〈一般研究課題〉　レビュー文で言及される低頻度の観点に対する
効率的な検出手法の開発

助 成 研 究 者　名古屋工業大学　　菊地　真人

レビュー文で言及される低頻度の観点に対する効率的な検出手法の開発

菊地　真人
（名古屋工業大学）

Development of Efficient Detection Method for Low-frequency Aspect Categories from Review Texts

Masato Kikuchi
(Nagoya Institute of Technology)

Abstract：

Aspect category detection (ACD), a task for detecting the aspect categories from texts, helps to extract the information customers need from review sentences automatically. BERT-pair was proposed, effectively solving this task by constructing auxiliary sentences from categories and classifying pairs of review and auxiliary sentences. In contrast, several product genres with multiple components include various low-frequency aspect categories in their reviews. Nevertheless, BERT-pair struggles to detect these categories; thus, the ACD performance for such products is poor. Moreover, to the best of our knowledge, no study on low-frequency category detection exists. Thus, we propose improved methods of BERT-pair to increase the performance for low-frequency categories. Our methods replace the review sentences in "false sentence pairs" that train BERT-pairs with similar sentences. This replacement enhances the diversity of sentences for low-frequency categories without additional training costs. Our experiment demonstrates that one of our methods increases the macro-F1 value for low-frequency categories by up to 0.11 points while maintaining the macro-F1 value for all categories.

1. はじめに

観点検出(Aspect Category Detection, ACD)は，各製品ジャンルに定義された観点を入力文から

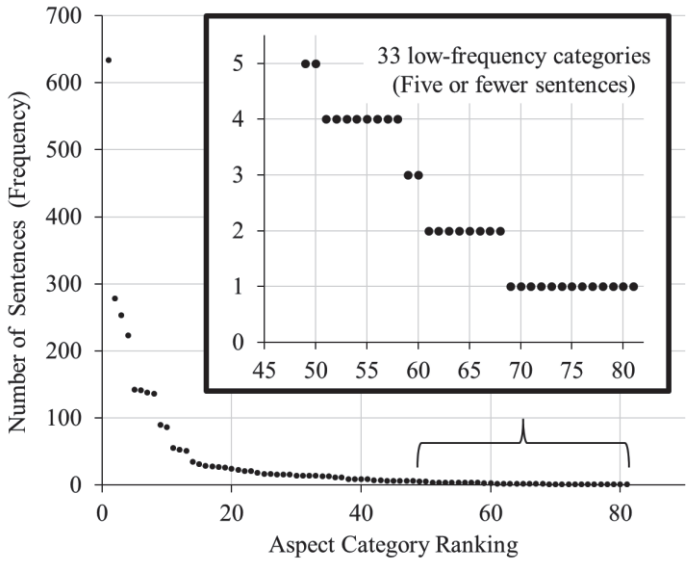

図1．ラップトップにおける観点の頻度分布．横軸は学習データセットにおける各観点を文頻度の降順で並べたランキングを表し，縦軸は各観点の文頻度を表す．

検出することを目的とする［1］．本タスクが成功すれば，製品レビューサイトに投稿された膨大な数のレビューから，使いやすさ，デザイン，配送プロセスなどの多様な観点に関する言及を効率的に抽出できる．例えば，「このノートパソコンは見た目は良いが，キーボードはあまり良くない」というレビュー文から，LAPTOP#DESIGN FEATURESとKEYBOARD#GENERALという観点を検出できる．

Bidirectional Encoder Representations from Transformers（BERT）[2]に基づいた検出手法BERT-pairモデル[3]およびその派生手法は，良好な性能を示す[4]．BERT-pairの重要な点は，元文と補助文のペアを生成することである．例えば，[“Not easy to carry,” LAPTOP#PORTABILITY]というペアでは，“Not easy to carry”が元文（レビュー文），LAPTOP#PORTABILITYが観点からなる補助文である．その後，この文ペアに対して，両文の関係が真か偽かを二値分類する．なおBERT-pairでは，製品ジャンルに含まれる全観点に対して文ペアを生成する必要がある．言い換えれば，1つの元文に対して考えうる全観点とのペアを生成する．このペア生成方法は豊富な学習例を提供するが，学習データの規模が大きくなる傾向がある．

特定の製品ジャンルでは，低頻度の観点が多くあり，これらの観点の検出が重要な課題である．例えば，ラップトップに関する既存の学習データセット[5]には2,039のレビュー文と81の観点が含まれる．図1に示すように，81の観点のうち33の観点はデータセット中に5文以下しか含まれていない．しかし我々の知る限り，既存手法の中で低頻度の観点検出に特化した手法はない．さらに，BERT-pairおよびその派生手法を用いても，学習に使用可能な文が限られていることや文ペアの多様性が乏しいことから，低頻度の観点に対する検出性能が不十分な可能性がある．

本研究では，BERT-pairの改良手法として，学習に用いる元文と補助文のペアにおいて，元文を類似文に置換する手法を提案する．これにより，学習に使用するペア数を一定に保ち，学習効率の低下を抑制できる．また，学習に用いる「負例ペア」のみを置換する．負例ペアとは，元文と補助文の関係が誤りである，つまり元文が誤った観点に対応するペアである．BERT-pairでは，こうした負例ペアと正例ペアの両方を用いて学習例を充実させている．正例ペアに対しても類似文への置換

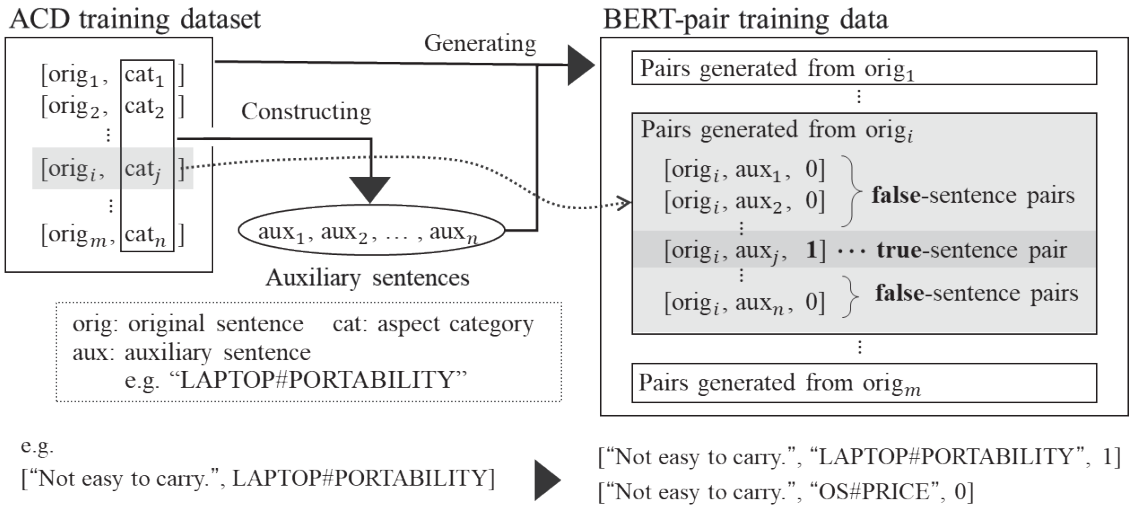

図2．BERT-pairにおける文ペア生成プロセス．origは元文，auxは補助文，catは観点を表す．

を行うことは可能であるが，正例ペアは検出対象と直接関係するため，正確性が求められる．このため，正例ペアの置換はACDの性能低下を招くリスクがある。加えて，低頻度の観点の検出では，生成される負例ペアの方が正例ペアよりもはるかに多く，かつ多様性に欠ける傾向がある．よって本研究では，負例ペアの元文を類似文に置換し，学習効率を維持しつつ，学習例の多様性を高めることを目指す．評価実験により，提案手法は全観点に対するマクロF1値をほとんど低下させず，低頻度の観点に対するマクロF1値を最大0.11ポイント向上させることを報告する．なお，本研究の成果は国際会議IEA/AIE 2024にて発表した[6].

2. ACD手法

2.1 BERT-pair

BERT-pairは，BERTベースの分類器を用い，ACDの学習データセットから生成された元文と補助文のペアを入力とすることでACDの性能を向上させる．図2は，BERT-pairの学習における文ペア生成プロセスを示している．BERT-pairを用いるにあたり，ACDの学習データセットに基づいて観点集合 $\{cat_1,cat_2,\cdots,cat_n\}$ を作成し，各観点に基づいて補助文を構築する．たとえば，データセット中の要素[“Not easy to carry”，LAPTOP#PORTABILITY]から文ペアを生成する場合，“Not easy to carry”が元文，LAPTOP#PORTABILITY が補助文となる．この2文はデータセット内で関連づけられているため，この文ペアには1が割り当てられる．図2に示すように，BERT-pairは与えられた元文に対して，全観点に対する補助文との文ペアを生成する．例えば，ラップトップドメインの学習データセットを用いると，81の観点が含まれているため，全ての元文に対して81通りの文ペアが学習されることになる．しかし，これらの文ペアの大半はデータセットに含まれておらず，関連のない組み合わせであるため，それらには0が割り当てられる．1が割り当てられた文ペアを正例ペア(図中のtrue-sentence pair)，0が割り当てられた文ペアを負例ペア(図中のfalse-sentence pair)と呼ぶ.

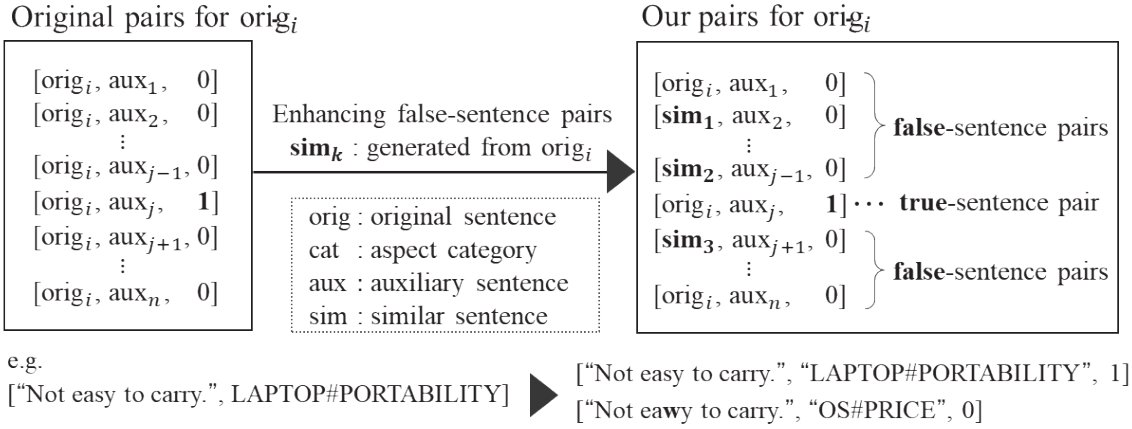

図3. 提案手法.

2.2 提案手法

負例ペアの元文を置換することで，データサイズを維持したままBERT-pairの学習データの多様性を高めることができる．これにより，特に出現頻度の低い観点における検出性能の向上が期待される．提案手法では，BERT-pairの学習データを生成する際に，図3に示すように，ラベル0のペア(すなわち負例ペア)に含まれる元文(orig)を類似文(sim)に置き換える．これは，負例ペアが正例ペアに比べて検出性能への寄与が小さいと考えられること，および低頻度の観点では負例ペアの数が正例ペアの数よりも多いという観察に基づく．$orig_i$ に由来する負例ペアに含まれる元文は，特定の割合 θ でランダムに類似文に置き換えられる．本研究では，θ の異なる値を用いて実験を行う．図中では，3つの元文がそれぞれ sim_1，sim_2，sim_3 に置換される．提案手法では，この処理をBERT-pairの学習データに含まれる全ての $\{orig_i\}_{i=1}^{m}$ に対して適用する．

BERT-pair によって生成された各ペアは，元文が各観点に属しているかを判定するために用いられる．したがって，類似文の生成方法および置換方法は，ACDの性能に大きく影響する．本研究では2つの類似文生成手法を用いる．1つはKA [7]と呼ばれる単純な手法であり，もう1つは大規模言語モデル(LLM)に基づく手法である．いずれの手法も，1つの文を入力として受け取り，そこから類似文を生成する．KAは，キーボード上の距離に基づいて元の文中の文字をランダムに置換することで類似文を生成する．LLMに基づく手法では，OpenAI APIを介してChatGPT(gpt-3.5-turbo-0613)を用い，類似文を生成する．なお，紙面の都合で詳細は省略するが，様々な類似文生成手法を比較する予備実験を行い，それらの結果よりKAとLLMを用いることにした．

3. 評価実験

3.1 実験設定

実験では，類似文生成手法および置換率 θ がACDの性能に与える影響を比較する．提案手法を信頼できる方法で評価するために，ラップトップドメインの学習データセットおよびゴールドアノテーションデータセットを用い，5つの学習データセットと1つの評価データセットを作成した．

まず，元の学習データおよび評価データから，評価データ中に5文未満しか含まれない観点に属する全ての文を除外した．これは，提案手法が低頻度の観点における検出性能を重視しているとはいえ，評価データがある程度の量に達していなければ，置換の効果を正確に評価することができな

表1: 学習データおよび評価データの情報.

	TR_0	TR_1	TR_2	TR_3	TR_4	TS
#sentences	1222	1437	1514	1531	1390	535
#categories	29	29	29	29	29	29

いためである．本実験では，この前処理済み評価データを使用した．次に，残された観点の中からランダムに5つの観点を選出し，学習データにおいて疑似的に低頻度化した．その際，選出された各観点における文頻度が学習データ中で5文となるように，対応するレビュー文をランダムに削除した．この手順によって疑似的に低頻度化した観点を含む学習データセットをTR 0と呼ぶ．同様の手順を繰り返すことで，TR 1から TR 4まで，計5つの学習データセットを作成した．そして，TR 0から TR 4をそれぞれ学習に用いて，ACDの性能を測定した．表1に，これらの学習データおよび評価データの情報を示す．“#sentences”および“#categories”は，各データセットにおける文頻度，観点の種類数を表している．TSは，上記で述べた前処理済みの評価データである．

各学習データセットを用いた評価は，Google Colaboratory上で利用可能なNVIDIA T4 GPUを用いて実施した．学習データセットのうち90%を学習用，残りの10%を検証用として使用した．BERTのアーキテクチャにはHuggingfaceのBERTForSequenceClassificationを用い，BERTに単純な線形層を追加した構成とした．エポック数は5，バッチサイズは24とした．検証データにおける損失が最小となるように学習されたモデルを評価に用いた．損失関数にはクロスエントロピー損失を用い，学習率は2e-5，最適化手法にはAdamWを採用した．

全観点に対するマイクロF1値とマクロF1値，マクロF1値がゼロではない観点数，および以下の頻度範囲ごとのマクロF1値を用いて，提案手法(KA-BERT-pairおよびLLM-BERT-pair)を評価した．学習データセットにおいて20文を超える文を含む観点を高頻度観点，6文から20文までを含むものを中頻度観点，5文以下のものを低頻度観点と定義する．最終的なマイクロF1値とマクロF1値は，TR 0からTR 4までのそれぞれの値の平均をとって算出した．生成された負例ペアにおける元文は，KAあるいはLLMを用いて類似文に置換される．置換率 θ は，各元文から生成された負例ペアのうち，25%，50%，75%，および100%を試行した．

3.2 実験結果

図4は，全観点に対するマイクロF1値およびマクロF1値を示す．マイクロF1値は両手法でほぼ同一であり，θ=75%までは約0.04のわずかな減少にとどまったが，100%では0.37まで低下した．KA-BERT-pairにおいては，マクロF1値は75%まではほとんど変化せず，100%時点で0.43から0.08ポイントの減少が見られた．表2はF1値が0より大きい観点数，すなわち検出された観点数を示す．KA-BERT-pairは，θ の増加に伴って複数の観点を検出した．その一方で，LLM-BERT-pairでは，検出された観点数は25%および50%で減少したが，75%以降で増加した．図5は，各頻度帯の観点に対するマクロF1値を示す．高頻度帯では，マイクロF1値と同様に，75%まではわずかな減少にとどまったが，100%で大幅な減少が見られた．中頻度帯では，高頻度帯と同様の傾向を示しつつも，100%時の減少幅はやや小さかった．低頻度帯では，KA-BERT-pairのマクロF1値は θ の増加にともない大きく向上し，LLM-BERT-pairでも75%以降でマクロF1値が向上した．

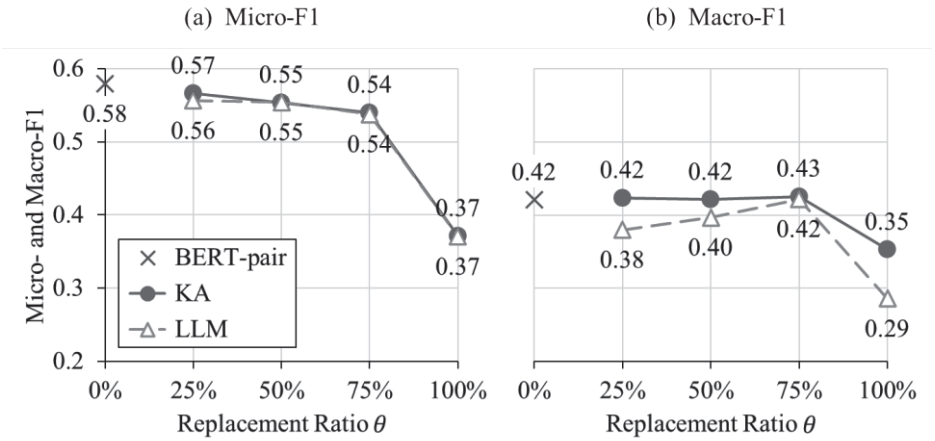

図4. 全観点に対するマイクロF1値およびマクロF1値.

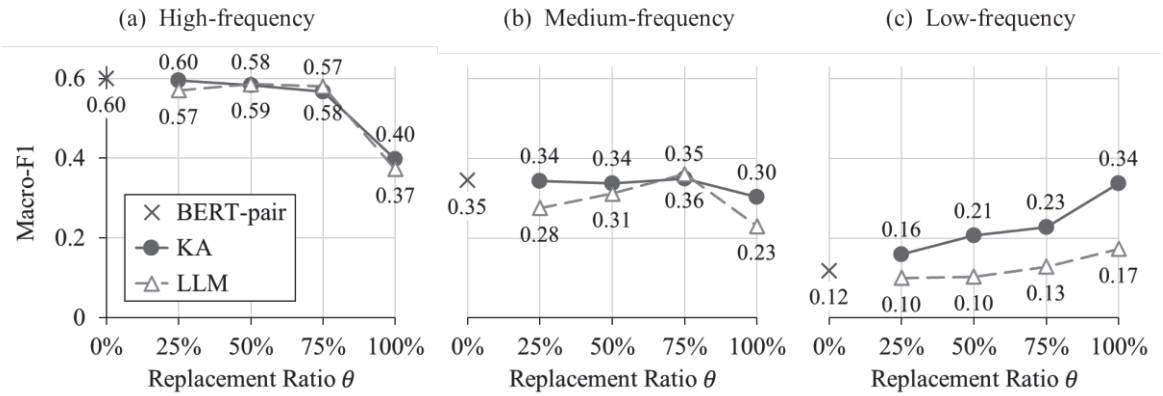

図5. 各頻度帯の観点に対するマクロF1値.

表2: 検出された観点の種類数(各学習データを用いた5回の試行の平均値).

手法	置換率 θ				
	0%	25%	50%	75%	100%
BERT-pair	22.6	---	---	---	---
KA	---	24.0	24.0	25.6	28.8
LLM	---	21.2	22.4	24.8	27.2

4. おわりに

本研究では，低頻度の観点に対する検出性能を向上させるために，BERT-pairに基づく2つのACD手法KA-BERT-pairおよびLLM-BERT-pairを提案した．提案手法の主要なアイデアは，負例ペアにおける元文を類似文に置換することである．本手法は，学習データの量を増加させず学習の効率性を維持しながら，学習に利用する文の多様性を高めるという利点を有している．評価実験において，KA-BERT-pairは，マイクロF1値，マクロF1値，および検出された観点数の全てにおいてLLM-BERT-pairを上回る性能を示した．特に θ=75% では，全観点に対するマクロF1値を維持しつつ，低頻度の観点におけるマクロF1値が0.11ポイント向上した．

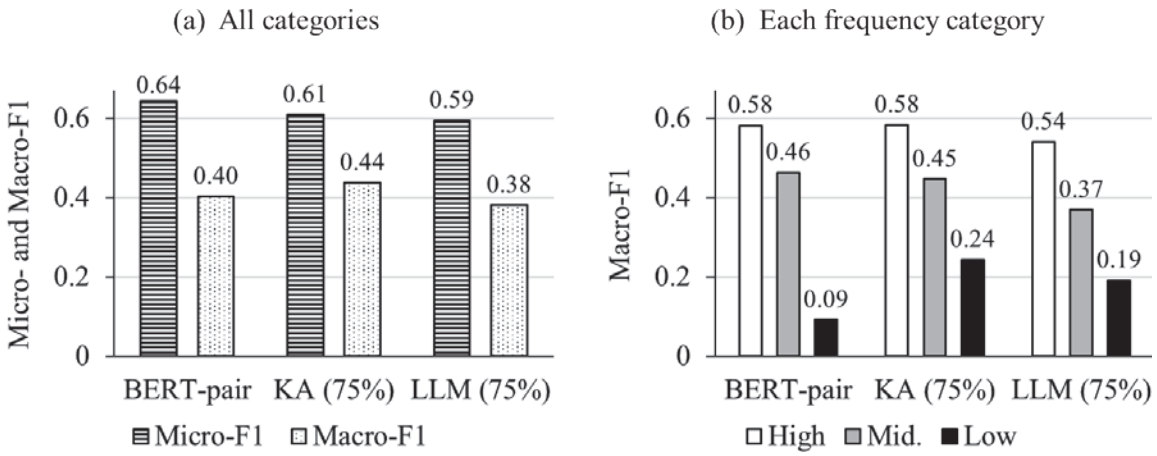

図6. ラップトップの実データセットに対するF1値.

なお紙面の都合により詳細は省略するが，ラップトップに基づく実データセットでは，θ=75%のKA-BERT-pairにより，低頻度の観点に対するマクロF1値が最大で0.15ポイント向上し(図6(b))，全観点におけるマクロF1値は最高の0.44ポイントを達成した(図6(a))．また，BERT-pairと提案手法との計算時間の差は10%未満であった．さらに，負例ペアをランダムに削除するアプローチも試し，BERT-pairと同等の検出精度を維持しながら，計算時間を削減できることを確認した．今後の課題として，負例ペアの置換と削除を組み合わせることにより，学習を効率化させながら，低頻度の観点に対する検出性能の向上を目指す．加えて，最新のLLMに対して本研究で得た知見を応用し，多様な観点からのレビュー分析システムの確立を目指す．

参考文献

[1] Pontiki, M., et al.: SemEval- 2016 Task 5: Aspect based sentiment analysis. In: Proc. SemEval-2016. pp. 19–30 (2016)

[2] Devlin, J., et al.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. NAACL-HLT 2019. pp. 4171–4186 (2019)

[3] Sun, C., Huang, L., Qiu, X.: Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proc. NAACL-HLT 2019. pp. 380–385 (2019)

[4] Chebolu, S.U.S., et al.: Survey on aspect category detection. ACM Computing Surveys 55(7, 132), 1–37 (2022)

[5] SemEval-2016 Task 5, https://alt.qcri.org/semeval2016/task5/, 2025年5月

[6] Kikuchi, M., Anda, S., Ozono, T.: Enhancing False-Sentence Pairs of BERT-pair for Low-Frequency Aspect Category Detection. In: Proc. IEA/AIE 2024. pp. 135–146 (2024)

[7] nlpaug.augmenter.char.keyboard - Read the Docs, https://nlpaug.readthedocs.io/en/latest/augmenter/char/keyboard.html, 2025年5月