公開

研 究 終 了 報 告 書

「言葉が埋め込まれた空間の形と言葉の意味の接続」

研究期間： 2020 年 11 月～2023 年 3 月
研 究 者： 横井 祥

1． 研究のねらい

単語や自然言語文に計算機可読な意味表現を与え、これを通じて言語に関わる様々な意味の計算をおこなうことが自然言語処理・計算言語学分野の大目標である。この研究では、近年飛躍的に発展した言語の埋込表現の技術、とりわけシンボルである単語を連続空間で柔らかく表現する技術に着目し、**言語の意味が埋込空間の形にいかに反映されるのか**について、理論・実験の両面から研究をおこなう。

単語の表現に関しては 2010 年代初頭に単語埋め込みと呼ばれる基盤技術が確立した。このトークンを連続空間で表現するという考え方は今ではトランスフォーマーと呼ばれるより複雑なアーキテクチャに受け継がれ、ほとんどすべての自然言語処理タスクの必須の道具となっている。一方で、こうしたアプローチでなぜ言語という複雑な対象をかなり高い精度で処理できるのか、あるいはその内部で言語の意味がいかなる空間の性質に置き換えられるのかについては、今のところ残念ながらほとんど何もわかっていない。結果として、稀に発生する不可解な挙動がなぜ起きるのか、あるいは現状のモデルの表現力は十分足りているのかといった実用的にも重要な問題に対して、経験的ないし発見的なアプローチに頼らざるを得ない状況である。たとえば単語ベクトルを足し合わせれば良い文の表現を作れることが知られており、翻って単語ベクトルの長さにある種の重要度がエンコードされていることが分かっているが、それはなぜなのか、重要度というのは言語いかなる特性を反映した概念なのかなど、重要な疑問はすべて未解決の状態で放置されている。したがってひとつめの研究のねらいは、**単語の意味が単語埋込空間の幾何的特性としていかに反映されるのか**を確かめることである。

さらに、**単語よりも大きな単位のテキスト**、たとえば文の意味表現については、今のところ表現の決定打が存在していない。単語と文の関係に関しては、構成性と呼ばれる概念が歴史の評価に耐えた重要な立脚点と言えよう。古来より指摘されているように、われわれ人間は（驚くべきことに）有限な語彙を組み合わせて全く新しい考えを表現することができる。言語において単語は原始的な部品だということだ。したがって文を計算機で表現する際も、文自体を直接的に単位とするのではなく、単語の表現に帰着・還元することが肝要だと考えられる。実際、単語ベクトル集合の平均（加法構成性）、あるいは単語ベクトル集合同士の最適輸送コストなど、単語がベクトルで表現されていることを活用した手法群が経験的に大きな成功を収めている。しかしこちらについても、文の意味の計算を単語埋込のいかなる性質に帰着させるべきかについての理論的な理解は全くと言って良いほど進んでいない。足し算という演算で表現できるのはいかなる種類の意味合成なのか、それは拡張できるのか、単語ベクトル集合を直接表現する際に単に連続空間の点群として扱えば良いのか、それともより好ましい表現があるのか、いずれも未解決の重要な問題である。したがってふたつめの研究の狙いは、**文の意味を単語が埋め込まれた空間で計算する**適切な方法を探究することである。

2.　研究成果

（1）概要

単語が埋め込まれた空間の幾何的特性を解明し、またこれを利用してより大きな単位のテキストの意味を計算する技法を探究するという大きく二種類の目標について、それぞれで複数の研究成果を得ることができた。ここで取り上げる研究以外にも様々な関連研究を実施することができ、またその一部は顕著な結果として「5. 主な研究成果リスト」に取り上げたが、このセクションではふたつのねらいに直接的に関係するものに絞り、それぞれで 3 点ずつのサブプロジェクトを取り上げることにする。

【単語が埋め込まれた空間の形を知る】

単語の言語的な特徴と単語が埋め込まれた空間の特徴の関係性については、まず、単語の「意味の強さ」と単語ベクトルの**ノルム**に関係について理論的な説明を与えた（下記研究テーマ A1）。

＜一部非公開＞

単語列としての自然言語文をニューラルネットワークで処理する方法として、最近ではトランスフォーマーと呼ばれる機構がほとんどすべての自然言語処理アプリケーションの実現のために第一に選択される強力な手段となっている。いわゆる「単語ベクトル」とは異なるこの機構内での単語の連続表現の扱いについても、予測への寄与をトークンの埋込の**ノルム**に帰着させる統一的な方法を開発し、これによって BERT 等の代表的なモデルの内部機構についての深い理解を得ることにつながった（下記研究テーマ A3）。

【単語が埋め込まれた空間で文の意味の違いを計算する】

単語が埋め込まれた空間での文の意味の計算については、まず、**数単語の意味の合成**を単語埋込上の演算として実現するための手法を作った（下記研究テーマ B1）。単語の意味の合成としては、AND の計算が埋込の和で近似的に成り立つことが理論的にも経験的にも知られている。たとえば、v(woman) + v(royal) ≈ v(queen) となる。本研究ではこれを OR や NOT に一般化する手法を与えた。たとえば case が多義語で box（箱）と instance（事例）の両方の意味を持つことや（OR）、hate と love が対義語であること（NOT）を、埋込空間で適切に計算する方法を提案した。

さらに**文レベルの計算あるいは単語集合に対する意味の計算**を実現するため、単語集合を単語埋込の張る線型部分空間で表現するフレームワークを提案した（下記研究テーマ B2）。量子論理の考え方を援用した提案法は、単語集合に対する集合演算と線型部分空間がなすオーソモジュラー束上の演算とを対応させる。結果として、シンボル集合に対する演算を埋込空間の柔らかさを活用しながら良いところどりで実現することができる。ここでも下記研究テーマ A1 で示した「単語の持つ情報の量 ≈ 単語ベクトルのノルム」の関係を活用した。

また単語ベクトル集合を埋込空間上の離散分布だと見なすことで、文の意味の計算に最適輸送の枠組みを自然に適用できる。さらに広く、言語の再帰的構造と近年の埋込表現の組み合わせは計算最適輸送と非常に相性がよい。この自然言語処理と最適輸送の豊かな

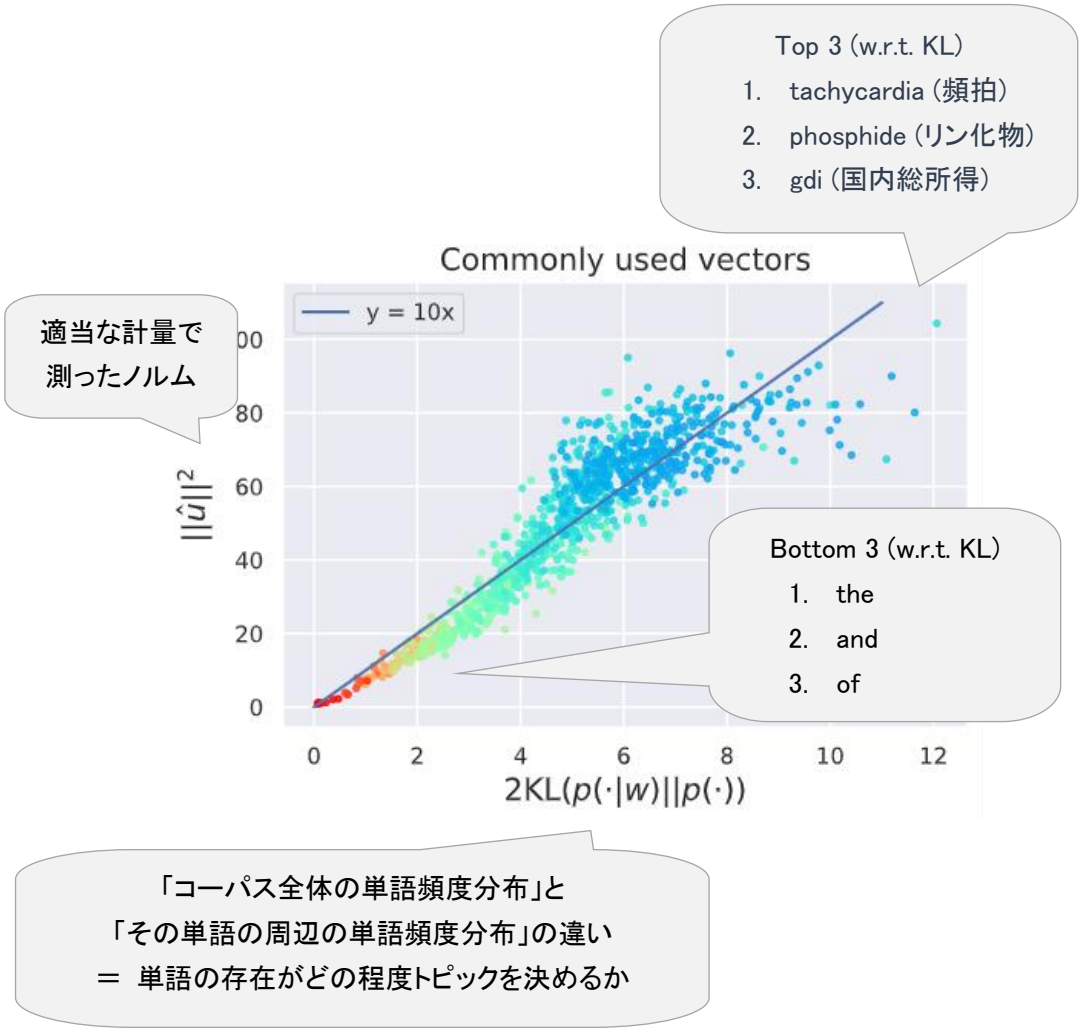
関係について、自分自身の研究を含めた大がかりな文献調査をおこない、複数の大規模なチュートリアル講演をおこなった（下記 B3）。

（2）詳細

■ 研究テーマ A： 単語が埋め込まれた空間の形を知る

□ 研究テーマ A1： 単語ベクトルの長さに意味の強さがエンコードされる機序の解明

単語ベクトルの長さと単語の意味の強さが実際に対応づくことを理論的・経験的に示した。はじめに対照学習を用いた典型的な学習アルゴリズムから得られる単語ベクトル空間を情報幾何の言葉で特徴づけ、単語ベクトルのノルムと「その単語が存在することで周辺単語の分布



がどの程度歪むか」が対応することを示した。また、単語頻度の交絡を除去してもこの性質が担保されていることを経験的に示した。

図： 単語の持つ情報の量と単語ベクトルの（適当な計量の下での）ノルムの理論的関係が経験的にも非常によく成り立っている様子。

文献・発表

- *大山 百々勢, *横井 祥, *下平 英寿. 単語ベクトルの長さは意味の強さを表す. 言語処理学会第 28 回年次大会 (NLP), Online, March 2022. (* equally contributed)
  [受賞] 若手奨励賞 (12/280).
- 大山 百々勢, 横井 祥, 下平 英寿. 単語ベクトルの長さは KL 情報量で解釈可能な意味の強さを表す. 第 25 回情報論的学習理論ワークショップ (IBIS), 筑波, November 2022.
  [受賞] 優秀プレゼンテーション賞ファイナリスト.
- Momose Oyama, Sho Yokoi, Hidetoshi Shimodaira. Norm of word embedding encodes information gain. arXiv preprint. (under review)

□ 研究テーマ A2: 単語ベクトル空間の等方性の測り方と単語頻度分布

学習済単語埋め込みに簡単な後処理を加えることで後段タスクの性能が向上することが知られており、とりわけ埋め込み空間の等方性が重要な指針となっている。

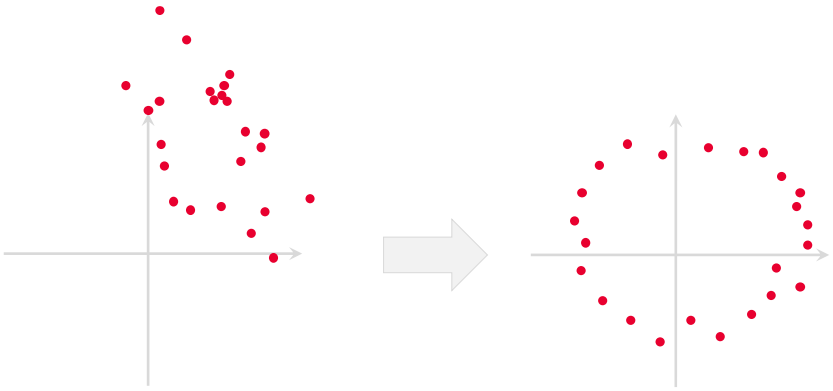


図: 単語ベクトル空間の等方化

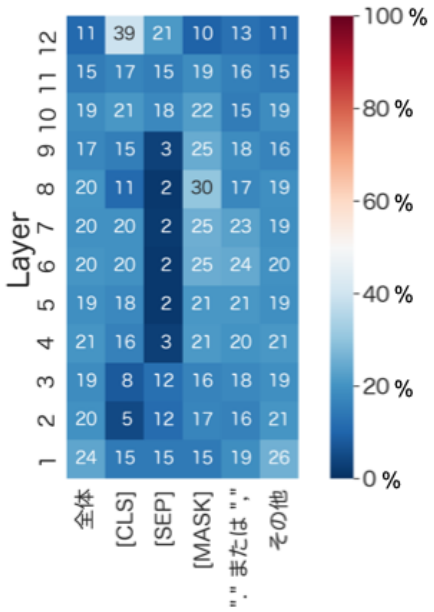
この研究では、既存の多くの後処理が単語埋め込み行列をデータ行列のように扱っていること、つまり経験的に冪分布に従う単語頻度分布が一様だと暗黙的に仮定していることを指摘し、単語頻度を考慮した等方化のアルゴリズムを提案した。この**単語レベル**の幾何的性質の改善をうながす提案法は、**文レベル**の計算をおこなう後段タスクに著しい性能改善をもたらした。また、単語頻度を考慮して計算した等方性の尺度が後段タスクの性能と高い正の相関を持つことを示した。

文献・発表

- 横井 祥, 下平 英寿. 単語埋め込みの確率的等方化. 言語処理学会第 27 回年次大会 (NLP), Online, March 2021.
  [受賞] 優秀賞 (8/361).

□ 研究テーマ A3: トランスフォーマー内部での「混ざり具合」のノルムを介した分析

トランスフォーマーという名前の注意機構つきニューラルネットを用いて各単語の周辺文脈を動的に考慮しながら潜在表現を作る方法が、自然言語処理全域で大きな経験的効果を発揮している。この研究では、当該のニューラルネットが単語の埋め込み表現を更新する際に文脈情報をいかに混ぜ合わせているかについて、内部表現のノルムを用いて定量的・定性的に評価する方法を提供した。とりわけ、注意機構で混ぜ合わせられた文脈情報が、残差結合および層正規化でそのほとんどがキャンセルされること、また BERT と呼ばれる穴埋め言語モデルは高頻度語を無視する（周辺単語の情報によって上書きされやすい）ことを発見した。



図：出力ベクトルに対する周辺文脈の寄与度。注意機構によってなされる「混ぜる作用」は残差結合とレイヤー正規化によりほとんどキャンセルされることがわかる。

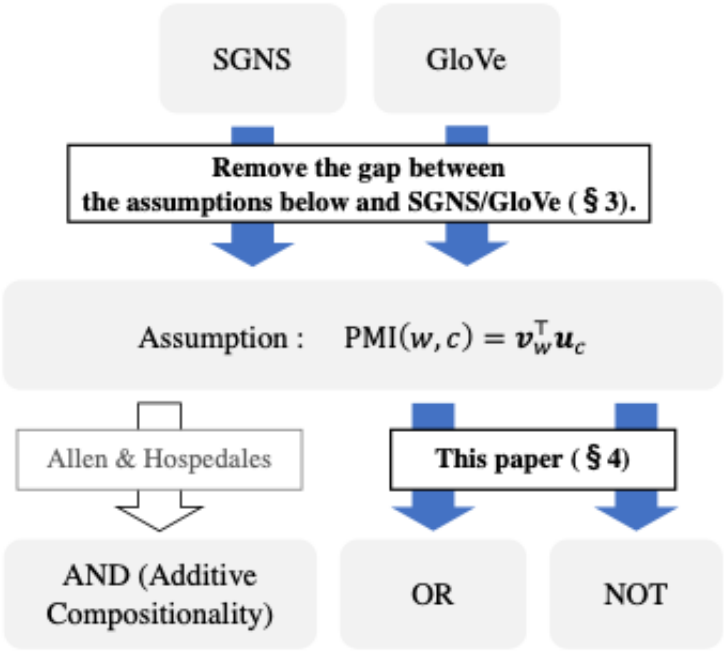
文献・発表

- 小林 悟郎, 栗林 樹生, 横井 祥, 乾 健太郎. Transformer の文脈を混ぜる作用と混ぜない作用. 言語処理学会第 27 回年次大会（NLP), Online, March 2021.
  [受賞] 若手奨励賞（11/247).
- Goro Kobayashi, Tatsuki Kuribayashi, Sho Yokoi, Kentaro Inui. Incorporating Residual and Normalization Layers into Analysis of Masked Language Models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.4547–4568, 2021.
- 小林 悟郎, 栗林 樹生, 横井 祥, 乾 健太郎. Transformer におけるフィードフォワードネットの作用. 言語処理学会第 28 回年次大会（NLP), Online, March 2022.
  [受賞] 委員特別賞.

■ 研究テーマ B: 単語が埋め込まれた空間で文の意味の違いを計算する

□ 研究テーマ B1： AND に対応する単語埋込の加法構成性の OR, NOT への拡張

単語ベクトルの和で句のベクトルを作るヒューリスティック（加法構成）に対する既存の理論は、単語ベクトルの内積と自己相互情報量の関係に仮定を置いている。我々は、実際に学習された単語埋め込みがこの仮定を見たさないこと、また単語頻度を考慮した中心化を加えるだけでこの仮定をよく満たすようになることを示した。すなわち、 (1)埋め込み空間全体に一定のバイアスが乗っていること、(2)加法構成においてこのバイアスが蓄積されてしまうこと、(3)重み付き中心化でバイアスを除去できること、の 3 点を理論的に理解することができた。これは研究テーマ A2 で示した頻度重み付き処理の別の観点からの利点とも言える。さらに既存の理論が示す AND の構成法だけではなく、OR や NOT の構成法を提案し経験的にも動くことを示した。



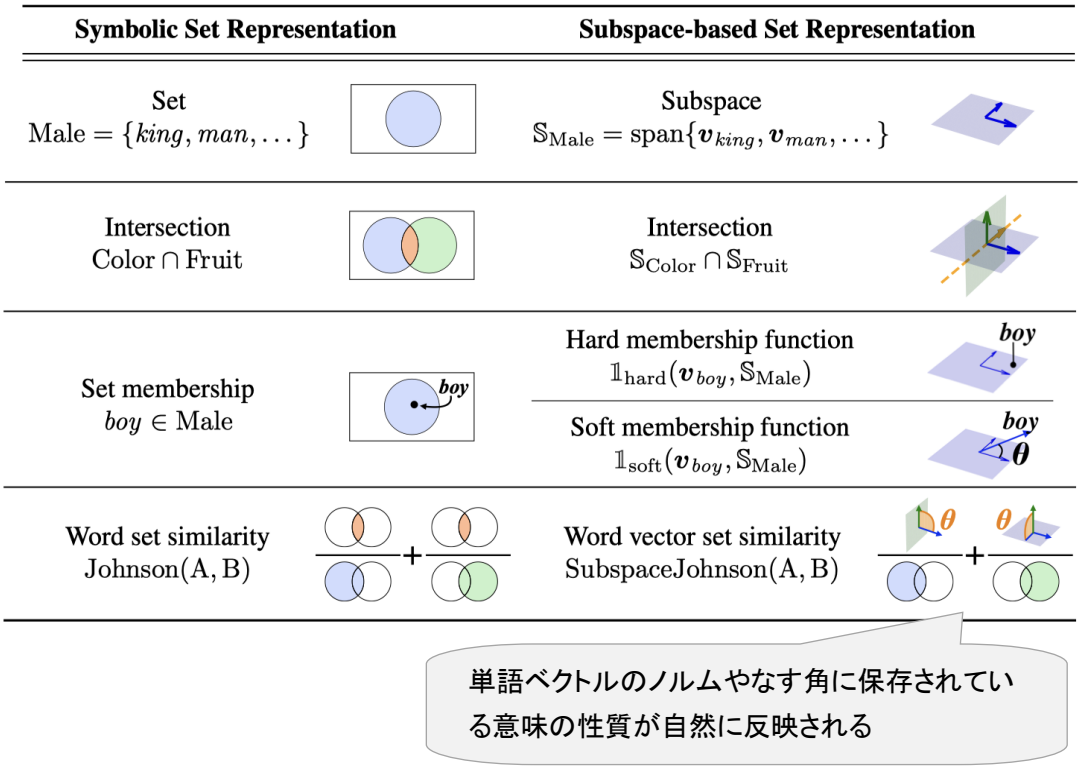
図： 単語頻度で重みづけた中心化をおこなうことで実際の単語ベクトルに加法構成がよく動くことを理論的に保証できる。

文献・発表

- 内藤 雅博, 横井 祥, 下平 英寿. 単語埋め込みによる論理演算. 言語処理学会第 27 回年次大会 (NLP), Online, March 2021.
  [受賞] 委員特別賞.
- Masahiro Naito, Sho Yokoi, Geewook Kim, Hidetoshi Shimodaira. Revisiting Additive Compositionality: AND, OR and NOT Operations with Word Embeddings. The ACL-IJCNLP 2021 Student Research Workshop, 11 pages, 2021.

□ 研究テーマ B2：単語埋込の張る線形部分空間による文表現

単語より大きな単位のテキストに対する意味計算の重要なサブセットとして、単語集合に対する集合演算がある。この研究では単語埋め込み空間上で、もとの単語（シンボル）の集合に対する集合演算を実現するためのフレームワークを提案した。提案法では、単語集合と単語埋込空間の線形部分空間を対応させ、量子論理の枠組みに基づき集合間演算を実現した。実際この枠組みによって、近い概念を示す単語集合を拡張するタスクや、文間の意味類似度を推定するタスクで高い性能が達成された。



図：単語集合の上の演算と単語ベクトルが張る線形部分空間上の演算の対応。

文献・発表

- 石橋 陽一，横井 祥，須藤 克仁，中村 哲．線型部分空間に基づく学習済み単語埋込空間上の集合演算．言語処理学会第 28 回年次大会（NLP），Online，March 2022．［受賞］若手奨励賞（12/280）．
- Yoichi Ishibashi, Sho Yokoi, Katsuhito Sudoh, Satoshi Nakamura. Subspace-based Set Operations on a Pre-trained Word Embedding Space. arXiv preprint. (under review)

□ B3：自然言語処理と最適輸送の関係性についてのアウトリーチ活動

「1. 研究の狙い」にも記したように、文についての意味の計算、とくに 2 文以上が関係する計算においては最適輸送が有用な道具となっている。これは、意味の計算においてその部品である単語同士のアラインメントが重要な役割を果たし、さらにそれらの単語については良い連続表現が開発されているからというふたつの理由による。
この自然言語処理と計算最適輸送の相性の良さおよび今後の発展について、著者自身の研究（関連文献参照）を越えて包括的な理解を得るため大規模なサーベイと、いくつかのチュートリアル講演をおこなった。

発表

- 横井 祥.「最適輸送の使い方」. 第 3 回 0x セミナー 最適輸送の情報科学における進展 招待講演. 2021 年 6 月.
- 横井 祥.「最適輸送と自然言語処理」. 言語処理学会第 28 回年次大会 チュートリアル講演. 2022 年 3 月.
- 横井 祥.「構造を持った言語データと最適輸送」. NAIST DSC NLP Seminar 2022 Summer チュートリアル講演. 2022 年 9 月.
- 横井 祥.「ChatGPT と自然言語処理 / 言語の意味の計算と最適輸送」. Workshop OT 2023. 2023 年 3 月.

関連文献

- Sho Yokoi, Ryo Takahashi, Reina Akama, Jun Suzuki, Kentaro Inui. Word Rotator's Distance. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.2944–2960, 2020.
- Ayato Toyokuni, Sho Yokoi, Hisashi Kashima, Makoto Yamada. Computationally Efficient Wasserstein Loss for Structured Labels. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, pp.1–7, 2021.
- Hiroaki Yamagiwa, Sho Yokoi, Hidetoshi Shimodaira. Improving word mover's distance by leveraging self-attention matrix. arXiv preprint (under review).

3. 今後の展開

＜一部非公開＞

4. 自己評価

研究期間の自己評価として満足できる点のひとつは、**内外の研究者との協働**によってひとりではとても成し遂げられなかった魅力的な研究やアウトリーチ活動をいくつも実施できたことだ。本報告書でもたくさんの共同研究を取り上げたが、本報告書に挙げた以外にも、ACT-X をきっかけにはじまった（まだ論文にできる結果は出ていないが間もなく出るであろうと期待できる）研究プロジェクトがすでに 10 以上も走っている。来年、再来年と少しずつ成果を世に問うていきたい。人脈という言葉は陳腐だが、数理・情報という大きな冠に偽りのない

多様で魅力的な研究者のかたがたと関われたことは、今後の研究者人生の糧となると確信できる。協業の効果にひとつだけ触れるなら、専門的な話題にも関わらず 500 人を越える聴衆を集め Slack（メッセンジャー）で数百のリアルタイムコメント・質問をいただくという思いがけぬ大盛況を博したチュートリアル講演（上記報告書 B3）を挙げたい。これも ACT-X で出会った数学者のかたとの企画トークに端を発するもので、さらにこの講演をご覧になったかたがたから多岐に渡る魅力的なお仕事を打診していただくことができ、さらにこの新しいお仕事でも多くの ACT-X 関係者と関わることとなった。

研究期間の自己評価として大きな反省のひとつは、**研究成果の文書化（論文化）**の遅さだろう。まだ査読中の仕事（研究テーマ B2, A1 など）はもちろん、国内会議の予稿の論文化が完了していない仕事（研究テーマ A2 など）まであり、プロの研究者として恥ずかしい思いである。上に書いた自己評価の満足できる面、つまり協業は、決して闇雲に抱える仕事を増やすための仕組みではなく、各研究者が抱える重要な複数の研究課題を協業を通して高度にかつ効率的に実現するための仕組みであろう。先述したチュートリアル講演に関しても、サーベイ論文化、国際会議でのチュートリアルの実施など、目指すべき先の目標はたくさん存在する。手元の案件の管理やスケジューリングなどの基本的なセルフマネジメントに関して、今回の反省を来年以後に切に活かしていきたい。

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：6件（別途，査読中プレプリント3件、国内会議での発表および論文13件）

1. Goro Kobayashi, Tatsuki Kuribayashi, Sho Yokoi, Kentaro Inui. Incorporating Residual and Normalization Layers into Analysis of Masked Language Models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021. pp.4547-4568., 2021.
トランスフォーマーと呼ばれる深層学習アーキテクチャが単語の埋め込み表現を更新する際に文脈情報がいかに混ぜ合わせられているかを定量的・定性的に評価する方法を提供した。注意機構で混ぜ合わせられた文脈情報が、残差結合および層正規化でそのほとんどがキャンセルされること、また BERT と呼ばれる穴埋め言語モデルは高頻度語を無視する（周辺単語の情報によって上書きされやすい）ことがわかった。
2. Kazuaki Hanawa, Sho Yokoi, Satoshi Hara, Kentaro Inui. Evaluation of Similarity-based Explanations. The Ninth International Conference on Learning Representations (ICLR). 2021. 26 pages.
自然言語処理に限らず機械学習を用いたモデル全般に対してその解釈性が重要なトピックとなっている。この研究では訓練事例とテスト事例の類似性の測りかた対して規範的な指針を示した。自然言語処理と画像処理にまたがる多くのデータセットと手法を用いた検証により、表現学習および埋め込み空間の幾何について 2 点の発見があった。1 点目の発見は、単

語埋め込みのように中間表現自体を特徴量にするよりも、勾配ベクトルを特徴量にした方が解釈性の観点ではすぐれた性能を示すこと。2 点目の発見は、勾配ベクトルを特徴量とした場合、そのノルムが対象の外れ値である度合いを示し結果として勾配ベクトル同士の余弦が事例同士の類似性をよく捉えたことである。
3. *大山 百々勢, *横井 祥, *下平 英寿. 単語ベクトルの長さは意味の強さを表す. 言語処理学会 第 28 回年次大会 発表論文集. 2022. pp.1050-1055. (* equally contributed) / Momose Oyama, Sho Yokoi, Hidetoshi Shimodaira. Norm of word embedding encodes information gain. arXiv preprint. 2022. 9 pages. under review.
単語ベクトルの長さが単語の意味の強さを表すことを理論的・経験的に示した。はじめに対照学習を用いた典型的な学習アルゴリズムから得られる単語ベクトル空間を情報幾何の言葉で特徴づけ、単語ベクトルのノルムと「その単語が存在することで周辺単語の分布がどの程度歪むか」が対応することを示した。また、単語頻度の交絡を除去してもこの性質が担保されていることを経験的に示した。

（2）特許出願

研究期間全出願件数：0 件（特許公開前のものも含む）

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

- 招待講演
  - 横井 祥.「言葉の形を教えてくれる自然言語処理」. IPSJ-ONE 2022 招待講演. 2022 年 3 月.
  - 横井 祥.「最適輸送と自然言語処理」. 言語処理学会第 28 回年次大会 チュートリアル講演. 2022 年 3 月.
  - 横井 祥.「構造を持った言語データと最適輸送」. NAIST DSC NLP Seminar 2022 Summer チュートリアル講演. 2022 年 9 月.
  - ほか計 13 件.
- 受賞
  - 言語処理学会第 27 回年次大会 優秀賞.
    横井 祥, 下平 英寿.「単語埋め込みの確率的等方化」.
  - ほか計 5 件.
  - これらとは別に, 共著研究の共著者への受賞 6 件.