

音楽推薦におけるセマンティックギャップと認識ギャップの解消 Eliminating Semantic Gaps and Recognition Gaps in Music Recommendations

岡田 祐久[†] 島川 博光[†] 原田 史子[†]
Tasuku Okada Hirotsu Shimakawa Fumiko Harada

1. はじめに

音楽ストリーミングサービスの普及により、ユーザの多様な嗜好に対応した推薦システムの重要性が高まっている。しかし、従来手法ではユーザの曖昧な要求や感情的ニュアンスを的確に捉えることが難しく、推薦結果との間に「認識ギャップ」が生じる問題がある。さらに、推薦理由が不明瞭であることから、ユーザの納得感や信頼性を損なう恐れもある。本研究では、BERT を用いた分類モデルに対し、Triplet Loss による意味表現の学習を導入し、ユーザ意図に沿った Attention 配置と推薦結果の向上を目指す。加えて、Attention の可視化により推薦根拠を明示し、ユーザとのギャップを視覚的に把握・修正できる仕組みを構築する。

2. 関連研究

2.1 Attention 機構

Attention とは、AI がタスクをこなす際に「そのタスクには文のどの単語が重要か、どの単語に注目すべきか」を決めるための機構である[1]。この機構により、従来モデルの「長い文になると精度が落ちる」という欠点を克服し、最先端技術である BERT[2]、GPT[3]などの起源となった。本研究では Attention を可視化した AttentionMap を用いることでユーザに対し「どこに着目して推薦を行ったか」について説明することでユーザがモデルとの意図の差、認識ギャップを理解し、適切なフィードバックを行える仕組みをつくる。

2.2 Triplet Loss を用いた表現学習

Triplet Loss は、画像認識や自然言語処理など、埋め込み表現の学習において広く用いられる距離学習手法である。Anchor (基準)・Positive (類似)・Negative (非類似) の 3 つのサンプル間の距離関係を保つように学習を行うことで、意味的に近いデータは近く、異なるデータは遠くなるような埋め込み空間を獲得できる。本研究では、ユーザからのクエリ文に対し意味的に近い・遠いコメントを選定し、Triplet Loss により BERT の表現空間を調整することで、Attention の配置をユーザの意図に近づけることを試みる。

3. 提案手法

本研究では、ユーザの要求に基づいて楽曲を推薦する分類タスクにおいて、モデルの推薦根拠がユーザの意図と乖離する「認識ギャップ」の問題を解消するための Attention の変更手法として、Triplet Loss を用いた学習を導入する。

従来のアプローチでは、事前学習済みの BERT を用い、コメントを入力として曲名を分類するモデルをそのまま学習する手法が一般的である。しかし、この方法では Attention が文中の意味的に重要でない語に向けられること

が多く、ユーザの意図と一致しない推薦が行われる可能性がある。

そこで本研究では、認識ギャップを解消するための分類モデルの性能向上と Attention の人為的変更を目的として、Triplet Loss による意味表現の事前学習を実施する。この Triplet Loss により、ユーザの要求に対する意味空間上の距離関係が学習され、よりユーザ意図に沿った特徴抽出が可能になり、また、Attention を意図したように変更することができる。

3.1 Triplet Loss による事前学習

第 2 章で示した Triplet Loss 学習を利用して BERT を学習する。これにより、入力文の意味的構造をより鋭敏に捉える埋め込み表現が構築され、Attention の配置にも変化が生じる。本研究では、ChatGPT により作成された 50 件のクエリ文と YouTube のコメント 2500 件を用意した。50 個のクエリ文に対し、正例 3 件、easy 負例 2 件、hard 負例 1 件を手動で構成し、Triplet Loss 学習を行う。

3.2 分類タスクによるファインチューニング

Triplet Loss により事前学習された BERT の重みを初期値として用い、音楽推薦の分類タスクのための学習を行う。学習データには 2024 年のカラオケ DAM ランキング[4]の上位 50 曲について、MV に投稿されたコメントをいいねが多い順に 50 コメントずつ、計 2500 コメントを収集し、コメントを説明変数、楽曲名を目的変数とする分類タスクを行う。この学習により、クエリ文を与えた際、最も確率の高い楽曲が出力として与えられる。

4. 評価と考察

Triplet Loss 学習の有用性を評価するために、本研究では 2 つの実験を行う。1 つ目は Attention 比較実験であり、2 つ目は推薦結果比較実験である。比較モデルとして、ファインチューニングなしの BERT モデル (シンプル BERT)、分類学習のみを行った BERT モデル (分類 BERT) を用意し、Triplet Loss 学習と分類学習を行った BERT モデル (提案手法) を含めた 3 手法について、クエリ文を与えた際の AttentionMap と推薦曲を比較する実験を行った。なお、シンプル BERT には分類機能がないため、AttentionMap の比較のみを行う。また、シンプル BERT に関しては日本語テキストの処理に広く用いられている cl-tohoku/bert-base-japanese [5]を、分類 BERT に関しては第 3 章の分類タスクによるファインチューニングで示した学習を行う。

4.1 Attention 比較実験

実験 1 として同一のクエリ文を与えた際の AttentionMap の比較を行った。図 1 はその結果である。

まずシンプル BERT について、クエリ文の単語には Attention はほとんど向いておらず、SEP にのみ強くかかっていることが分かる。このことからシンプル BERT はクエ

リ文の意味処理ができておらず、どこに注目すればよいか理解していないといえる。

次に分類 BERT について、クエリ文中のすべての自立語に対し、高い Attention がかかっていることが分かる。このことからシンプル BERT に比べ 1 つ 1 つの単語に意識して推薦できることが分かる。しかし、「静か」、「サビ」、「盛り」などの推薦に重要な単語だけでなく、「部分」という分類には有効でなさそうな単語にも Attention が強く向いていることから、まだ注意が散乱しており、推薦には不十分である。

最後に提案手法について、分類 BERT と同様に、シンプル BERT に比べて自立語に Attention がかかっているのがわかる。さらに、「静か」、「サビ」、「盛り」のみに Attention が強くかかっており、そのほかの自立語にはあまりかかっていないことから推薦に重要な単語のみに着目できているといえる。

以上より、提案手法はシンプル BERT に比べ、単語に着目しており、さらに分類 BERT に比べ、より推薦する上で重要な単語のみに着目できているといえる。

シンプルBERT:

[CLS] 静かな部分もあり、サビは盛り##上がるような曲 [SEP]

分類BERT:

[CLS] 静かな部分もあり、サビは盛り##上がるような曲 [SEP]

提案手法:

[CLS] 静かな部分もあり、サビは盛り##上がるような曲 [SEP]

図 1 3手法の AttentionMap の比較

4.2 推薦結果比較実験

実験 2 として内容に少しの差異がある 2 つのクエリ文を作成し、分類 BERT と提案手法に与えた際の Attention と推薦結果を比較した。図 2、表 1 はそれぞれ 2 つのクエリ文 α 、 β を与えた際の Attention と推薦結果である。図 2 の分類 BERT について、クエリ α では「失恋」、「元気」など、重要な単語に着目できていることが分かる。しかしわずかに内容を変えたクエリ β では「失恋」にのみ大きな Attention がかかっており、重要単語である「楽しかった」という部分に着目できていないことが分かる。

さらに表 1 について、赤は歌詞や曲調が前向きな曲、青は後悔を歌った曲、紫は過去の思い出を歌った曲である。なお、スコアは分類確率を表している。クエリ内容との対応から、クエリ α に対しては赤の曲、クエリ β に対しては紫、次いで青が推薦され、それぞれのクエリで結果が異なることが望ましい。しかし、分類 BERT の結果より、クエリ間で分類確率の高い曲がほとんど変化していないことが分かる。加えて、両クエリとも青色の曲が分類確率の大半を占めており、特にクエリ α に対しては不適切な曲が上位であることが分かる。これらの結果から、分類 BERT は「失恋」のみに強く着目しており、クエリの変更点である「元気」と「楽しかった思い出」という区別がつかず無視していることが分かる。

次に図 2 の提案手法について、クエリ α では分類学習と同じく「失恋」、「元気」など、重要な単語に着目できていることが分かる。さらに、クエリ β では「失恋」、「楽しかった」という部分に Attention が強くかかっており、

両クエリで重要な単語に着目できていることが分かる。さらに表 1 について、クエリ α では分類 BERT に比べ、分類確率が高いものの中に赤色の曲が多く含まれていることが分かる。また、クエリ β では赤色の曲が減少しており、さらに紫色の曲も追加されていることが分かる。これらのことから、提案手法は「失恋」という要素のみに注目しておらず、「元気が出る曲」と「楽しかった思い出に浸れる曲」という細かいニュアンスの違いを少し理解できているということが分かる。

以上より、TripletLoss 学習により、通常の学習では学習できないような「人間の感覚」をモデルが学習できたり、また Attention を理想的な形に変更できることが分かる。

分類BERT

クエリ α [CLS] 失恋 後悔 後に 元気 が 出る 曲 [SEP]

クエリ β [CLS] 失恋 後悔 後に 楽し ## かった 思い出 に 浸 ## れる 曲 [SEP]

提案手法

クエリ α [CLS] 失恋 後悔 後に 元気 が 出る 曲 [SEP]

クエリ β [CLS] 失恋 後悔 後に 楽し ## かった 思い出 に 浸 ## れる 曲 [SEP]

図 2 手法別クエリ変更と Attention の変化

表 1 手法別クエリ変更と推薦曲の変化

	分類BERT		提案手法	
	クエリ α	クエリ β	クエリ α	クエリ β
1 最後の雨	0.95	最後の雨	0.70	サウダージ
2 花束の代わりにメロディーを	0.01	366日	0.13	チェリー
3 366日	0.01	Pretender	0.05	最後の雨
4 Pretender	0.00	カブトムシ	0.01	First Love
5 ビリミオン	0.00	ランデブー	0.01	ドライブラワー
6 LOVE YOU	0.00	サウダージ	0.01	Pretender
7 カブトムシ	0.00	花束の代わりにメロディーを	0.01	チェリー
8 シングルベッド	0.00	恋人ごっこ	0.01	ビリミオン
9 チェリー	0.00	糸	0.00	シャルル
10 世界が終わるまでは...	0.00	ベテルギウス	0.00	Pretender
				高嶺の花子さん
				366日
				小さな恋の歌
				アイノカタチ

5. おわりに

本研究では、ユーザの要求文に対する推薦精度と説明可能性の向上、Attention の人為的変更を目的として、BERT に対する TripletLoss による事前学習と分類タスクによるファインチューニングを組み合わせた推薦手法を提案した。

評価実験では、従来手法と比較して、Attention の配置や推薦結果において提案手法がより適切な出力を行う傾向が示された。今後は、Triplet とコメントデータを増やすことで、より柔軟な推薦システムの実現を目指す。

参考文献

- [1] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*.
- [2] Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language
- [3] OpenAI. (2022, Nov.) Chatgpt: Optimizing language models for dialogue. <https://openai.com/blog/chatgpt/>
- [4] [カラオケランキング TOP100\(最新/週間/月間\)](https://mora.jp/topics/news/karaoke2410) | [カラオケ DAM 公式](https://mora.jp/topics/news/karaoke2410) <https://mora.jp/topics/news/karaoke2410> 参照 2024-11-01
- [5] Tohoku University Inui-Suzuki Laboratory, "cl-tohoku/bert-base-japanese", Hugging Face, <https://huggingface.co/cl-tohoku/bert-base-japanese>