

大規模言語モデルを用いた役職判定技術の 複数モデル評価

Multi-Model Evaluation of Job Title Classification Techniques

Using Large Language Models

小川 秀貴¹ 高阪 崇久¹ 小林 史弥¹

OGAWA HIDEKI¹ KOSAKA TAKAHISA¹ KOBAYASHI FUMIYA¹

1. はじめに

効率的なマーケティングにおいては、顧客属性を正しく把握することが重要である。例えば、自社の商品やサービスの提案の際には、その属性に応じて必要となる商品やサービスが異なるため、顧客属性に適した提案を行うことで、顧客に対してより質の高い提案が可能となる。

また、商品やサービスの利用においては、顧客の役職によって、サービスや商品の導入に向けた障壁の度合いが異なると考えられる。例えば、一般的には、部長や課長といったポジションは決裁権限を持つことも多く、そういった方々に直接アプローチすることができれば、迅速な意思決定のもと商品やサービスの受注につながる可能性が高いと見込まれる。しかし、顧客の役職情報は、企業によってかなりばらつきがあり、その情報をマーケティングに有用な形に落とし込むのは容易ではない。

そこで本稿では、近年急速に利活用が進んでいる大規模言語モデル (以降、LLM)[1] を利用し、企業ごとに表記の異なる役職情報に対して、マーケティングに利活用が可能な形式で役職判定をする技術を提案する。また、複数の LLM のモデルによる評価も行い、どのモデルが一番良いかについても述べる。

2. 大規模言語モデルによる役職判定

2.1 役職の定義

役職判定では、以下の 3 つの役職のいずれかに分類を行う。

- 経営層
- 中間管理職

• 担当者

正解ラベルの付与に際して、経営層は、代表取締役や専務など経営に携わる役職とし、中間管理職は、経営層ではない、部長や課長、係長、主査などとし、担当者は経営層や中間管理職に該当しないそれ以外と定義を行った。

2.2 プロンプト

役職判定では、カンマで区切られた部署名のリストを含むプロンプトを LLM に入力することで、部署名を定義に基づく 3 分類のいずれかに分類を行うことで実施する。また、プロンプトに一度に入力する部署の数は最大で 10 個までとする。

使用するプロンプトは以下のとおり。

Instruction

以下の役職をいずれかに分類してください。

ただし、回答は数字のみとし、それ以外の文字列を含めないでください。

また、出力形式は、Output format に準拠してください。

0: 経営層 1: 中間管理職 2: 担当者

Input

本部長,CEO,IT 担当

Output format

```
{  
  "本部長": ,  
  "CEO": ,  
  "IT 担当":  
}
```

¹ 東日本電信電話株式会社
NIPPON TELEGRAPH AND TELEPHONE EAST CORPORATION

3. 実験

3.1 データセット

今回の実験で用いたデータセットは、実際に存在する企業の部署名が合計 292 件であり、役職の内訳は表 1 の通りである。なお、役職の正解ラベルは、部署名から筆者らによって手動でラベル付与を実施した。

役職	人数
経営層	74
中間管理職	181
担当者	37

表 1 データセットの内訳

3.2 使用した LLM

今回使用した LLM のモデルは、以下の 3 つである。

- gpt-4o-mini[2]
- claude3-haiku[3]
- llama-3.3-70b[4]

4. 評価

4.1 正解率

各モデルごとに予測された結果から計算された正解率は、表 2 のとおり。

モデル	正解率 (%)
gpt-4o-mini	85
claude-3-haiku	80
llama-3.3-70b	76

表 2 モデル別正解率

4.2 エラー分析

以下の 3 つのケースについて、エラー分析を実施した。

4.2.1 全ての LLM で誤分類したケース

全ての LLM で誤判定したケースは、19 件存在した。「次世代事業開発統括アドバイザー」や、「教授」の 2 件を除く、17 件ではすべての LLM で誤判定した結果が一致しており、LLM での差分はほとんど見られなかった。

4.2.2 一番精度の良いモデルのみ正解したケース

一番正解率の高い gpt-4o-mini でのみ正解したケースに着目する。

gpt-4o-mini でのみ正解し、その他の 2 モデルで不正解となったのは、14 件存在した。そのうち、主査やチーフといった中間管理職を担当者と誤判定したケースが 9 件、それ以外が 5 件という内訳となった。

4.2.3 一番正解率の高いモデルで中間管理職を担当者と誤判定したケース

一番正解率の高い gpt-4o-mini で、中間管理職を担当者と誤判定したケースは、18 件存在した。そのうち、課長・係長、主査やチーフを担当者と誤判定したケースが、12 件であった。残る 6 件は、「シニアリサーチャー」や「シニアコンサルタント」といった、人による判断をした場合にも選択に迷うことが十分考えられるケースであった。

4.3 考察

エラー分析の結果、以下の知見が得られた。

- モデルによる正解率の差は見られる
- モデルによる改善は見込めないケース（人による判断でも難しいケース）が存在する
- 主査やチーフといった役職は誤判定しやすい傾向

上記を踏まえ、さらなる改善のために、正解率が良いモデルを利用したうえで、誤判定をしやすことが経験的にわかっているケースや、判断が難しい役職について、プロンプトにおいて Few-shot で例示をすることが考えられる。

5. おわりに

本稿では、LLM を利用した役職判定技術を提案した。複数モデルによる評価の結果、一番正答率の高い gpt-4o-mini というモデルでは、85%を得ることができた。

また、今後に向けては、考察で挙げられたような手法によるさらなる精度の改善や、LLM が判断に迷うケースにおいて、複数ラベルを付与するといった運用面での対処などが想定される。

参考文献

- [1] Gartner 2024 Hype Cycle for Emerging Technologies Highlights Developer Productivity, Total Experience, AI and Security, 入手先 (<https://www.gartner.com/en/newsroom/press-releases/2024-08-21-gartner-2024-hype-cycle-for-emerging-technologies-highlights-developer-productivity-total-experience-ai-and-security>) (2025.04.17).
- [2] GPT-4o mini, 入手先 (<https://platform.openai.com/docs/models/gpt-4o-mini>) (2025.04.17).
- [3] Claude 3 Haiku: our fastest model yet, 入手先 (<https://www.anthropic.com/news/claude-3-haiku>) (2025.04.17).
- [4] meta-llama Llama-3.3-70B-Instruct, 入手先 (<https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct>) (2025.04.17).