

情報ボトルネック法を用いた視線・脳情報デコード手法の提案

原田宥都 大関洋平

東京大学大学院 総合文化研究科言語情報科学専攻

{harada-yuto, oseki}@g.ecc.u-tokyo.ac.jp

概要

人間がテキストを読む際の視線や脳活動の記録は、これまで主に人間の認知機能の理解のために用いられてきたが、近年ではこのような人間の認知データが自然言語処理タスクにおいても有益であることが分かってきた。しかし、認知データは特定の言語処理タスクにとって多くの無関係な情報を含むため、効率的に用いるのが難しく、また、なぜタスクの改善に寄与するのかについてもほとんど議論がなされていない。本稿では、視線や脳情報から特定のタスクに有益な情報のみを取り出す認知情報ボトルネック法を提案し、その性能を検証する。

1 はじめに

眼球運動や脳活動の記録は、人間の持つ言語能力がどのように機能するのかを理解する上で、重要な役割を担ってきた [1, 2]。近年の研究では、アイトラッキングデータや EEG により記録された脳波を用いて自然言語処理タスクを改善することができると報告されており [3]。この効果は固有表現認識や関係抽出、センチメント分析など、複数のタスクにおいて一貫して確認されている。これらの結果から、視線や脳波などの情報が統語的情報、語彙的情報、あるいは言語処理の際の認知負荷など様々な情報を含んでおり、それらが自然言語処理タスクにおいて有益な特徴となっていることが予想される。しかしながら、それぞれの認知データは言語処理の全ての過程を反映した情報であり、特定のタスクにとって無関係な情報を含んでいる。特に脳波は一般的に高次元なデータであり、特定のタスクでどの脳領域の情報があるか、あるいはどの周波数帯域の脳波が有用か、などを事前に知ることは困難である。

そこで、本研究では、認知データから特定のタスクにとって有益な情報のみを取り出すための手法として、認知情報ボトルネック法を提案する。具体的には、一般的な情報ボトルネック法 [4] を継承し、

それを同時に記録された視線と脳波に適用することで、マルチモーダル設定での学習に用いることのできる単一の埋め込み表現へ変換する。情報ボトルネック法の目的関数はタスクの精度を最大化しつつ入力情報をなるべく圧縮するものであり、視線や脳波の持つ様々な情報からタスクにとって必要な情報のみを取り出すことができる。またその際には次元数を削減することができるため、次元数に大きく差のある視線と脳波を同時に学習に用いることが容易になる。

実験では、提案手法を用いて圧縮した認知データを用いることで、圧縮を行わなかった認知データを用いた時よりもタスクの精度が改善されることを検証した。その結果、圧縮した認知データは、行わなかったものやその他の単純な圧縮手法に比べて精度を改善することが分かった。

2 関連研究

2.1 人間の認知データを用いた言語情報のデコーディング

視線や脳波などの人間の認知データには、言語処理の過程など豊富な情報がエンコードされることが知られている。例えば、自然言語処理分野における脳情報デコーディングの研究には、脳情報のみから発話内容を推定するスピーチデコーディング [5] や、脳情報から認識している単語の品詞を推定するもの [6]、脳情報から読んでいる物語文の内容について推定するもの [7] など、様々な研究が存在する。近年では自然言語処理タスクの改善を目的とした脳情報デコーディングも行われており、特に Hollenstein ら [8] は視線と脳波を同時に記録したデータセット [9] を作成し、マルチモーダル設定での実験を行っている。彼らの実験では、周波数帯域を絞って脳波の一部を用いることで、脳波をそのまま用いるよりも精度が向上することが報告されている。タスクに対する認知データのノイズを除去するための手法は現在も検討されている途中であり、標準的な手法は

いまだ定まっていない。

2.2 情報ボトルネック法

情報ボトルネック法 (Information Bottleneck, IB) [4] とは、情報理論の分野に起源を持つ手法であり、近年では機械学習において利用されている。ある信号 $x \in X$ から別の信号 $y \in Y$ を予測する際に、予測に無関係である冗長な情報を取り除いた X の圧縮表現 T を見つけることが目的であり、次のように記述できる。

$$\mathcal{L}_{IB} = \beta I(X, T) - I(T, Y) \quad (1)$$

これは、相互情報量 $I(X, T)$ をなるべく最小化しながら $I(Y, T)$ を最大化すること、つまり、圧縮表現 T が X についての情報を最大限破棄しつつも、 T が Y を予測するために必要な情報は最大限保持するということを意味している。

実際に機械学習で用いる際は、[10]において(1)の効率的な変分推定が提案されており、以下のように導出された目的関数を用いる..

$$\mathcal{L}_{VIB} = \beta \mathbb{E}_x [\text{KL}[p_\theta(t|x), r(t)]] + \mathbb{E}_{t \sim p_\theta(t|x)} [-\log q_\phi(y|t)] \quad (2)$$

ここで $r(t)$ は $p(t)$ の変分近似、 $q_\phi(y|t)$ は $p(y|t)$ の変分近似である。 $p_\theta(t|x)$ と $q_\phi(y|t)$ は独自のパラメータセットを持つニューラルネット、つまりエンコーダとデコーダであることを意味している。

3 実験

3.1 認知情報ボトルネック法

本研究で用いるモデルの概要図を図1に示す。実験では固有表現認識タスクを行う。

3.1.1 入力特徴

視線特徴 視線特徴は固視回数、第一固視時間、総読み時間、注視時間、視線通過時間の5つからなる。1単語あたり5つの数値を用いた5次元のベクトルとして扱った。視線情報を用いて固有認識表現の精度を改善した先行研究[11]ではより細かく特徴を整理し、1単語あたり全17の特徴から視線特徴を構成しているため、本実験の視線特徴はより基礎的な構成であり、そのままの値ではタスクに有益にはなりづらいと予想される。

脳波特徴 本実験では、105個の電極から記録されたEEGデータを脳波特徴として用いた。視線特

徴との対応を取るため、脳波特徴は総読み時間に対応する数値を取り出している。また、周波数帯域ごとにそれぞれ θ (4-8 Hz), α (8.5-13 Hz), β (13.5-30 Hz), γ (30.5-49.5 Hz) とデータを4つに分割している。そのため、1単語あたり420次元のベクトルとして脳波特徴を扱っている。

単語の埋め込み表現 視線特徴、脳波特徴と共に、言語モデルによる単語の埋め込み表現を用いる。視線情報を用いて固有認識表現の精度を改善した先行研究[11]ではGlove[12]を用いているため、比較のために本実験ではGloveを用いて実装した。そのため、1単語あたり100次元のベクトルとして単語の埋め込み表現を扱っている。

ノイズと認知データの比較 認知データを入力特徴として用いた際にタスク精度が改善された場合、それが本当に認知データの影響であるかどうかを確認するために、同じ次元数のノイズベクトルを用いて比較実験を行う。視線特徴に対しては5次元の、脳波特徴に対しては420次元のランダムノイズを用意する。

3.1.2 モデル内部のアーキテクチャ

情報ボトルネック法による圧縮 認知データである視線特徴と脳波特徴については、情報ボトルネック法を用いた圧縮を行う。最適化は固有表現認識のデコーダの出力精度を改善するように行われる。視線特徴については5次元から3次元へ圧縮し、脳波特徴については420次元から50次元へ圧縮する。これにより、情報ボトルネック法で圧縮された表現は固有表現認識のために必要な情報は残しつつ、それ以外の冗長な情報を捨てるように学習する。また、情報ボトルネック法を用いた結果と比較するために、単純な主成分分析を用いた次元削減も行う。

BiLSTM-CRF デコーダ 視線特徴、脳波特徴、単語の埋め込み表現は結合し、1つの入力特徴として扱う。モデルの後段では、視線情報を用いて固有認識表現の精度を改善した先行研究[11]との比較を行うために、同様のモデルであるBiLSTM-CRF[13]を用いている。

3.2 データセット

本実験では、データセットとしてZurich Cognitive Language Processing Corpus (以下ZuCo)[9]を用いた。このコーパスはアイトラッキングとEEGの記録を同時に行ったデータセットであり、12人の被験者に

れる。

	P(%)	R(%)	F(%)
Glove	76.14	77.01	76.57
Glove + EEG	79.12	66.06	72.00
Glove + Noise(dim=420)	75.28	61.47	67.68
Glove + IB(EEG)	83.84	76.15	79.81

表2 脳波特徴を用いた固有表現認識の結果

次に、脳波特徴を用いた実験の結果を表2に示す。この実験では、脳波特徴とノイズの次元数が共に420次元であり、Gloveベクトルの100次元と比べても大きい。次元数の大きいノイズを含む入力特徴に大きく影響を受けたためか、Glove + EEG, Glove + Noiseの設定ではどちらも精度が大きく下がっている。420次元の脳波特徴(EEG)を扱うにはモデルのサイズが十分でなかったと考えられる。しかし、こちらの実験設定においても情報ボトルネック法を用いて脳は特徴を圧縮した結果が一番精度が高くなった。また、視線特徴を用いた際の結果と比べても改善幅はより大きかった。主成分分析による圧縮と比べても精度が高く、情報ボトルネック法を用いたモデルの方が、ノイズの多い脳波の特徴からタスクに有用な情報を取り出すのに適しているのではないかと考えられる。

	P(%)	R(%)	F(%)
Glove	76.14	77.01	76.57
Glove + Gaze + EEG	82.72	61.47	70.53
Glove + Noise(dim=425)	77.91	61.47	68.72
Glove + IB(Gaze) + IB(EEG)	81.37	76.15	78.67

表3 視線特徴と脳波特徴の両方を用いた固有表現認識の結果

最後に、視線特徴と脳波特徴の両方を用いた実験の結果を表3に示す。この実験では、脳波特徴のみを用いた時の結果と同様に、次元数の大きいノイズを含む入力特徴に大きく影響を受けたためか、Glove + Gaze + EEG, Glove + Noiseの設定ではどちらも大きく精度が下がっている。また、情報ボトルネック法によって視線特徴と脳波特徴の両方を圧縮しているGlove + IB(Gaze) + IB(EEG)の設定では、Glove + IB(EEG)よりもやや精度が低くなった。用いている特徴としては後者の方が少ないので、情報ボトルネック法を用いた学習はあまり精度が安定していない可能性がある。

3.5 考察

実験全体を通して言えることとして、今回用いたモデルでは、入力特徴を全て単純に結合しているだけであるため、特徴の次元数の差が大きく影響してしまうということが挙げられる。モデルアーキテクチャの改善方法として、認知データを用いたマルチモーダル設定での学習であると捉え、それぞれの入力特徴の重要性を次元数に左右されない形で扱うことが望ましいかもしれない。また、認知データをそのまま用いても精度が改善されるとは限らないということも分かった。それと同時に、情報ボトルネック法による圧縮や、単純な主成分分析など、必要に応じた次元削減はこのような認知データを扱う際には重要な要素であった。今回の情報ボトルネック法のハイパーパラメータ設定については、より詳しく探索することで、さらに性能を向上させることができるかもしれない。

今後は、Gloveベクトルよりもより高性能なBERT[14]等の単語埋め込み表現を用いた際にも、今回のような認知データによる改善が見られるのか、といったことや、BiLSTM-CRFデコーダをさらに最先端の固有表現認識デコーダに変更した場合にどのような影響があるのか、といった、実験設定による結果への影響を調べる必要がある。それに加えて、情報ボトルネック法によって認知データから取り出された情報はどのような内容であるのかを分析によって調べることで、なぜ認知データが自然言語処理タスクの改善に寄与することがあるのかを調べていく。

4 結論

本稿では、視線情報や脳活動情報といった人間の認知データを、自然言語処理タスクの改善により効率的に用いるための手法である認知情報ボトルネック法を提案した。視線特徴や脳波特徴と単語埋め込み表現を単一の入力特徴として扱えるように変換し、認知データのノイズを除きつつ次元数を調整することができる本手法は、実験の結果、人間の認知データを用いて固有表現認識タスクの精度を改善することに成功した。今後はより詳細な実験設定を検証しつつ、圧縮された認知データの内容を分析することで、人間の認知データがどのようにして言語処理タスクの改善に役立つのかを調べていく。

謝辞

本研究は、JST さきがけ JPMJPR21C2 の支援を受けたものです。

参考文献

- [1] Brian Murphy, Leila Wehbe, and Alona Fyshe. Decoding language from the brain. **Language, cognition, and computational models**, pp. 53–80, 2018.
- [2] Shouyu Ling, Andy CH Lee, Blair C Armstrong, and Adrian Nestor. How are visual words represented? insights from eeg-based visual word decoding, feature derivation and image reconstruction. **Human brain mapping**, Vol. 40, No. 17, pp. 5056–5068, 2019.
- [3] Nora Hollenstein, Maria Barrett, Marius Troendle, Francesco Bigioli, Nicolas Langer, and Ce Zhang. Advancing nlp with cognitive language processing signals. **arXiv preprint arXiv:1904.02682**, 2019.
- [4] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. **arXiv preprint physics/0004057**, 2000.
- [5] David A Moses, Sean L Metzger, Jessie R Liu, Gopala K Anumanchipalli, Joseph G Makin, Pengfei F Sun, Josh Chartier, Maximilian E Dougherty, Patricia M Liu, Gary M Abrams, et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. **New England Journal of Medicine**, Vol. 385, No. 3, pp. 217–227, 2021.
- [6] Alex Murphy, Bernd Bohnet, Ryan McDonald, and Uta Noppeney. Decoding part-of-speech from human eeg signals. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2201–2210, 2022.
- [7] Leila Wehbe, Brian Murphy, Partha Talukdar, Alona Fyshe, Aaditya Ramdas, and Tom Mitchell. Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. **PloS one**, Vol. 9, No. 11, p. e112575, 2014.
- [8] Nora Hollenstein, Cedric Renggli, Benjamin Glaus, Maria Barrett, Marius Troendle, Nicolas Langer, and Ce Zhang. Decoding eeg brain activity for multi-modal natural language processing. **Frontiers in Human Neuroscience**, p. 378, 2021.
- [9] Nora Hollenstein, Jonathan Rotsztein, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. **Scientific data**, Vol. 5, No. 1, pp. 1–13, 2018.
- [10] Alex Alemi, Ian Fischer, Josh Dillon, and Kevin Murphy. Deep variational information bottleneck. In **Proceedings of the International Conference on Learning Representations**, 2016.
- [11] Nora Hollenstein and Ce Zhang. Entity recognition at first sight: Improving ner with eye movement information. **arXiv preprint arXiv:1902.10068**, 2019.
- [12] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In **Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)**, pp. 1532–1543, 2014.
- [13] Meizhi Ju, Makoto Miwa, and Sophia Ananiadou. A neural layered model for nested named entity recognition. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 1446–1459. Association for Computational Linguistics, 2018.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186. Association for Computational Linguistics, June 2019.