

大規模言語モデルを活用した 大規模医療用語辞書メンテナンスの効率化

大槻優佳¹ 矢田竣太郎^{1,2} 西山智弘¹ 工藤紀子¹
川端京子¹ 藤牧貴子¹ 永井宥之¹ 若宮翔子¹, 荒牧英治¹
¹ 奈良先端科学技術大学院大学 ² 筑波大学

{otsuki.yuka.pb0, hiro.nagai}@naist.ac.jp yada@slis.tsukuba.ac.jp
{nishiyama.tomohiro.ns5, noriko.kudo, kyoko.kawabata, takako.hujimaki,
wakamiya, aramaki}@is.naist.jp

概要

自然言語処理において、最も早くから利用されてきたリソースが辞書である。辞書は多様なタスクに有用であるが、構築やメンテナンスに要するコストが課題である。我々は、辞書データへの人手の修正履歴を活用し、未修正の用語に対して修正を自動提案するシステムを構築することで、質を担保しつつ低コストでメンテナンスが可能になると考えた。本研究では、構築中の50万語を超える大規模医療辞書において、医療用語のメタデータを自動修正する手法を提案する。実験では、専門的な医学知識を必要とする遺伝子バイオマーカの表記ゆれの修正性能を検証した結果を報告する。本手法は、多くの辞書メンテナンス負担軽減に貢献するものである。

1 はじめに

電子カルテや病理レポートなどの医療データは、医療AIや情報抽出ツールの開発基盤となる重要なリソースである。これらのデータの統一性を確保するために辞書やオントロジーの役割が重要視されている[1, 2]。例えば、SNOMED-CT[3]は国際的な医療概念リソースとして広く利用されているが、維持管理には年間運営費6億円を超える多大なコストを要する[4]。国内においても万病辞書(J-MeDic)などの医療辞書が存在するが、人手による修正作業のコストが課題である[5]。さらに、新語の発見と追加は広く行われているが、用語の定義や関連情報の修正および充実するメンテナンスは十分に行われていないのが現状である[6]。我々は、医療AIの実現を目指す第3期戦略的イノベーション創造プログラム(SIP3)「統合型ヘルスケアシステムの構築」

にて、大規模医療辞書JMED-DICT¹⁾を構築しているが、メンテナンスの課題は同様に当てはまる[7]。

本研究では、辞書の代表的な応用先の1つである医療文書からの情報抽出に焦点を当て、大規模言語モデル(LLM)を用いた辞書の維持管理の効率化を検討する。医療文書からの情報抽出の例として、病理レポートに頻繁に出現するバイオマーカの抽出が挙げられる[8]。バイオマーカは、疾患の診断、進行状況の評価、治療反応性の予測、さらには予後の見通しを立てるための指標として臨床利用されている重要な患者情報である[9, 10]が、頭文字による略称や数字記号の組み合わせで表現されることが多く、表記ゆれが多い[11, 12]。また、この略称は、一部の有名なものを除いて医療者なら誰でも知っているという常識レベルではなく、通常は専門としている領域について必要なものを把握しているという高度な専門知識であり、LLMの医療応用の格好のベンチマークになると考えている。例えば、乳がん治療に関するバイオマーカ「HER2」の表記には、「HER-2」や、医学的には同義である「c-erbB-2」「SK-BR-3」などが用いられる。表記ゆれの解消方法には様々なアプローチがあるが、その中でも、表記ゆれを修正し、正しい参照先が明示された辞書を作成して利用する方法は、最も単純かつ分かりやすい[1, 13, 14]。辞書を活用することで、同義語や類似表現を効率的に統一できるだけでなく、参照先が明示されているため解釈の一貫性も確保できる[15]。

本研究では、このような辞書活用の重要性を踏まえ、自己改善型辞書管理システム **Sustainable Dictionary Growing System (SDGs)** (図1)を提案する。SDGsは、Human-in-the-Loopのアプローチを取

1) <https://sip3-d2.naist.jp/>

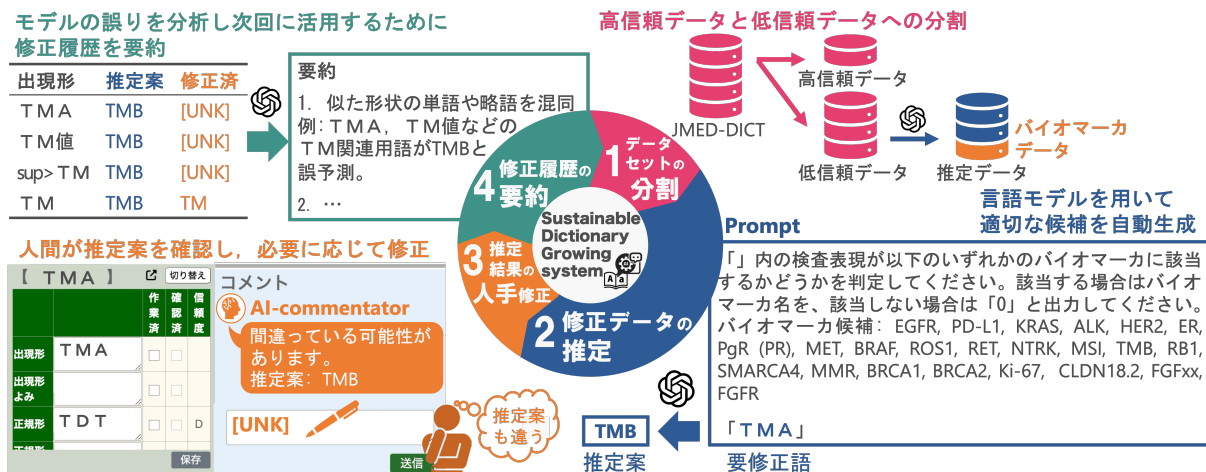


図 1: SDGs サイクルの流れ（バイオマーカーの例）。右上：高信頼データと低信頼データへの分割。右下：修正データ推定の生成プロセス。左下：推定案の人手修正インターフェース。左上：修正履歴の要約。

り入れた自己改善型の辞書管理システムであり、生成モデルを活用して辞書項目を推定し、人手による確認と修正を組み合わせることで半自動的な辞書メンテナンスを目指す。これによりコスト削減と修正精度の向上を両立し、完全自動や完全手動のメンテナンスの課題を克服する実用的な解決策となりうる。完全自動のメンテナンスでは、生成モデルの誤りや不適切な推定結果がそのまま反映されるリスクがあり [16], 完全手動のメンテナンスでは、時間と労力の負担が大きい [5]。半自動化はこれら両方の欠点を補い、効率的かつ精度の高い辞書メンテナンスを可能にする。

SDGs のサイクルは、図 1 に示すように、ステップ 1（データセットの分割）、ステップ 2（修正データの推定）、ステップ 3（推定結果の人手修正）、ステップ 4（修正履歴の要約）で構成される。特に、ステップ 4 は、膨大な文脈情報を扱う必要があり、従来の言語モデルでは難しかった。LLM の進化により、ステップ 4 が可能となったことで、これまで実現が困難だった辞書管理のサイクルを完成させることができる。本研究では、構築中の JMEDI-DICT を対象に、バイオマーカー関連の出現形を標準的な表現である正規形へ修正することで、提案手法の有効性を検証した。

表 1: 検査表現辞書の一部

出現形	出現形よみ	正規形	出現形 EN
血圧	けつあつ	血圧	blood pressure
心拍数	しんぱくすう	心拍数	pulse rate
TMA	ていーえむえー	TDT	TMA
TM値	ていーえむち	PT 活性	TM value

2 提案手法

2.1 実験材料

大規模医療辞書 JMEDI-DICT²⁾に含まれる検査表現辞書（表 1）を用いる。この辞書には、45,623 語の検査表現（出現形）が収載されており、代表的な臨床検査項目名（正規形と呼ぶ）に紐付け可能である [7]。辞書内の各種情報には作業状況を示すフラグが付与されている。

2.2 データセットの分割

推定対象を特定するためにデータセットを分割する。推定する項目の作業フラグ（例えば、正規形を修正する場合、正規形_flag）に基づき、データセットを人手確認済の「高信頼データ」、それ以外を「低信頼データ」に分割する。低信頼データに修正データの推定と人手修正を適用する。

2.3 修正データの推定

本ステップでは、言語モデルを活用して効率的に低信頼データの推定案を 1 つずつ生成する。具体的には、OpenAI API 経由で、生成モデルの GPT-4o mini³⁾をゼロショット設定で使用し、修正データを推定する。次に、生成された推定案の中から人手修正が必要となる用語を絞り込む。これにより、無関係または冗長な候補の提示を抑制し、後の人手修正の作業効率を向上できる。

2) <https://sip3-d2.naist.jp/data.html>

3) <https://platform.openai.com/docs/models/gpt-4o-mini>

2.4 推定結果の人手修正

言語モデルが生成した推定案には誤りが含まれる場合があるため、人手で確認し修正することでデータの一貫性と正確性を確保する。なお、適切な正規形が不明な場合は「[UNK]」とする。これにより、修正の質を担保するとともに、次のサイクルに向けたデータを蓄積する。

2.5 修正履歴の要約

人手修正履歴を分析することで、ステップ2のモデル改善とステップ3の修正作業のための知見を得るとともに、人手作業を効率化するための指針を提供する。OpenAI API が提供する GPT-4o⁴⁾を用いて修正履歴を要約し、誤りパターンと人手修正のポイントを抽出する。タスクに必要な情報4点（1. 出現形と正規形の定義、2. タスクの説明、3. 誤答例、4. 現行プロンプトとその改善点）を含めたプロンプトを用いる。これにより、次のサイクルに向けたモデル改善案や修正作業時の注意点が得られる。

3 実験と結果

3.1 設定と評価指標

バイオマーカとして、近年のがん患者の病理レポートや大学・製薬企業の研究者からの意見を参考にし、臨床現場や研究開発の双方にとって有用性の高いものを選定した。

推定性能検証のため、今回は全データを低信頼データとして扱った。プロンプトには、タスクに必要な情報3点（バイオマーカ候補一覧、非該当時の対応、出力形式）を含めた（図1）。この結果、45,623語のうち対象バイオマーカと判定された540語を以降のステップにおける対象語とした。

正規形（バイオマーカ）の推定性能の評価では、ステップ3にて人手で修正された正規形を正解とし、推定された正規形と比較する。全体の正解率は、540語の出現形のうち、正しい正規形を予測した割合とした。各バイオマーカの正解率は、修正後の正規形に対し、モデルが正しく予測した割合とした。理想的なモデルは、全ての出現形に対して正しい正規形を予測するとともに、不要な正規形を一切出力しないことである。

表 2: 検査値の正規形（標準的表現）推定正解率。
Total: 全体の正解率

正規形	正解率	Correct/Num
ALK	100.0%	8/8
BRAF	94.4%	17/18
BRCA1, BRCA2	83.3%	10/12
EGFR	100.0%	79/79
ER, PgR (PR)	100.0%	71/71
FGFR	100.0%	11/11
FGFxx	100.0%	17/17
HER2	100.0%	35/35
KRAS	100.0%	41/41
Ki-67	100.0%	37/37
MET	100.0%	5/5
MMR	100.0%	3/3
MSI	100.0%	14/14
PD-L1	100.0%	18/18
RB1	100.0%	3/3
RET	100.0%	3/3
ROS1	100.0%	3/3
TMB	100.0%	7/7
[UNK]	0.0%	0/87
Others	0.0%	0/68
Total	70.7%	382/540
w/o [UNK]	84.3%	382/453

3.2 修正データの推定結果と人手修正

修正データの推定性能の結果を表2に示す。人手修正後の正規形が実験対象外のバイオマーカや検査値である語、または複数のバイオマーカに該当する語は「Others」とした。人手修正後の正規形を正解として、GPT4o-miniの推定結果と一致する語をカウントした。

各バイオマーカの正解率は多くの場合で100%に達しており、高精度であることが確認された。しかし、[UNK]と修正されたものは87語あり、その出現形には「K i -」「e G F R <」など、出現形の語として適さないものを誤判定するケースが多いことがわかる。ただし、過剰に抽出された誤りの確認は容易であるため、人手作業との連携が進めやすいと言える。このような過検出を除くと、バイオマーカの正規化タスクにおいて、GPT-4o miniは実用に耐えうる性能を示した。この結果はLLMが一定の医学知識を内包していることを示唆しており、今後、

4) <https://platform.openai.com/docs/models#gpt-4o>

バイオマーカ以外の医療用語についても検証を予定している。

3.3 修正履歴の要約

3.2 節の結果をもとに、540 の対象出現形のうち不正解だった 158 用語を対象として修正履歴を要約した。誤答パターンの整理や現在の GPT-4o mini のモデルプロンプトの改善点、作業者が修正時に留意すべき点を挙げるプロンプト (A.1) を設計した。

要約結果の例を A.2 に示す。応答の一貫性と比較可能性を確保するために、temperature を 0.5 に設定し、5 回試行した。その結果、フォーマットに若干の差異が見られたが、内容的な整合性が確認された。以下、医学的妥当性・技術的妥当性・人手作業フィードバックの有用性の観点で質的に評価する。

医学的妥当性 6 つの誤答パターン【類似表記の誤認識】【関連性のある用語の混同】【部分一致による誤認識】【略語の誤解釈】【複数候補の不適切な選択】【未知の用語の誤認識】に適切に整理されている。このうちの 3 つのパターンの例を検証する。

【類似表記の誤認識】TM が TMB と誤認識されているケースが挙げられた。JMED-DICT において、TM は、内皮細胞の細胞膜に存在しトロンビンと結合する糖タンパク質の Thrombomodulin を意味する。TMB は、Tumor Mutation Burden (腫瘍遺伝子変異量) の略で、バイオマーカの一つである。似た略語や表記が誤認識されているというモデルの分析結果は適切といえる。

【関連性のある用語の混同】プロゲステロンを PgR と誤認識しているケースが挙げられた。PgR は、Progesterone Receptor (プロゲステロン受容体) の略であり、関連用語間の誤りのため、関連性のある用語の混同というモデルの指摘は適切である。

【略語の誤解釈】METs を MET と誤認識しているケースが挙げられた。METs は、代謝当量 (Metabolic equivalents) の略であり、運動強度を表す指標である。MET は、バイオマーカの他に Medical emergency team の略としても使用される。よって、MET が複数の意味を持つ場合に誤認識しているというモデルの指摘は正しいことが確認された。

技術的妥当性 要約技術の観点で述べると、モデルによる誤答パターンの分類が適切なのか懸念がある。前述の TM を TMB と誤認識するケースと、METs を MET と誤認識するケースはどちらも略語に関する誤答パターンだが、異なる誤認識の背景 (類

似する略語と複数の意味を持つ略語の誤解釈) と分類していた。誤答パターンの分類の粒度として基準が曖昧である。改善されたプロンプト例には、類似する出現形の区別や文脈を考慮した解析、略語の誤認識防止、複数正規形への対応、[UNK] の適切な出力などの改善点が反映されている点は評価できる。

人手作業フィードバックの有用性 人手による修正作業を効率化するための注目点として挙げている内容は、作業による確認の結果、妥当であると判断された。しかし、「出現形が略語の場合、文脈に基づいて正規形を判断する」など作業者がすでに心がけていることを指摘していることが多く、事例に即した新たな知見がなかったことが懸念される。

4 おわりに

医療辞書の効率的な維持管理を目的として、Human-in-the-Loop を取り入れた自己改善型辞書管理システム Sustainable Dictionary Growing System (SDGs) を提案し、バイオマーカ関連の表記ゆれを正規化する実験を通じてその有効性を検証した。バイオマーカはアルファベットなどの記号表現で、専門の医療者でしか理解できないものも多く含まれるため、この正規化は高度な医学知識を必要とする。提案システムに採用した GPT-4o mini は高い正解率を達成し、修正履歴の要約・分析によりモデルの誤り傾向や人手修正のポイントに関する知見を得た。2 サイクル以降では、前サイクルのステップ 3 にて確認されたデータが高信頼データとして扱われ、さらにステップ 4 で生成された改善プロンプトを活用することで、前サイクルで検出できなかったバイオマーカの検出が可能となる。このように、本システムを繰り返し運用することで言語モデルの性能向上と修正作業の負担軽減が期待される。

今後の課題として、出現形よみや検査分類など他の辞書項目への適用可能性の検証が挙げられる。さらに、医療知識をモデルに組み込むことで、誤りの減少や修正精度の向上が期待される。また、手作業による修正作業時間の計測や、モデル性能向上のための評価指標の検討を行い、より実用的で高精度なシステムの実現に向けた取り組みが必要である。

本研究の成果は、医療分野における辞書メンテナンスの効率化と品質向上に貢献するだけでなく、他分野での辞書管理にも応用可能な汎用性を有している。今後もシステムの改良と応用範囲の拡大に努め、医療 AI のさらなる発展を支援していきたい。

謝辞

本研究は、「戦略的イノベーション創造プログラム (SIP)」 「統合型ヘルスケアシステムの構築」 JPJ012425, JST CREST 「リアルワールドテキスト処理の深化によるデータ駆動型探査」 JPMJCR22N1 の補助を受けて行った。

参考文献

- [1] E G Brown, L Wood, and S Wood. The medical dictionary for regulatory activities (MedDRA). **Drug Saf.**, Vol. 20, No. 2, pp. 109–117, February 1999.
- [2] Mirjana Ivanović and Zoran Budimac. An overview of ontologies and data resources in medical domains. **Expert Syst. Appl.**, Vol. 41, No. 11, pp. 5158–5166, September 2014.
- [3] Kevin P Donnelly. SNOMED-CT: The advanced terminology and coding system for eHealth. **Stud. Health Technol. Inform.**, Vol. 121, pp. 279–290, 2006.
- [4] Kimikazu Kashiwagi. Introduction of SNOMED-CT: History and future vision. **J. Inf. Process. Manag.**, Vol. 51, No. 4, pp. 243–250, 2008.
- [5] Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Shoko Wakamiya, Tomohide Iwao, and Eiji Aramaki. J-MeDic: A japanese disease name dictionary based on real clinical usage. **LREC**, May 2018.
- [6] Marcin P Joachimiak, Mark A Miller, J Harry Caufield, Ryan Ly, Nomi L Harris, Andrew Tritt, Christopher J Mungall, and Kristofer E Bouchard. The artificial intelligence ontology: LLM-assisted construction of AI concept hierarchies. **arXiv [cs.LG]**, April 2024.
- [7] 永井宥之, 西山智弘, 大槻優佳, 藤牧貴子, 川端京子, 工藤紀子, 山崎由佳, 白石暖哉, 梶原智之, 進藤裕之, 河添悦昌, 今井健, 矢田竣太郎, 若宮翔子, 荒牧英治. Jmed-dict: 大規模医療用語辞書の構築. 言語処理学会第 31 回年次大会 (NLP2025), 2025.
- [8] Jeongeun Lee, Hyun-Je Song, Eunsil Yoon, Seong-Bae Park, Sung-Hye Park, Jeong-Wook Seo, Peom Park, and Jinwook Choi. Automated extraction of biomarker information from pathology reports. **BMC Med. Inform. Decis. Mak.**, Vol. 18, No. 1, p. 29, May 2018.
- [9] Kyle Strimbu and Jorge A Tavel. What are biomarkers? **Curr. Opin. HIV AIDS**, Vol. 5, No. 6, pp. 463–466, November 2010.
- [10] Robert M Califf. Biomarker definitions and their applications. **Exp. Biol. Med. (Maywood)**, Vol. 243, No. 3, pp. 213–221, February 2018.
- [11] À Bravo, M Cases, N Queralt-Rosinach, F Sanz, and L I Furlong. A knowledge-driven approach to extract disease-related biomarkers from the literature. **Biomed Res. Int.**, Vol. 2014, No. 1, p. 253128, April 2014.
- [12] Erfan Younesi, Luca Toldo, Bernd Müller, Christoph M Friedrich, Natalia Novac, Alexander Scheer, Martin Hofmann-Apitius, and Juliane Fluck. Mining biomarker information in biomedical literature. **BMC Med. Inform. Decis. Mak.**, Vol. 12, No. 1, p. 148, December 2012.
- [13] Olivier Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. **Nucleic Acids Res.**, Vol. 32, No. Database issue, pp. D267–70, January 2004.
- [14] Aldo Gangemi, Domenico M Pisanelli, and Geri Steve. An overview of the ONIONS project: Applying ontologies to the integration of medical terminologies. **Data Knowl. Eng.**, Vol. 31, No. 2, pp. 183–220, September 1999.
- [15] C Duclos, A Burgun, J B Lamy, P Landais, J M Rodrigues, L Soualmia, and P Zweigenbaum. Medical vocabulary, terminological resources and information coding in the health domain. In **Medical Informatics, e-Health**, pp. 11–41. Springer Paris, Paris, 2014.
- [16] Zhaopeng Feng, Yan Zhang, Hao Li, Bei Wu, Jiayu Liao, Wenqiang Liu, Jun Lang, Yang Feng, Jian Wu, and Zuozhu Liu. TEaR: Improving LLM-based machine translation with systematic self-refinement. **arXiv [cs.CL]**, February 2024.
- [17] H Weiler and BH Isermann. Thrombomodulin. **Journal of thrombosis and haemostasis**, Vol. 1, No. 7, pp. 1515–1524, 2003.
- [18] Dan Sha, Zhaohui Jin, Jan Budczies, Klaus Kluck, Albrecht Stenzinger, and Frank A Sinicrope. Tumor mutational burden as a predictive biomarker in solid tumors. **Cancer discovery**, Vol. 10, No. 12, pp. 1808–1825, 2020.
- [19] Caroline H Diep, Hannah Ahrendt, and Carol A Lange. Progesterone induces progesterone receptor gene (pgr) expression via rapid activation of protein kinase pathways required for cooperative estrogen receptor alpha (er) and progesterone receptor (pr) genomic action at er/pr target genes. **Steroids**, Vol. 114, pp. 48–58, 2016.
- [20] Maurice Jetté, Ken Sidney, and G Blümchen. Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity. **Clinical cardiology**, Vol. 13, No. 8, pp. 555–565, 1990.
- [21] MERIT Study Investigators, et al. Introduction of the medical emergency team (met) system: a cluster-randomised controlled trial. **The Lancet**, Vol. 365, No. 9477, pp. 2091–2097, 2005.

A 付録

A.1 修正履歴の要約プロンプト

修正履歴の要約プロンプト

大規模言語モデルを用いて、臨床検査項目名の「出現形」に対して「正規形」を予測する言語処理タスクに取り組んでいます。

- ・ 出現形: 電子カルテ等の医療テキストから抽出されたそのままの表記としての検査項目名
- ・ 正規形: 厚生労働省が提示する医師国家試験出題基準の「主な検査項目の表記」に準拠した標準的な検査項目名

例えば、出現形「〜」に対し正規形「XX」を出力するタスクです。なお、正規形が不明な場合は「[UNK]」を出力し、複数の正規形が該当しうるのは「;」区切りで最大Nつ出力します。今回は検査項目としてバイオマーカに相当するものを対象にしています。

末尾に、モデルの誤答例を示します。「入力の実現形、出力された誤答、正しい正規形」の順となっています。「正しい正規形」は、出力された誤答を踏まえ、医療専門家（人間）が修正して得られたものです。

この誤答例を基に以下を検討してください。

1. 誤答をいくつかのパターンに整理してください
2. 以下に示す、現在のモデルプロンプトの改善点を挙げてください
3. 点2を踏まえ、改善されたプロンプト例を1つ挙げてください
4. 人手による修正作業を効率化するために、点1の誤答パターンを踏まえ、作業者がモデル出力結果について注目すべき点を挙げてください

現在使用しているプロンプト例（モデル：GPT-4o mini）

「」内の検査表現が以下のいずれかのバイオマーカに該当するかどうかを判定してください。該当する場合はバイオマーカ名を、該当しない場合は「0」と出力してください。

バイオマーカ候補: EGFR, PD-L1, KRAS, ALK, HER2, ER, PgR (PR), MET, BRAF, ROS1, RET, NTRK, MSI, TMB, RB1, SMARCA4, MMR, BRCA1, BRCA2, Ki-67, CLDN18.2, FGFxx, FGFR

誤答例

出現形「TMA」 誤答:「TMB」 正規形:「[UNK]」
出現形「TM値」 誤答:「TMB」 正規形:「[UNK]」
(後略)

図 2: 要約プロンプト

図 3: 誤答パターンの整理や推定プロンプトの改善点、人手修正時の留意点を挙げるプロンプト

A.2 修正履歴の要約結果

修正履歴の要約結果

1. 誤答パターンの整理:
 - **類似表記の誤認識**: 似たような略語や表記（例:「T M」を「TMB」と誤認識）。
 - **関連性のある用語の混同**: 関連するが異なる用語を混同（例:「プロゲステロン」を「PgR (PR)」と誤認識）。
 - **部分一致による誤認識**: 出現形の一部がバイオマーカ名と一致するための誤認識（例:「R O S」を「ROS1」と誤認識）。
 - **略語の誤解釈**: 略語が複数の意味を持つ場合の誤解釈（例:「M E T s」を「MET」と誤認識）。
 - **複数候補の不適切な選択**: 複数の可能性がある場合に誤った選択（例:「R A S」を「KRAS」と誤認識）。
 - **未知の用語の誤認識**: [UNK]が適切な場合に誤ったバイオマーカを選択（例:「TMA」を「TMB」と誤認識）。
2. 現在のモデルプロンプトの改善点:
 - 出現形が類似しているが異なる意味を持つ用語を区別する能力を向上させる。
 - 略語や部分一致の誤認識を防ぐため、文脈を考慮した解析を促す。
 - 正規形が複数ある場合の適切な出力をサポートする。
 - 未知の用語に対して[UNK]を適切に出力するようにする。
3. 改善されたプロンプト例:

「」内の検査表現が以下のいずれかのバイオマーカに該当するかどうかを文脈を考慮して判定してください。該当する場合はバイオマーカ名を、該当しない場合は「[UNK]」を出力してください。類似した表記や略語に注意し、正規形が複数ある場合は「;」で区切って最大Nつ出力してください。

バイオマーカ候補:
EGFR, PD-L1, KRAS, ALK, HER2, ER, PgR (PR), MET, BRAF, ROS1, RET, NTRK, MSI, TMB, RB1, SMARCA4, MMR, BRCA1, BRCA2, Ki-67, CLDN18.2, FGFxx, FGFR
4. 人手による修正作業を効率化するための注目点:
 - 出現形が略語の場合、他の可能性を考慮する。
 - 類似表記がある場合、文脈に基づいて正規形を判断する。
 - 出現形が複数の正規形に該当する可能性がある場合、全ての選択肢を検討する。
 - 未知の用語に対しては、無理にバイオマーカを当てはめず、適切に[UNK]を使用する。

図 4: GPT-4o による要約結果の例：誤答パターンの整理、現在のプロンプトの改善点と例、人手修正の注意点