

アスペクト推定における拡張データの生成と選別

深田悠晴¹ 嶋田和孝¹

¹九州工業大学 情報工学部

fukata.yusei417@mail.kyutech.jp

shimada@ai.kyutech.ac.jp

概要

本論文では、事前学習済みモデルを用いたアスペクト推定において、LLMを用いた拡張データの生成とその生成したデータの選別に取り組む。事前学習済みモデルを用いたアスペクト推定には大量のアスペクトラベル付きデータが必要となる。そこで、LLMを用いた拡張データの生成を行う。本論文の提案手法として、RAGを利用して拡張データの生成を行い、その拡張データの選別を行うことでアスペクト推定の精度向上を目指す。実験の結果から、提案手法は実際のデータが少数の場合有効であることを確認した。

1 はじめに

インターネット上では飲食店や通販サイトなどにおいてレビューの投稿が増加している。投稿されたレビューには、サービスや価格に対してなど、様々な観点（以降、アスペクト）に対して意見が述べられている。それぞれのレビュー文がどのアスペクトについて述べているのかが分かると、ユーザーや企業にとって有益である。例えば、企業はユーザーの意見を分析しやすくなり、商品のマーケティングに役立つ。また、ユーザーにとってはレビューがアスペクトごとに整理されていることで、関心のあるアスペクトについての情報を効率的に取得できるという利点がある。

しかし、大量にあるレビュー文から各レビュー文がどのアスペクトについて述べているかを人手で判断することは非常に労力がかかる。そこで、レビュー文に含まれるアスペクトを自動でラベル付けする、アスペクト推定が注目されている [1]。アスペクト推定は、アスペクトのラベルが付いたデータでモデルを学習させる手法が有効である [2, 3]。そのような手法ではアスペクト推定モデルの学習には大量のラベル付きデータが必要となる。

しかし、大量のデータを集め、ラベル付けの作業を行うことはコストがかかる。そのような作業を低減するために、少量のラベル付きデータを用いて訓練データを拡張する方法がある。データ拡張の手法として、文中の単語の置換や削除を行うもの [4] が考えられる。しかし、アスペクト推定においては、単語の置換や削除を行うことは拡張元のデータのアスペクトが変更されたり、消えてしまう可能性があると考えられる。例えば、単語の置換により「この旅館の風呂は広くて快適だった。」という文が「この旅館の部屋は広くて快適だった。」と変更された場合、「風呂」というアスペクトの要素が「部屋」という別のアスペクトに置換されて意味が大きく変わっている。

アスペクト推定におけるデータ拡張の試みとして、川寄ら [5] は大規模言語モデル (Large Language Model: LLM) を用いたデータ拡張を行っている。LLMは大量のデータで事前学習が行われた言語モデルであり、自然言語で指示を入力するプロンプトを用いることで簡単に文章生成が可能である。川寄らの実験の結果、実際のデータに対して、LLMが生成した拡張データの割合が小さいとき精度向上が見られている。しかし、拡張データの割合を増やすと精度が低下している。これは、LLMが生成した拡張データが実際のデータと内容が異なるためデータの質が低下したことが原因であると考えられる。その要因としてLLMは大量のデータで汎用的な学習を行っているが、アスペクト推定タスクに特化した学習は行っていないためであると考えられる。

そこで、本研究ではRAG [6] (Retrieval-Augmented Generation) を活用する方法を提案する。RAGとはWebやデータベースのような外部情報を検索し、その検索結果をLLMに与えることで、LLMの生成の質を高める仕組みである。具体的には、アスペクトの情報やレビュー対象の商品名、サービス名などを検索に用い、検索結果を外部情報としてLLMの生

成時に与える。この手法により、LLMの生成するデータが実際のレビューデータに似ているものになると考えられる。

一方で、LLMは多様な文章を生成するため、LLMが生成した拡張データの中にはモデルの学習に有効なものとはそうでないものが存在する可能性がある。そこで、本研究では生成した拡張データの中から、モデルの精度向上に寄与する可能性の高いデータを選別する手法の有効性を調査する。

拡張データの選別手法として、能動学習 [7] を活用する方法と、文の類似性を利用する方法を検討する。能動学習は機械学習のモデル自身が学習に有効なデータを選択し、そのデータを訓練データに加えて学習する手法である。本論文では LLM の生成した拡張データの中から学習に有効なものをアスペクト推定モデル自身が選別することにより、アスペクト推定モデルの精度向上が期待される。文の類似性を利用する手法では、拡張データと実際のデータで文の類似度が高いものを選別する。これにより、拡張データの中から実際のデータに似ているものを選別されるため、モデルの精度向上が期待される。

2 関連研究

2.1 RAG

RAG は LLM が文章生成を行う際に、Web やデータベースなどの外部情報を参照する仕組みである。LLM の文章生成の際に外部情報を与えることで、LLM がその外部情報をもとに生成するため、LLM に生成を行う際に参照させたい情報を確実に与えることが可能である。

RAG は、QA タスク [6] や、医療、法律等の専門知識を問われるタスク [8] で利用され、精度が向上している。これらの研究のように RAG で外部情報を与えると LLM はその情報を活かして生成を行う。そのため、RAG を用いて、アスペクトやレビュー対象の情報を与えることで、LLM が生成するレビュー文が実際のレビュー文に近くなると考えられる。

2.2 能動学習

能動学習とはラベルがついていないデータの中からモデル自身が学習に有効なデータを選択する。そして、そのデータに人手でラベル付けを行い、モデルの学習に用いるといった手順を行う。これらの手順によってデータの中からモデルを効率的に学習で

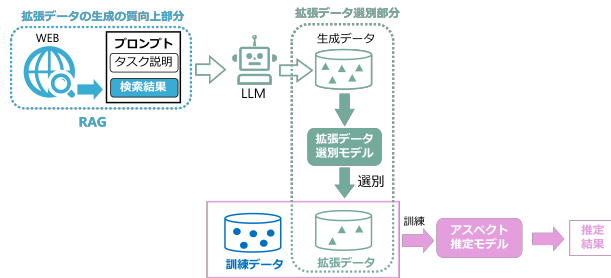


図 1: 提案手法の流れ。

きるデータを選別する。これにより、ラベル付けのコストを小さくしつつ、タスクの精度を向上させることを期待する手法である。

能動学習は様々なタスクにおいて有効である [9, 10, 11]. Manju ら [12] はアスペクト抽出タスクにおいて、強化学習と能動学習を組み合わせた手法を提案している。全データのうち 9~13% のデータのみをモデルの学習に使用し、全データを学習に使用したモデルと同等の精度を実現している。本研究では、能動学習の特性を活かし、LLM が生成する拡張データの中からモデルの精度を向上させるデータを選別することを期待する。

3 提案手法

本章では、RAG を用いた拡張データの生成とそのデータの選別手法を提案する。拡張データの生成と選別を行う流れを図 1 に示す。3.1 節では RAG を用いた拡張データの生成について述べる。3.2 節では拡張データの選別手法について述べる。

3.1 RAG を用いた拡張データの生成

RAG とは Web やデータベースのような外部情報を検索し、その検索結果を LLM に与えることで、LLM の生成の質を高める仕組みである。本研究では、アスペクト推定タスクにおける拡張データの生成において、RAG を利用する。RAG を利用することで LLM が生成する拡張データが実際のデータの質に近づくことを期待する。

RAG を用いた拡張データの生成を行う流れを図 2 に示す。外部情報として商品名やサービス名とアスペクト名で検索した情報を利用する。その検索結果を LLM に入力することで、実際のレビュー文に近い内容のデータを生成することを目指す。LLM への入力として、プロンプトを使用し拡張データを生成させる指示を与える。検索した外部情報は図 2 に示すように、プロンプトの一部に入力する。

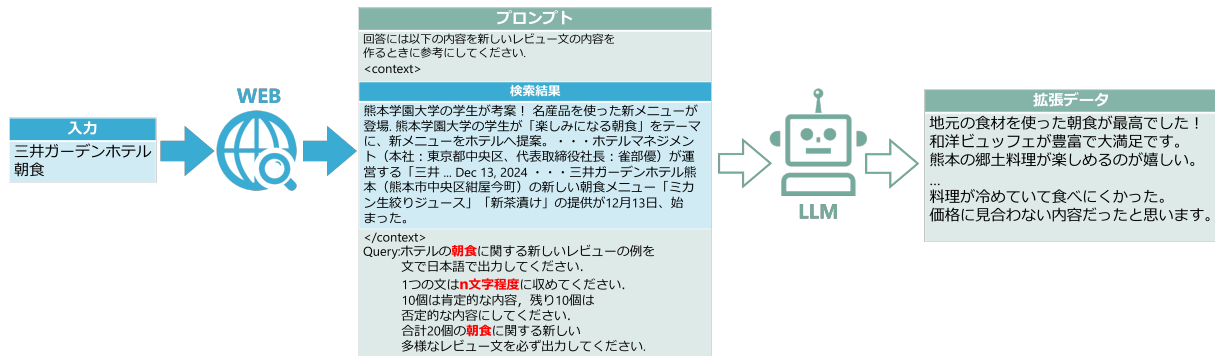


図 2: RAG を用いた拡張データの生成の流れ. ホテルの朝食について拡張データの生成を行う場合の具体例.

レビュー文の偏りをなくするため、1回の生成で肯定的なレビュー文 10 件、否定的なレビュー文 10 件の合計 20 件生成させる。文の長さにも多様性を持たせるため、図 2 のプロンプト部分の赤字で示している n 文字程度の箇所を変更し、異なる長さの文を生成させる。この調整を行う理由は LLM が同程度の長さの文ばかりを生成し、それらを拡張データとして追加することでモデルが文の長さによって推測を行うモデルになる可能性があるためである

3.2 拡張データの選別

ここでは拡張データの選別手法の説明と、各手法を本論文の実験に適用するための実験設定について説明する。本論文では、能動学習の考え方をを用いた選別手法を提案する。その手法により、モデルにとって効率的に学習できるデータを選別することで、アスペクト推定の精度向上を期待する。また、文の類似性を考慮した選別手法も提案する。その手法により、拡張データの中から実際のデータに似ているものを選別することで、アスペクト推定の精度向上を期待する。

不確実性サンプリング 能動学習の手法のうちの一つに不確実性サンプリングがある。不確実性サンプリングとは機械学習モデルが最も推測に悩んだデータが効率的に学習できるデータであるという考え方である。本論文では不確実性サンプリングの中でも、LC スコア (Least Confidence Score) を利用する。LC スコアはモデルの各ラベルの予測確率分布を p とすると、以下の (1) 式で表される。

$$LC\text{Score} = 1 - \max(p) \quad (1)$$

LC スコアはモデルが推測したラベルの予測確率が他ラベルと比べて、差が大きいほど小さく、モデルが推測したラベルの予測確率が他ラベルと拮抗し

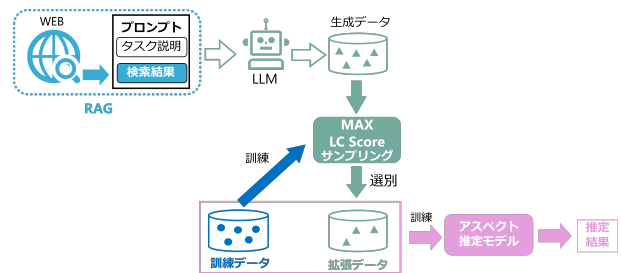


図 3: LC スコアによる不確実性サンプリングで拡張データを選別する際の流れ.

ているときは大きくなる。したがって、LC スコアが大きいデータを選ぶことにより、モデルが推測に悩んでいるデータを選ぶ。これにより、モデルが学習していないデータを追加で学習できるためモデルの精度向上が期待できる。

LC スコアによる拡張データの選別の手順を加えたものを図 3 に示す。図 3 のアスペクト推定モデルと、LC スコアによる選別モデルには、同一のモデルおよび訓練データを使用している。これにより、拡張データ追加前のアスペクト推定モデルが推測に悩んでいるデータを拡張データとして選別することでモデルの精度が向上すると考えられる。

多様性サンプリング 能動学習の手法のうちの一つに多様性サンプリング (Diversity Sampling) がある。多様性サンプリングとは、機械学習モデルに与える訓練データの多様性を確保する手法である。多様性があれば、機械学習においては重要なモデルの汎化性能の向上が期待できる。アスペクト推定においても、学習に利用したデータとは異なる表現などに対応する必要があるため、モデルの汎化性能は重要である。

今回用いた多様性サンプリングの手法を図 4 に示す。まず図 4 にあるようにテキストデータをベクトル化し、埋め込み (embedding) として表現

する。具体的には BERT [13] (Bidirectional Encoder Representations from Transformers) で embedding 化する。BERT はトークン単位で embedding 化を行う。そのため、BERT でテキスト全体の embedding を取り出す際は CLS トークンという特殊なトークンを使用することが一般的である。本研究でも、CLS トークンを用いてテキストデータを embedding として表現する。次に embedding 化したデータのクラスタリングを行う。そして、各クラスタの中心にあるものから順番にデータを選択する。クラスタリングは k-means 法を用いて行う。

k-means 法は、はじめにランダムにクラスタの重心の初期値を設定する。その重心からの距離をもとにデータをクラスタリングする。具体的には、各データは一番距離が近い重心のクラスタに分類される。次にそのクラスタリングしたデータをもとに各クラスタの重心を再計算する。そして、重心からの距離をもとにデータをクラスタリングする処理と各クラスタの重心を再計算する処理を変化がなくなるまで繰り返すという手続きにより、クラスタリングを行う手法である。k-means 法のクラスタ数 k を決める方法として、エルボー法を用いる。エルボー法ではクラスタ数 k を徐々に増やしながら、クラスタ内誤差平方和 (SSE) を計算する。SSE は以下の (2) 式で表され、各クラスタにおける中心点から各データ点までの距離の 2 乗を合計することで計算される。 k はクラスタ数、 C_i は i 番目のクラスタに含まれるデータの集合、 c_i は i 番目のクラスタの重心である。

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} (p - c_i)^2 \quad (2)$$

ほとんどの場合、クラスタ数 k を増やすと各クラスタにおける中心点から各データ点までの距離は小さくなるため、SSE は減少する。クラスタ数を増やした際に SSE の減少幅が小さくなる k を最適なクラスタ数とする手法である。

多様性サンプリングによる拡張データの選別の手順を加えたものを図 5 に示す。まず図 5 にあるようにアスペクト推定モデルの訓練データでクラスタリングを行う。そして、RAG を用いて生成したデータを、訓練データでクラスタリングされた各クラスタに分類する。そして、各クラスタの中心に近いものから順番に拡張データを選別する。

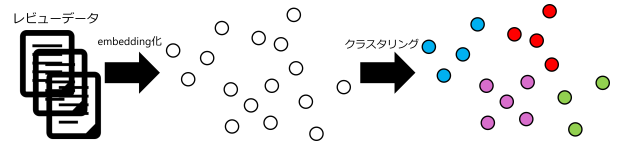


図 4: 多様性サンプリングの流れ。各レビュー文が embedding 化され、その embedding に基づいてクラスタリングが行われる。

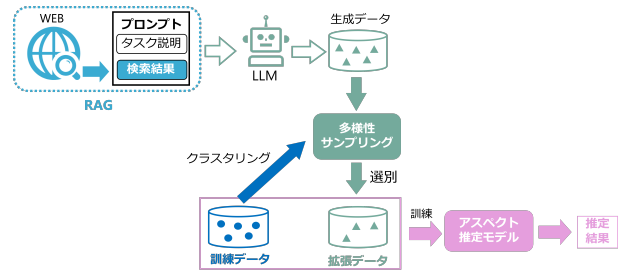


図 5: 多様性サンプリングで拡張データを選別する際の流れ。

文の類似性を利用した選別 本論文では、拡張データの選別手法として文の類似性を用いる。具体的には、拡張データの中から実際のデータと比べて文の類似度が高いものを選別する。これにより、拡張データの中から実際のデータに似ているものが選別されるため、モデルの精度向上が期待される。

文の類似性を測る方法として、BERT の CLS トークンを用いて文の embedding を取り出す手法が考えられる。しかし、CLS トークンは分類タスクを行うために最適化されているため、必ずしも文の embedding として適切かどうかはわからない。一方で、文の embedding として特化されたモデルも存在する。そこで、本論文では SBERT [14](SentenceBERT) を文の類似度測定に導入する。SBERT は文単位で embedding 化するため、文の意味を BERT より正確に測ることができる。

文の類似性を利用した拡張データを選別する手順を加えたものを図 6 に示す。まず図 6 にあるように訓練データを SBERT に入力する。そして、訓練データを文ごとにベクトル化し、平均ベクトルを求める。その平均ベクトルと拡張データの文ごとのベクトルの差を計算する。その差が小さいものを拡張データとして選別する。これにより、拡張データの中から文の内容が実際のデータに近いものを選別可能であると考えられる。

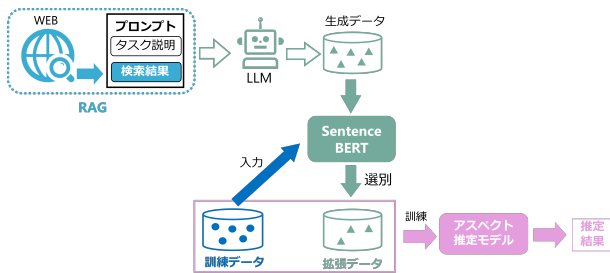


図 6: 文の類似性を利用して拡張データを選別する際の流れ。

4 実験

4.1 データセット

本論文で行う実験では、アスペクト推定を行うためのレビュー文として、楽天トラベルのアスペクト・センチメントタグ付きコーパス [15] を用いる。このデータは表 1 に示すように、ユーザーが宿泊した宿泊施設に対するレビュー文に対して、アスペクトタグをつけたものが 76624 件収録されている。アスペクトとして朝食、夕食、風呂、サービス、立地、設備・アメニティ、部屋の 7 観点がある。各アスペクトごとのデータ数は、朝食 15500 件、夕食 11449 件、風呂 10264 件、サービス 19913 件、立地 7262 件、設備・アメニティ 14704 件、部屋 11141 件である。

4.2 実験設定

本論文では、4.1 節で説明したデータセット 76624 件を訓練データ 30000 件、検証データ 1000 件、残り 45624 件をテストデータとして用いる。3.1 節で示した RAG で検索を行う単語としてホテル名とアスペクト名を用いる。ホテル名は楽天トラベルデータ [16] に含まれるホテルマスターデータを利用する。

本論文では、アスペクト推定を行うモデルとして BERT を用いる。BERT モデルは東北大学が公開しているモデル¹⁾を使用する。ハイパーパラメータは学習率は $1e-5$ 、バッチサイズは 16、エポック数は 15 とする。BERT モデルは各アスペクトごとに作成するため 7 つとなる。BERT モデルはレビュー文を入力として与え、アスペクトを含むかどうかの 2 値で出力する。モデルの評価指標は F1 値とする。

RAG を用いた拡張データの生成を行う LLM として gpt-4o-mini-2024-07-18²⁾ を用いる。gpt-4o-mini の

1) <https://huggingface.co/tohoku-nlp/bert-base-japanese-whole-word-masking>
 2) <https://platform.openai.com/docs/models/gpt-4o-mini>

表 1: アスペクトセンチメントタグ付きコーパス。

	朝食	夕食	風呂	サービス	立地	設備・アメニティ	部屋
朝食も温泉も最高!	○	×	○	×	×	×	×
スタッフの接客がひどい。	×	×	×	○	×	×	×
部屋は狭いが、アメニティは充実している。	×	×	×	×	×	○	○

RAGありのLLM	RAGなしのLLM
熊本の郷土料理が楽しめるのが嬉しい。	新鮮な魚が美味しかったです。
ミカン生絞りジュースが特に美味しかった。	朝食バイキングの種類が豊富です。
自然な味わいの新茶酒けに感激しました。	季節の野菜を使った料理が素晴らしい。
自家製プリンがとても美味しかったです。	和食の朝食が心温まる味でした。
黒豚のしゃぶしゃぶが絶品でした!	お粥がとても柔らかくて美味しいです。

(a) RAG ありの場合。 (b) RAG なしの場合。

図 7: 生成される拡張データの比較。RAG ありの場合は、外部情報に「三井ガーデンホテル 朝食」と検索した際の結果を利用した例。

ハイパーパラメータはデフォルトのものを使用する。文長の多様性を確保するため、gpt-4o-mini のプロンプトで指定する文字数を 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 と変更して、拡張データを生成する。また、実データの文字数分布に近づけるため、各文字数で生成する件数を調整する。RAG を用いて拡張データを生成させた場合と、LLM にそのまま拡張データを生成させた場合の比較を図 7 に示す。RAG を用いることで、LLM にそのまま生成させるよりも、図の赤字部分のように実際に商品やサービスを利用したような具体的な情報が増え、実際のレビュー文の内容に近づくと考えられる。

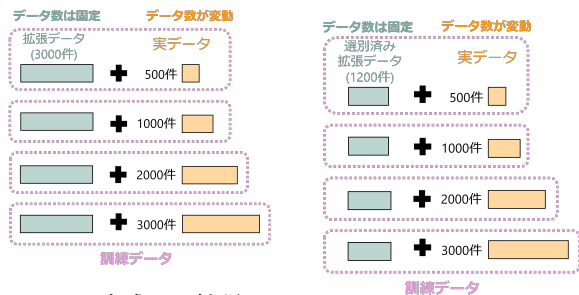
3.2 節に示した文の類似性をもとに選別する手法で用いる SBERT のモデルとして、日本語を扱える sonoisa³⁾を利用する。

一般的にデータ拡張は実データの数が増えると有効性が下がる傾向にある。そこで、本節では拡張データの数を固定し、実データの数を増やすことで、提案手法がどの程度有効か調査する。3.1 節で示した RAG を用いた LLM で生成する拡張データの数は 3000 件とする。それらの拡張データに対して 3.2 節で紹介した選別手法を適用し、1200 件選別する。実データの数による提案手法の精度に与える影響の差を調査するため、図 8 に示すように、拡張データの数を固定し、実データの数を 500 件、1000 件、2000 件、3000 件と変更した実験を行う。

4.3 実験結果

実験結果を表 2 に示す。表のそれぞれの数値は 7 つのアスペクト推定モデルの F1 値の平均である。LC スコアによる不確実性サンプリングを LC、多様

3) <https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens>



(a) RAG で生成した拡張データ (3000 件) をそのまま使う場合。 (b) 選別済みの拡張データ (1200 件) の場合。

図 8: 実データの数を考える場合。拡張データの数は固定し、実データの数のみを変更し、実データと拡張データを足したものをそれぞれアスペクト推定モデルの訓練に用いる。

性サンプリングを DS, 文の類似性を利用した選別を SS, ランダムサンプリングを RS とする。表 2 の結果から, RAG による拡張データを加え, 拡張データの選別を行わない場合, 実データの数に関わらず精度が悪化している。図 9 に示すように, LLM がプロンプトの指示に従わず, 指定のアスペクト以外のデータを生成することがあった。そのようなデータにより精度が大きく低下したと考えられる。

3.2 節で紹介した 3 つの手法のうち, 文の類似性を利用した手法は他の 2 つの手法と比べ精度が低下し, 特に実データが 500 件, 1000 件のときは, ランダムサンプリングより精度が悪かった。SBERT で拡張データと実データの文の類似性を求める際に各拡張データの embedding と実データ全体の平均の embedding を比べている。そのため, 拡張データの中から情報量の少ない文長の短いものが選別されやすかったと考えられる。その結果, BERT モデルが文長の短いものがアスペクトを含むと推測するモデルになった可能性が考えられる。

実データの数は 500 件使用し, LC スコアを用いた不確実性サンプリングによる選別を行った場合のみベースラインより精度が向上した。実データの数が少数の場合, 提案手法は LC スコアを用いて選別を行う際は有効であると考えられる。これは図 9 に示した例のような LLM がプロンプトの指示に従わないデータは LC スコアを計算する拡張データ追加前のモデルにおいても, 容易に該当アスペクトではないと予測できるためであると考えられる。そのため, LC スコアは小さくなり, 図 9 のようなデータは拡張データとして追加されないと考えられる。

表 2: 実データの数によるアスペクト推定の精度比較。精度は 7 つのアスペクト推定モデルの平均。

実データ数	ベース	データ拡張 (+3000)		データ拡張+選別 (+1200)		
		RAG	RAG + LC	RAG + DS	RAG+SS	RAG+RS
500	0.552	0.527	0.587	0.541	0.525	0.535
1000	0.610	0.572	0.599	0.581	0.550	0.570
2000	0.650	0.597	0.647	0.629	0.621	0.620
3000	0.674	0.633	0.671	0.653	0.643	0.634

館内のインテリアが優雅で癒されました。スタッフのサービスが親切で嬉しかった。温かいお布団でぐっすり眠れました。朝食後の散策が楽しみになる立地。スタッフの対応が遅くて不満でした。

図 9: 朝食を指定したが, LLM が従わない例。

5 おわりに

本論文では, アスペクト推定タスクにおいて, LLM を用いた拡張データの生成と選別を行った。RAG を用いた LLM による拡張データの生成手法と, 効率的に学習できるデータを選別できる能動学習の考え方と実際のデータに似ている文を選別できる文の類似性を利用した拡張データの選別手法を提案し, その有効性を確認した。

本論文の実験結果では, RAG を用いて生成した拡張データを訓練データに加え, 選別を行わない場合は精度が低下した。精度低下の原因は, LLM が指定されたアスペクトとは異なるデータを生成したためであると考えられる。

拡張データの選別手法として, 能動学習の中でも, 不確実性サンプリングと多様性サンプリングの考え方をそれぞれ利用した手法と文の類似性を用いた手法を提案した。実データの数が少数の際に不確実性サンプリングで選別を行う場合, 拡張データを追加する前のベースラインよりアスペクト推定の精度が向上した。そのため, 提案手法は実データの数が少数の場合, 有効であると考えられる。

しかし, 実データの数を増やすと, 提案手法はベースラインの精度を越えることができなかった。その原因は RAG を用いても, LLM の生成した拡張データが, 実際のデータの質に比べて劣るためであると考えられる。

今後の課題として, 精度低下の要因となった LLM の生成する拡張データの質を向上させる手法を模索する。また, LLM がプロンプトの指示に従わない生成を行ったこともアスペクト推定モデルの精度低下の一因となった。そのため, プロンプトの工夫などの改良により, この課題を解決する必要がある。

参考文献

- [1] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. Attention-based lstm for aspect-level sentiment classification. In **Proceedings of the 2016 conference on empirical methods in natural language processing**, pp. 606–615, 2016.
- [2] Jakub Šmíd, Pavel Priban, and Pavel Kral. LLaMA-based models for aspect-based sentiment analysis. In **Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis**, pp. 63–70, 2024.
- [3] Chi Sun, Luyao Huang, and Xipeng Qiu. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. 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. 380–385, 2019.
- [4] Jason Wei and Kai Zou. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 6382–6388, 2019.
- [5] 川崎慎之介, 嶋田和孝. アスペクト文判別のための大規模言語モデルを用いたデータ拡張の有効性. 火の国シンポジウム 2024, B-4-4, 2024.
- [6] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. **Advances in Neural Information Processing Systems**, pp. 9459–9474, 2020.
- [7] Robert (Munro) Monarch 著・上田隼也 訳・角野為耶 訳・伊藤寛祥 訳. Human-in-the-Loop 機械学習. 共立出版, 2023.
- [8] Joshua J Woo, Andrew J Yang, Reena J Olsen, Sayyida S Hasan, Danyal H Nawabi, Benedict U Nwachukwu, Riley J Williams III, and Prem N Ramkumar. Custom large language models improve accuracy: Comparing retrieval augmented generation and artificial intelligence agents to noncustom models for evidence-based medicine. **Arthroscopy: The Journal of Arthroscopic & Related Surgery**, 2024.
- [9] Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan. Multi-criteria-based active learning for named entity recognition. In **Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)**, pp. 589–596, 2004.
- [10] Yuekai Zhao, Haoran Zhang, Shuchang Zhou, and Zhihua Zhang. Active learning approaches to enhancing neural machine translation. In **Findings of the Association for Computational Linguistics: EMNLP 2020**, pp. 1796–1806, 2020.
- [11] 中島寛人, 嶋田和孝. 能動学習と局所対話への特化を考慮した対話文における言い淀み検出. 火の国シンポジウム 2024, B-4-3, 2024.
- [12] Manju Venugopalan and Deepa Gupta. A reinforced active learning approach for optimal sampling in aspect term extraction for sentiment analysis. **Expert Systems with Applications**, Vol. 209, p. 118228, 2022.
- [13] 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, 2019.
- [14] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3982–3992, 2019.
- [15] Yuki Nakayama, Koji Murakami, Gautam Kumar, Sudha Bhingardive, and Ikuko Hardaway. A large-scale Japanese dataset for aspect-based sentiment analysis. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 7014–7021, 2022.
- [16] 楽天グループ株式会社 (2020). 楽天トラベルデータ. <http://doi.org/10.32130/idr.2.2>, 2020.