「うまく答えてくれるといいなあ」という希望的観測の結果にすぎず, 何ら客観的裏付けを持ったものにはなりません.

したがって大事なのは, LLM に質問して「AI を使った」つもりになるのではなく, AI の背後にある統計的な原理を理解し, 適材適所で LLM を含んだ道具を使いこなすことだと考えられます. 実は多くの場合, 巨大な LLM は不要で単に SVM でよかった, といった話を耳にすることも増えてきました.

そこで本稿では, 筆者が他分野および企業との共同研究を通じて行った, 「教科書にない」二つの数理的な研究についてご紹介します. なお, これらは深層学習は使っていませんが, 筆者は深層学習を適切に部品として用いた研究ももちろん行っています. 詳しくは, 筆者の研究室サイト[*3)] をご覧ください.

*1) https://chat.openai.com/

*2) こうした状況は実は既視感があり, 2000 年代にサポートベクトルマシン (SVM) という, 機械学習の強力な分類器が現れたときも同じでした. 現在は SVM は, 特定の場合に大変有効なツールの一つとして知られています.

*3) http://clml.ism.ac.jp/

2. 「副詞」の意味の統計モデル

筆者は 10 年ほど前から, 要請を受けてロボティクスの分野と共同研究を行ってきました. ロボティクスは現在, 決められた動作を正確に行う従来のロボティクスから, より自律的に判断して自由な動作を行う知能ロボティクスに急速に進化しつつあります. 社会の中でロボットが様々な場面で人間をサポートするために, これは不可避な変化といえるでしょう.

このような中で, ロボットに与える指示も「この箱を向こうに運んで」と言うとガッチャン, ガッチャンと運んで終わりました, というこれまでの機械的な動作から, 「**丁寧に**運んであげて」「**しっかり**支えてあげて」といった, より人間に寄り添った表現を理解することが今後ますます求められると考えられます. ここで重要になるのは, 太字で書いたような**副詞**, つまり動作を修飾する言葉です. 「箱」「ボトル」のような実体がないこうした言葉を, 数学的にはどう理解したらいいのでしょうか?

一つ言えることは, 動作は最終的に体の各部の軌跡として表れますから, この軌跡の性質を副詞は表現している, ということです. たとえば「よぼよぼ」に動くとは細かいランダムウォークを含むということですし, 「すぱっと」動くとは区分的に直線状に動く, ということでしょう. ここで大事なのは, これはそれぞれの軌跡がどこを通るか, つまり軌跡の位相にはよらないということです. むしろ重要なのは軌跡の特徴であり, これは軌跡が (時間に対する) 関数のクラスとしてどのような集合に属するか, を記述していると考えられます.

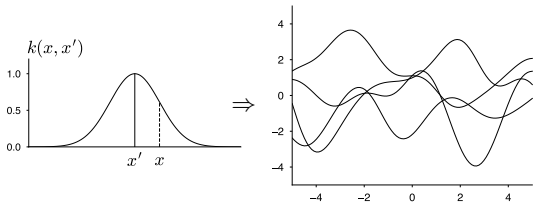


図 1: ガウス過程とカーネル. 最も基本的な RBF カーネル (左) と, そこからランダムに生成された関数 (右).

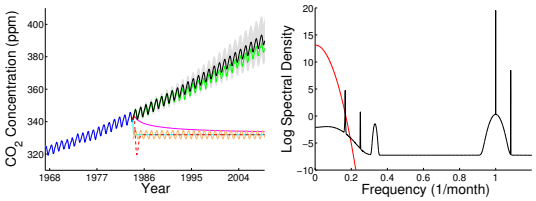


図 2: ガウス過程におけるスペクトル混合カーネル. 左のマウナロア山 CO_2 データの時系列の前半から, 周波数空間における右のようなカーネルが推定でき, これにより左図後半の太線のような高精度な外挿も可能になります. (論文[2] より引用)

2.1 軌跡の確率モデル

こうした軌跡をランダムに生成する基本的な確率過程として**ガウス過程**があり, 筆者も自然言語処理が専門ながら, 入門書[1] を出版しています. ガウス過程は図 1 左のような**カーネル**とよばれる, 入力空間の類似度を表す関数 $k(x,x')$ を与えれば, それから軌跡を表す関数, すなわち無限次元のベクトルを, 対応する無限次元の多変量ガウス分布からのサンプルとして出力します.

たとえば, 図 1 左の RBF カーネル

$$k(x,x') = \exp\left(-\frac{||x-x'||^2}{2\ell^2}\right) \qquad (1)$$

とよばれる標準的なカーネルを用いて, これを可能な D 次元の入力 (この場合は 1 次元の実数) x_i, x_j の間の共分散行列の要素 $K_{ij} = k(x_i, x_j)$ とした多変量ガウス分布からランダムにサンプリングすると, 図 1 右のような軌跡が得られます. カーネルは通常は, 式 (1) のような標準的なカーネルの中から選ぶか, その組み合わせを考えて, 組み合わせ係数をデータから学習します[1] が, 実はカーネル自体をデータから逆算することも可能です.

2013 年の論文で示されたスペクトル混合カーネル[2] というこの方法では, まずボホナーの定理から, 入力の差 $\tau = x - x'$ だけに依存する定常な基底関数 $k(x,x') = k(\tau)$ は逆フーリエ変換により,

$$k(\tau) = \int_{\mathbb{R}^D} \psi(s) e^{2\pi i s^T \tau} ds \qquad (2)$$

と表せることを利用します. ここで $\psi(s)$ は $k(\tau)$ の周波数空間における表現で, フーリエ変換の双

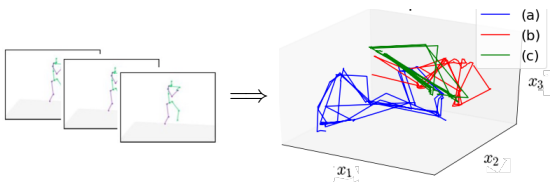


図 3: GPLVM による動作の非線形次元圧縮. 動画から抽出された (a)–(c) の異なる歩行動作が, 低次元の潜在空間の軌跡として表現されています.

対性から

$$\psi(s) = \int_{\mathbb{R}^D} k(\tau) e^{-2\pi i s^T \tau} d\tau \qquad (3)$$

で求めることができます.

式 (1) の RBF カーネルを式 (3) に代入して計算すると, RBF カーネルの周波数表現は

$$\psi(s) = (2\pi\ell^2)^{D/2} \exp(-2\pi^2\ell^2 s^2) \qquad (4)$$

と, 原点を中心とするガウス分布となることがわかります.

ということは, $\psi(s)$ について周波数空間での混合ガウス分布

$$\psi(s) = \sum_{i=1}^{Q} w_i \mathcal{N}(s|\mu_i, \sigma_i^2) \qquad (5)$$

を考えれば, 任意のカーネルを表現できるはずです. 式 (5) を式 (2) に代入すると, これはカーネルとして

$$k(\tau) = \sum_{i=1}^{Q} w_i \prod_{d=1}^{D} \exp\left(-2\pi^2\tau_d^2\sigma_{di}^2\right) \cos(2\pi\tau_d\mu_{di}) \qquad (6)$$

を用いることと等価で, このカーネルのハイパーパラメータ $\{w_i, \mu_i^{(d)}, \sigma_i^{(d)}\}_{i=1}^{Q}$ は, 通常のガウス過程のパラメータ最適化で求めることができます. これを**スペクトル混合カーネル**とよびます.

たとえば原論文の著者ら[2]は, 図 2 左の有名なハワイ・マウナロア山の CO_2 濃度の時系列データについてスペクトル混合カーネルを当てはめ, 図 2 右のように 14,6,4,2,1 ヶ月の各周期に対応するピークが周波数領域で得られ, 非常に正確な外挿が可能になることを報告しています.

2.2 副詞とカーネル周波数の対応モデル

ロボティクスの場合, 与えられるデータは各関節角の時系列になります. これは高次元 (40〜90 次元程度) なため, GPLVM (ガウス過程潜在変数モデル)[1]とよばれる非線形な主成分分析で 3 次元に圧縮したものを, 図 3 に示しました. それぞれの軌跡 (a)(b)(c) が, 異なる歩行動作に対応しています. ここから式 (6) を用いて, 各軌跡を出力したカーネルの周波数表現を求めたものが図 4 です. ゆっくり歩いている (b) に比べ, 動きの速い (c) は周波数 (の逆数) が高周波に集中しており, (a) はその中間となっていることがわかります.

図 4 を入力とすれば, 動画に付けられた「ゆっくり」「生き生きと」「波のような」といった副詞と, スペクトル混合カーネルを通じて得られた周波数の観測値を対応づける統計モデルを考えることができます. このためにわれわれが提案した[3]のが図 5 に表したモデルで, これはトピックモデルとして知られる LDA (潜在ディリクレ配分法)[4], およびその無限次元版である HDP(階層ディリクレ過程) の拡張となっているため, 以下 HDP–SMLDA と呼ぶことにします.

HDP–SMLDA では, 各動画 $n = 1, \cdots, N$ について, 式 (6) から得られる動作の周波数 $\{\mu_i^{(d)}\}_{i=1}^{Q}$ および, その動画に付与された M 個の副詞 $\mathbf{w}_n = \{w_1^{(n)}, \cdots, w_M^{(n)}\}$ が次のようにして生成されたと考えます.*4)

0. Draw $G_0 \sim \mathrm{DP}(\eta, H)$. (無限次元の事前分布)

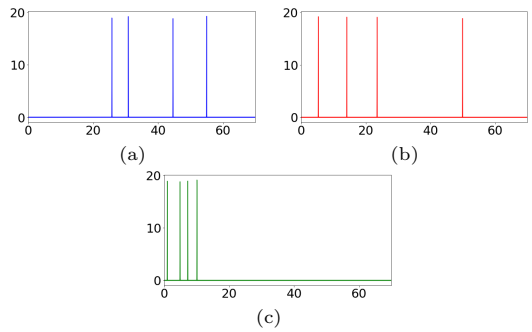


図 4: 図 3 の各動作について, 1 次元目の軌跡を解析したスペクトル混合カーネルによる周波数表現. 縦軸, 横軸はそれぞれ推定された 4 個のガウス分布の確率密度, 平均を表す.

*4) 図 4 の分散は非常に小さいため, ここでは $\mu_i^{(d)}$ だけを考えています.

表 1: AIST++ データに対する HDP-SMLDA で得られた潜在トピック別の副詞上位 5 語.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
激しく	楽しそうに	規則正しく	しなやかに	力強く	踊るように	慣れたように	テンポ良く
力強く	リズミカルに	テンポよく	優雅に	激しい	ステップを踏み	安定的に	スタイリッシュに
はっきりと	軽やかに	躍動的に	なめらかに	激しく	嬉しそうに	くねくねと	気持ち良さそうに
熱心に	弾むように	生き生きと	軽やかに	素早く	躍動するように	キビキビと	流れるように
上品に	元気に	大胆に	くるくると	大胆に	つまらなそうに	ダイナミックに	格好良く
Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16
ゆったりと	ダイナミックに	はずむように	格好よく	キビキビと	小刻みに	確かめるように	軽く
滑らかに	激しく	ひろがるように	カクカクと	機械のように	回るように	ひょうきんに	揺れているような
ゆっくりと	くねくねと	たどたどしく	おおらかに	コミカルに	細かく	丁寧に	波のような
機械的に	おおきく	ぐんぐんと	楽しそうな	しっかりと	クルクルと	慎重そうに	細かい動作で
ゆるやかに	キレよく	落ち着いた	機械のように	ロボットのように	リズム感よく	探すように	ロボットのような

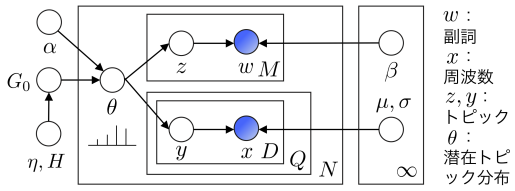


図 5: HDP-SMLDA のグラフィカルモデル. 動画ごとに存在する潜在トピック分布 θ から, 動画に付与された副詞 w と動作のカーネル周波数 x が生成され, 両者は潜在トピックを介して結びついています. これらの対応はすべて未知で, 観測されるのは色のついた副詞と周波数の集合だけです.

1. For $n = 1, \cdots, N$,
   a. Draw $\theta_n \sim \mathrm{DP}(\alpha, G_0)$. (動画ごとのトピック分布を生成)
   b. For $i = 1, \cdots, M$ (M 個の副詞について)
      * Draw $z \sim \theta_n$. (副詞のトピックを生成)
      * Draw $w_{ni} \sim \beta_z$. (トピックから副詞を生成)
   c. For $j = 1, \cdots, Q$ (Q 個の周波数について)
      * Draw $y \sim \theta_n$. (周波数のトピックを生成)
      * Draw $x_{nj} \sim \mathcal{N}(\mu_y^{(j)}, \sigma_y^{(j)2})$. (トピック別の正規分布から周波数を生成)

ここでθは, 和が1になる$\theta = [0.2, 0.6, 0, 0.1, \cdots]$のような無限次元の潜在トピック分布で, 各動画がトピックとよばれる混合要素にどれくらいの確率で属しているのかを表しています. この θ からランダムにトピック z を選び, z ごとの副詞の確率分布 β_z から,「生き生きと」のような副詞 w が生成されたとします. これによって, 副詞が意味的にクラスタ分けされることになります.

HDP–SMLDA はこの際に, 観測された動作の周波数についても, θ からトピック y を選び, 周波数 $x = \mu_i^{(d)}$ が y 番目のガウス分布 $\mathcal{N}(\mu_y, \sigma_y^2)$ から生成されたとする混合ガウス分布を考えます. このトピックは副詞のものと共有されているため,「この副詞ならば (x の次元ごとに) この周波数」というように, 副詞のクラスタと周波数の各ガウス分布が対応するのが特徴です.

以上のモデルにおいて, トピック z ごとの副詞の分布 β_z, ガウス分布の平均と分散 μ_y, σ_y^2, 各動画のトピック分布 θ はすべて未知のため, マルコフ連鎖モンテカルロ (MCMC) 法を用いて, すべてをサンプリングして推定します.「トピック」の数およびその際のハイパーパラメータ α についても, 階層ディリクレ過程 (HDP)[5]の理論を用いて自動的に決定します.

表 1 に, 産業技術総合研究所が公開しているダンス動画のデータセット AIST++[6]に, クラウドソーシングによって副詞をアノテーションしたデータについて HDP–SMLDA を適用し, 得られたトピック別の副詞の分布の上位語を示しました. この場合, 推定されたトピック数は約 16 になりました. 明らかに, 意味的に関連の深い副詞がクラスタにまとまっており, 本手法が動画から副詞の意味を学習していることがわかります. このモデルでは副詞の各トピックの裏に, Q 次元 (この場合は $Q = 3$) の周波数空間のガウス分布が対応していることが特徴で, それを図 6 に示しました. それぞれの副詞のトピックが, Q 次元の異なる周波数の組み合わせ $\{\mu_i\}_{i=1}^{Q}$ を持っていることがわかります.

表 2: 正解の副詞と, 推定された副詞の上位 7 語.

正解の副詞	$Q=4$ の場合	$Q=10$ の場合
情熱的に	力強く	テンポ良く
陽気に	激しい	スムーズに
テンポ良く	激しく	スタイリッシュに
スムーズに	大胆に	流れるように
流れるように	堂々と	陽気に
力強く	キビキビと	悲しそうに
大胆に	ダイナミックに	気持ちよさそうに

これにより, 逆に動作のカーネル周波数がわかれば, そこから θ を推定することで, その動作を表す副詞を確率つきで求めることができ, その逆も可能になります. 表 2 に, 図 7 のジャズバレエの動画から計算した, 副詞の上位語を示しました. 観測周波数の数 Q が増えるほど副詞の意味が詳細になりますが, クラウドソーシングで付与された「正解」の副詞と比べると, むしろ Q を増やしすぎない方が人間の判断に近いことがわかります.

なお, 単純に動画の時系列から副詞を予測するニューラルネット (LSTM) および多層パーセプトロン (MLP) を使った場合と, 副詞の予測性能を比べたものが表 3 で, 提案手法のような動作の数理を考慮しないニューラルネットは, 予測性能の面でも提案法より劣っていることがわかります.

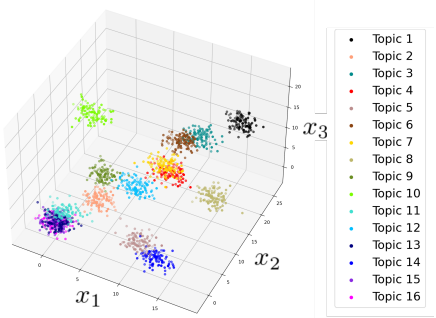


図 6: トピックと動作特徴の関係. 表 1 の各潜在トピックが, 動作の周波数空間における (ここでは 3 次元の) それぞれの多変量ガウス分布に対応しています.

表 3: 各モデルにおける評価時のパープレキシティ.

モデル	LSTM	MLP ($Q=4/10$)	HDP-SMLDA ($Q=4/10$)
100 Walks	210	253 / 284	**89** / 117
AIST++	1068	994 / 1027	**320** / 382

図 7: 評価用の動画. この動画では, ダンサーはジャズバレエを踊っています.

3. 広告メールと言葉の関係

もう一つ, ブラックボックスの深層学習が不得意なものの一つに, 純粋に統計的な分析があります. 筆者は (株)NTT ドコモ, (株) 電通, (株)NTT アドの共同出資会社である (株)D2C の自然言語処理に関する研究アドバイザーを長く続けています. 我々の共同研究[7]では, ドコモ社の巨大なデータから, 興味深い傾向を発見しました.

(株)NTT ドコモのスマートフォン向けのメール配信サービス「メッセージ S(スペシャル)」[*5]は, 登録者 3300 万人を持つ巨大な広告配信サービスです. もちろん, 広告であるからには広告のクリック率 (CTR) や, 購入につながったコンバージョン率 (CVR) が問題になりますが, これらの確率は通常きわめて低く, 広告の効果が直後とは限らないことを考えると, そもそもメール広告を開封してもらえるかが入り口として重要だと考えられます.

メッセージ S では, 配信当日にメールを開いたかが記録されるため,「どういうメールならば開いてもらえるのか」に興味が生まれます. ユーザーからみると, 開封前のメールの情報はタイトル文程度です. そこで,「どんな言葉が使われていれば, 開封率が上がる/下がるのか」が具体的な焦点になるでしょう. 後で示すように, これはユーザーの特徴によってかなり異なっており,「どういったユーザーが, どの言葉に ± の反応を示すのか」という, 興味深い統計的な知見が得られます.

3.1 分析データ

メッセージ S の配信ログのうち, 2021/9/1〜2022/1/31 の 5 ヶ月分を学習データ, 2022/2/1〜

*5) https://www.ntt.com/business/services/message_s.html

2022/5/31 の 4 ヶ月分をテストデータとして用いました．ログは 1 日あたり約 1000 万件と膨大で，学習データ全体では 15 億件にものぼるため，ランダムサンプリングによってそれぞれ 500 万件を抽出して分析しています．UI の違いから Android ユーザーのみを対象とし，開封率が 80% 以上および 20% 以下の極端なユーザーは除外しています．なお，データは完全に匿名化されており，統計的な分析のみを行っています．

3.2 データの前処理

単純な形態素解析のみを行ってしまうと，「話題沸騰」のような言葉が「話題/沸騰」と切れてしまい，「話題」「沸騰」のそれぞれしか分析できなくなってしまいます．そこで，後で説明する正規化自己相互情報量 (NPMI) を用いて，教師なしで統計的なフレーズ化を行います．形態素解析で得られた単語の連続 vw について，$\mathrm{NPMI}(v,w) \geq 0.5$ であれば vw を一つの単語としてつなげる，という処理を行います．これを n 回行えば，2^n 語までの単語列がフレーズ化できますが，予備実験の結果からここでは $n=1$ としました．本研究では広告の種類数がさほど多くないため，使用したフレーズの総数は $N=762$ となりました．

3.3 統計分析

上で統計的にフレーズ化された，メールのタイトルに含まれる単語を w としましょう．配信ログには，ユーザーの情報として性別および年代 (10 代, 20 代, …) が含まれているため，(30 代, 女) のような (年代, 性別) のペアを特徴量 f とします．

単語 w を持つメールが特徴量 f を持つユーザーによって開封されたとき，頻度 $Y(f,w)$ に 1 を加えることにすると，データ全体は下の式 (7) のような行列 $\mathbf{Y}$ で表すことができます．これから，どのような情報が読みとれるでしょうか？

$$\mathbf{Y} = \begin{array}{c} \\ f_1 \\ f_2 \\ \vdots \\ f_M \end{array} \begin{array}{c} \begin{array}{cccc} w_1 & w_2 & \cdots & w_N \end{array} \\ \begin{pmatrix} & \vdots & & \\ \cdots & Y(f,w) & & \\ & & & \\ & & & \end{pmatrix} \end{array} \tag{7}$$

ユーザーの特徴 f と単語 w の関係を分析する最も素直な方法は，この二者の**自己相互情報量** (Pointwise Mutual Information, PMI)

$$\mathrm{PMI}(f,w) = \log \frac{p(f,w)}{p(f)p(w)} \tag{8}$$

を計算することでしょう．これは，f と w が独立な場合 (分母) に比べて実際の観測値 (分子) がどれくらい生じやすいのかを測っており，独立な場合は $p(v,w)=p(v)p(w)$ ですから，値は $\log 1 = 0$ になります．式 (8) をさらに $p(f,w)$ で期待値をとったものは情報理論で相互情報量とよばれているため，その各要素である式 (8) は自己相互情報量とよばれています．[*6)]

ただし，式 (8) は f や w が稀で確率が小さいほど，値が大きくなってしまうという問題があります．単純に式 (8) の大きい方から (f,w) のペアを探すと，ほとんど出ない言葉や，データの稀な 80 代の特徴などがクローズアップされることになってしまいます．そこで，PMI をその最大値である $-\log p(f,w)$ で割った

$$\mathrm{NPMI}(f,w) = \log \frac{p(f,w)}{p(f)p(w)} \Big/ \left(-\log p(f,w)\right) \tag{9}$$

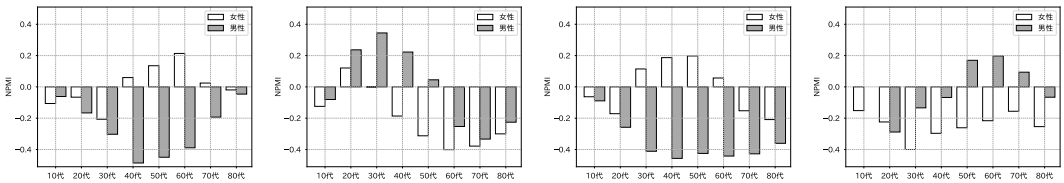
を**正規化自己相互情報量** (NPMI)[8] といいます．この指標は稀な f や w の影響を抑え，PMI と異なり

$$-1 \leq \mathrm{NPMI}(f,w) \leq 1 \tag{10}$$

をみたします．f と w の出現が完全に同期するとき 1，逆になるとき -1 になります．詳しい議論は，原論文[8] を参照してください．

そこで，式 (7) の $\mathbf{Y}$ について式 (9) の NPMI を単語 w ごとに計算すると，(10 代, 男), (10 代, 女), …の各特徴 f について，w がどのように相関しているかを求めることができます．これを可視化したのが図 8 で，単語によって，開封するユーザーの特徴がかなり異なっていることがわかります．「研

*6) この比は，f と w の実際の分布が，独立な場合とどのくらい違うかの**程度**を測っているとみることができます．データ量が充分多い場合は，完全に独立な場合を帰無仮説とする検定はほとんど棄却されるため，無意味になります．



「研究成分」　「転職」　「シワ」　「売却」

図 8: 単語を含むメールタイトルがどの性別・年代に強く影響を与えているかを, NPMI の棒グラフで表したもの. 横軸は年代, 縦軸は NPMI で, 黒白が男性・女性を表します. 正の値が大きいほど開封に肯定的な影響を, 負の値が大きいほど否定的な影響を与えています. 年代・性別で大きな差があることがわかります.

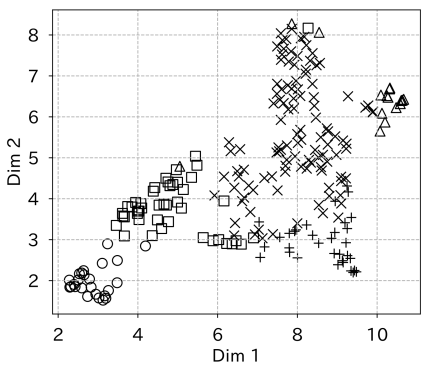


図 9: UMAP によって, 各単語の NPMI 反応ベクトルを可視化したもの. 各点は単語を表しています. マーカーで元の空間でのクラスタを表しています.

究成分」は 40 代以降の女性にしか効かないこと, 「転職」はほぼ男性 40 代までに効くこと, 「シワ」が効くのは女性で 50 代程度までであること, などを読み取ることができます.

なお, 図 8 は各単語 w を, 特徴 $f_1, f_2, \cdots, f_M$ との NPMI を並べたベクトル

$$v(w) = (\mathrm{NPMI}(w, f_1), \cdots, \mathrm{NPMI}(w, f_M))$$

と結びつけられるということですから, このとき $v(w)$ には一定のパターンがあるはずです. そこで, $v(w)$ の集合を 2 次元に可視化したのが図 9 で, 左下の○は女性 > 男性に効く言葉, 左中の□は男性 > 女性の言葉, 右下の + は若い世代 > 高齢者に効く言葉, 右上の×は影響が弱い言葉となっています.

4. おわりに

本稿では, 実際の問題で必要となった二つの「教科書にない」自然言語処理の数理モデルについて解説しました. これらの分析を LLM が自動的に行ってくれるとは考えにくく, 仮に今後できても, それが正しいかについて, 人間の側にやはり数理的な理解が必要になるでしょう. LLM のような巨大なモデルも, 「全体として何を最適化しているのか」という数学的な枠組の中に取り込まれる日が来ると考えられます. 本稿以外にも可能なモデルは無数にあり, それらを探究していくことが, すなわち言語を知ることにもつながってくるでしょう.

参考文献

1) 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.
2) Andrew Gordon Wilson and Ryan Prescott Adams. Gaussian Process Kernels for Pattern Discovery and Extrapolation. ICML 2013, pages 1067–1075, 2013.
3) 谷口巴, 持橋大地, 長野匡隼, 中村友昭, 長井隆行, 稲邑哲也, 小林一郎. ガウス過程を用いた周波数スペクトル分析による副詞の理解. 電子情報通信学会 PRMU 研究会 PRMU2021-74, pages 91–96, 2022.
4) David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
5) Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. Hierarchical Dirichlet Processes. *JASA*, 101(476):1566–1581, 2006.
6) Ruilong Li, Shan Yang, David A. Ross, and Angjoo Kanazawa. AI Choreographer: Music Conditioned 3D Dance Generation with AIST++. ICCV 2021, 2021.
7) 吉井健敏, 城田晃希, 市川匠, 佐野雄一 (株式会社 D2C), 持橋大地. メール型広告におけるタイトルが開封に与える影響. 情報処理学会研究報告 情報基礎とアクセス技術研究会 (IFAT) 2022-IFAT-148, pages 1–11, 2022.
8) Gerlof Bouma. Normalized (Pointwise) Mutual Information in Collocation Extraction. Proceedings of GSCL, pages 31–40, 2009.

（もちはし だいち，統計数理研究所）