

大規模言語モデルのためのサービスに対する口コミを用いた入力自動補完の検討 Automated Input Supplement which Used User Reviews for Servises of Large Language Models

笹本 まこと¹⁾ 堀川 達平²⁾ 北山 大輔¹⁾
Makoto Sasamoto Tappei Horikawa Daisuke Kitayama

1 はじめに

ChatGPT の公開を皮切りに大規模言語モデルは、急速に利用が広まっている。人々の話題の中心となり、活用の場面がますます増えようとしている。現時点、ChatGPT はどんな入力に対してもそれなりの結果を返すことができる。しかし、入力が曖昧な場合はその限りではない。出力もありきたりな内容となり、入力した人の意図した出力にならないといったことが起こる。

例えば、あるサービスに対する問い合わせをする時に、サービスに関する知識の浅い人や、サービスを利用したことが無い、またはこれからサービスを利用するといった人々の入力は、曖昧な内容になりやすいと考えられる。こうしたユーザの入力に情報を補完することで、より具体的な出力が可能になり、ユーザの疑問解決や意思決定につながると考える。

Web 検索ではこのような時、曖昧な検索キーワードに対し、自動的なキーワード補完を行う。このような研究は以前から行われている [1]。本研究では、Web 検索ではなく、大規模言語モデルによるチャットを対象とする。具体的には、ユーザの入力に関連する情報を自動的に補完することで、大規模言語モデルが情報として持っている可能性のある、該当サービスの詳細情報を事前に与えるシステムを構築する。入力内容に対象サービスの専門的な内容を含めるためには、サービスに対する口コミデータを用いて、サービスの良い特徴・悪い特徴を自動的に抽出し、入力に自動補完する。

2 提案手法

本研究で提案するシステムの概要を図 1 に示す。この一連の流れに沿って、入力の自動補完を行う。

2.1 サービス名の抽出

ユーザからサービスに対する問いかけが入力された時、サービス名を判定するため、まず、固有表現抽出を行う。固有表現抽出には Google Natural Language API¹⁾を用いる。Google Natural Language API によってエンティティを識別し、それぞれにラベル付けをする。ラベル名が固有名詞、行事、組織、場所の順に当てはまるものがあつた時、それをサービス名の候補と判定する。ここで取り出した候補が口コミデータセット内においてどのカテゴリに分類されるのかを調べ、データセット中から該当単語を含む口コミを収集する。

ここで、口コミデータセットとして、株式会社マイスタースタジオが提供する「みんなの評判口コミデータセット」²⁾ [2]を用いる。このデータセットは、2017 年 1 月 16 日から 2022 年 7 月 12 日、約 4,700 の商品・サービスに対する全 166,376 件の口コミデータである。このデータセットでは、口コミ対象のサービス名が「商品

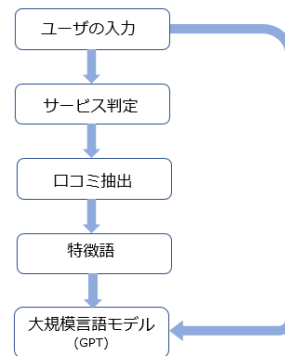


図 1 提案手法の概要図

表 1 補完情報付き入力文のテンプレート

[サービス名] は以下の特徴を持ちます。
特徴: [15 個の特徴語]
[ユーザの入力文]

名」の属性に、サービスのカテゴリが「カテゴリ名」の属性に格納されている。そのため、候補語に関して「商品名」「カテゴリ名」の順に該当するものがあるかを照合し、該当するものがあれば、それをサービス名として判定する。

2.2 口コミから特徴語の抽出

入力に該当するサービスの口コミから情報を抽出することで、Web 上位に表示される情報と比べて、サービスの短所や実際に利用した人の感想の情報を得ることができる。該当口コミに頻出する、サービスを実際に体験した人の後悔や注意点は、本システムのユーザにとっても有益な情報になると考えられる。収集した口コミの中から mecab を用いて名詞を抽出し、頻出順に上位 15 件の単語を取り出す。これを特徴語とする。この時、mecab の辞書としては ipadic-neologd を用いている。また、ストップワードは、著者が定義したものを使用している。ここで得られた 15 個の特徴語をユーザの元の入力文への補完情報とする。本システムでは、ユーザの問い合わせ入力と、補完情報である特徴語 15 個を表 1 のテンプレートに当てはめ、大規模言語モデルへの入力とする。単語が補完された入力からは、特徴語を含めた、より具体的なアドバイスが生成されると期待する。

3 実行例と考察

「ダイエット食品を選ぶときの注意点は？」という入力を例に、システムの動作を説明する。実行には、OpenAI が提供する gpt-3.5-turbo³⁾ を大規模言語モデルとして用いた。また、提案手法による補完を行う場合と、行わない場合の出力を比較し、提案手法の効果について考察する。

ユーザの入力に対し、「ダイエット食品」がサービス

3) <https://platform.openai.com/docs/api-reference/chat>

1) 工学院大学情報学部システム数理学科

2) 工学院大学大学院工学研究科情報学専攻

1) <https://cloud.google.com/natural-language?hl=ja>

2) <https://www.nii.ac.jp/dsc/idr/minhyo/>

表 2 入力補完前の入力

ダイエツト食品を選ぶときの注意点は？

表 3 入力補完前の出力

ダイエツト食品を選ぶときの注意点は以下の通りです。

1. 栄養バランスを確認する：ダイエツト食品は低カロリーであることが多いため、栄養バランスが偏りがちです。食品の栄養成分表を確認し、必要な栄養素が含まれているかを確認しましょう。
2. 食品添加物に注意する：ダイエツト食品には、甘味料や着色料、保存料などの食品添加物が含まれることがあります。可能な限り、自然な原材料で作られたものを選びましょう。
3. 食品の種類をバランスよく摂る：ダイエツト食品だけに頼らず、野菜や果物、たんぱく質を含む食品もバランスよく摂るようにしましょう。
4. 食品のカロリーを確認する：ダイエツト食品でも、食べ過ぎるとカロリーオーバーしてしまいます。食品のカロリーを確認し、適量を守るようにしましょう。
5. 食品の味や食感に注意する：ダイエツト食品は、味や食感が普通の食品と異なることがあります。自分が好きな味や食感のものを選ぶようにしましょう。

名の候補として判定され、データセット中のカテゴリ名に該当したため、サービス名として抽出された。ダイエツト食品に該当する口コミは、247 件存在し、その中から頻出語を抽出した結果「ダイエツト食品、プラン、味、スタッフ、ウエディング、前、ポイント、施術、契約、広告、商品、プロテインダイエツト、シリーズ、ダイエツト、今」の 15 語が得られた。

ユーザの問い合わせである表 2 だけを入力とした時の出力結果を表 3 に、本研究で提案する情報補完した入力を表 4、その出力結果を表 5 に示す。

本論文で提案する入力補完の前と後を比較すると、出力の内容の具体性が異なっている。表 2 の入力補完前の出力が基本的な内容であるのに対し、表 4 の入力補完後には、抽出した特徴語を含んだ、より具体的なアドバイスが出力された。

本システムが入力補完を行った 15 単語のうち、9 単語が出力に含まれている。補完情報が口コミに含まれる頻出単語であることから、「契約内容を確認する」、「広告に惑わされない」といった実際に多く起こるとされるトラブルについても教示している。入力文に関係が薄いとされる単語の「ウエディング、ポイント、施術、プロテインダイエツト、シリーズ、今」は出力には反映されなかった。

他の例で試したところ、問い合わせ内容と補完キーワード群の関連性が低いケースの場合、出力に全く影響を与えないことも確認された。そのため、サービス名に対する一般的な補完情報のみならず、ユーザの問い合わせ内容に関連する特化した補完情報を抽出することが必要になると考えられる。また、補完情報として「前」や「今」のように一般的すぎる語句が抽出されていることから、語句の抽出手法の改良も必要になると考えられる。

4 まとめ

本論文では、大規模言語モデルへの自動入力補完を行うシステムを提案した。多くの人の体験談を含む口コミデータから特徴語を抽出し、入力補完を行うことでより具体的な内容の出力が得られることを確認した。今後は、特徴語抽出の精度の向上させることを課題とし、一

表 4 入力補完後の入力

ダイエツト食品は以下の特徴を持ちます。

特徴：ダイエツト食品、プラン、味、スタッフ、ウエディング、前、ポイント、施術、契約、広告、商品、プロテインダイエツト、シリーズ、ダイエツト、今
「ダイエツト食品を選ぶときの注意点は？」

表 5 入力補完後の出力

ダイエツト食品を選ぶときの注意点は以下の通りです。

1. 栄養バランスを確認する：ダイエツト食品は低カロリーであることが多いため、栄養バランスが偏りがちです。栄養成分表を確認し、必要な栄養素が含まれているかを確認しましょう。
2. 味にこだわる：ダイエツト食品は味が悪いと続けることが難しくなります。自分が好きな味や食材を選ぶことで、食べやすく続けやすいダイエツトができます。
3. スタッフのアドバイスを聞く：ダイエツト食品を販売する店舗やサイトには、専門のスタッフがいます。自分に合ったダイエツトプランや商品選びのアドバイスを受けることで、より効果的なダイエツトができます。
4. 契約内容を確認する：ダイエツト食品を販売する店舗やサイトによって、契約内容が異なります。契約前にしっかりと内容を確認し、自分に合ったプランを選びましょう。
5. 広告に惑わされない：ダイエツト食品の広告には、効果が大きめに表現されていることがあります。広告に惑わされず、実際の口コミや評価を確認することで、より正確な情報を得ることができます。

般にも普及し始めた大規模言語モデルの利用を補助するシステムを目指す。

謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより株式会社マイスタースタジオから提供を受けた「みんなの評判口コミデータセット」を利用した。本研究の一部は、2023 年度科研費基盤研究 (C) (課題番号：21K12147) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] 関口裕一郎 望月 崇由 鈴木 智也 今井 桂子 今井 良太. Web 検索サービスにおける多義的なクエリ推薦手法. 第 2 回データ工学と情報マネジメントに関するフォーラム論文集, pp. A1-2, 2010.
- [2] 株式会社マイスタースタジオ (2022). みんなの評判口コミデータセット. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.16.1>.