

Yahoo!知恵袋データを活用したフレイル対策のナレッジベース構築手法

大室昌也, 繁永直希, 本間壘斗, 木川真実, 石橋武史, 浪岡保男 (東京都立産業技術大学院大学)

背景・目的

フレイル対策の重要性

- フレイルとは、健康状態と要介護状態の中間に位置する虚弱な状態を指す。[1]
- 日本は超高齢社会を迎え、社会保障給付費の増大や介護人材不足が深刻化。フレイル対策（予防・改善）による健康寿命の延伸が重要である。[1]



地域社会や自治体の課題

- フレイルに関する認知度は低く、地域社会におけるナレッジの蓄積・共有も十分とは言えない。[1]
- 自治体の健康相談は電話や紙媒体が中心で、住民にとってアクセスしづらく、日常的な健康不安が表面化しにくいという課題がある。[2]

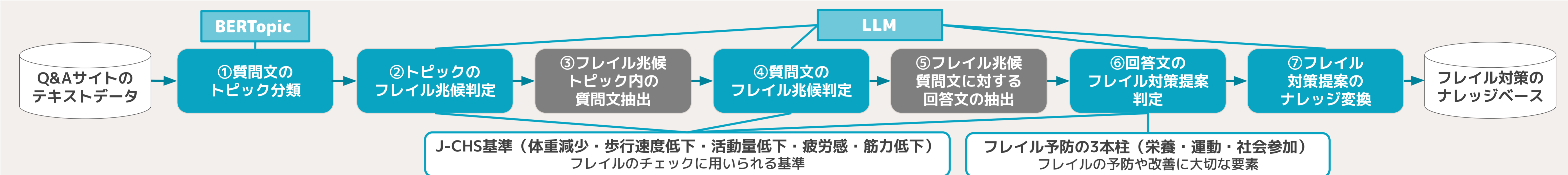


本研究の目的

- インターネット上のQ&Aサイトでは、健康相談や生活上の悩みが日常的に投稿されており、フレイル対策に役立つナレッジが潜在している可能性がある。[3]
- 本研究では、Q&Aサイトのテキストデータからフレイル対策に役立つナレッジを抽出し、地域社会や自治体の健康支援施策に活用可能なナレッジベース（ナレッジの集合体）の構築を試みる。

提案手法

本手法では、まずQ&Aサイトの質問文をBERTopicでトピック分類し、各トピックにフレイル兆候が含まれるかをLLMで判定する。次に、フレイル兆候ありと判定されたトピック内の質問文について、改めてフレイル兆候の有無をLLMで判定する。さらに、該当質問文に対する回答文を対象として、フレイル対策に有用な提案が含まれるかをLLMで判定し、その内容をナレッジとして可読性の高い文章へ変換する。なお、LLMでの判定にはJ-CHS基準[1]およびフレイル予防の3本柱[1]を活用する。



①の技術詳細) 文章埋め込み: senoia/sentence-bert-base-ja-mean-tokens-v2, 次元削減: UMAP, クラスタリング: k-means, 特徴語抽出: c-TF-IDF, ハイパーパラメータ最適化: Optuna
②以降の技術詳細) ローカルLLM: gpt-oss-20b

実験結果

本実験では、データセットとしてYahoo!知恵袋の「病気、症状」カテゴリの質問文49,092件と回答文88,898件を使用する。

BERTopic 質問文のトピック分類

→ 32トピックに分類

LLM トピックのフレイル兆候判定

→ 32トピック中8トピックがフレイル兆候あり 🧑‍🦕

- トピック 0 過食・嘔吐
- トピック 7 腰椎ヘルニアと日常・仕事への影響
- トピック 9 胸部のしこり・痛み・乳がん疑い
- トピック 10 足首・足の怪我・骨折・捻挫
- トピック 13 めまい・自律神経障害
- トピック 14 女子のめまい・立ちくらみ
- トピック 15 睡眠障害と金縛り
- トピック 28 血圧・体重・食欲・過食

LLM 質問文のフレイル兆候判定

→ 13,720件中1,438件がフレイル兆候あり 🧑‍🦕

例)

本文: 普通に歩けなくなってしまいました。2ヶ月前に諸事情により仕事を退職し引きこもりになりました。家ではほぼベッドの上で過ごすようになり、歩くのもトイレや風呂、ご飯の時のみです。そんな生活を2ヶ月程続けていたところ2週間前あたりから足の裏を床に付けて歩けなくなりました。つま先立ちをしないと歩けず、足の裏を全部つけようとする足首が思うように曲がらず後ろにバランスを崩して転倒しそうになります。(後略)

J-CHS基準の判定: 歩行速度低下・活動量低下

LLM 回答文のフレイル対策提案判定

→ 2,621件中557件がフレイル対策提案あり 🧑‍🦕

例)

本文: 理学療法士です。完全寝たきりの患者さんの関節は3日後から硬くなるという文献もある位なので、2ヶ月もほぼ寝る状態なら足首が硬くなってもおかしくないと思いますよ。(中略) アキレス腱やハムストのストレッチ、普段歩く時は踵から地面に足を付くなど意識していれば、改善すると思います。

J-CHS基準の判定: 活動量低下・筋力低下・歩行速度低下
フレイル予防の3本柱の判定: 運動

LLM フレイル対策提案のナレッジ変換

→ 557件中534件がナレッジ変換(文体統一・文脈付与など)成功 🧑‍🦕

例)

本文: 足首が硬くなっている場合は、アキレス腱とハムストリングスのストレッチを行い、歩行時はかかとから地面に足を付けるよう意識することが望ましいです。継続的に実施し、改善が見られない場合は医療機関で診察を受けることが望ましいです。

J-CHS基準の判定: 活動量低下・筋力低下・歩行速度低下
フレイル予防の3本柱の判定: 運動

評価

- ### パターン分類
- ナレッジ変換に成功した534件の回答文を15パターン(J-CHS基準×フレイル予防の3本柱)で分類した。
 - 「体重減少×栄養」・「活動量低下×運動」・「筋力低下×運動」が多く抽出されている一方、「社会参加」は総じて少なく、「歩行速度×社会参加」は0件であった。

※一つの回答文が複数の項目に該当する場合があるため、重複カウントあり

J-CHS基準\3本柱	栄養	運動	社会参加	合計
体重減少	205	54	9	268
歩行速度低下	19	79	0	98
活動量低下	103	267	22	392
疲労感	105	34	5	144
筋力低下	119	235	8	362
合計	551	669	44	1264

- ### 妥当性評価
- 各パターンからナレッジを3件ずつランダムに抽出し、5名でフレイル対策としての妥当性を評価した。
 - 平均妥当率・評価者間一致度(Fleiss' κ係数)から、本手法の有効性を確認した。
- ※「活動量低下×社会参加」は該当ナレッジが0件のため評価対象外

平均妥当率	κ係数
0.781	0.346

まとめ

- ### 主な結果
- BERTopicによるトピック分類と、J-CHS基準およびフレイル予防の3本柱を用いたLLM判定を組み合わせることで、Yahoo!知恵袋データからフレイル対策に有用なナレッジを抽出し、ナレッジベースとして体系的に構築可能であることを確認した。
- ### 今後の展望
- 手法の比較検証
 - データセットを用いず、LLM単独で同様のナレッジベースを構築し、本手法と比較検証する。
 - ナレッジの拡充
 - Yahoo!知恵袋の「病気、症状」以外のカテゴリにも対象を広げ、フレイル予防の3本柱における「社会参加」を促すナレッジを拡充する。
 - 社会実装
 - 本手法で構築したナレッジベースを活用し、LLM+RAG構成によるフレイル対策強化型の健康相談チャットボットを試作。地域社会や自治体の現場でどのように役立つか検討する。

謝辞

本研究では、国立情報学研究所の情報学研究データリポジトリ(IDR)によるデータセット提供サービスを通じて、LINEヤフー株式会社様からご提供いただいた「Yahoo!知恵袋データ(第3版)」(<https://doi.org/10.32130/idr.1.3>)を利用させていただきました。ここに記して謝意を表します。

参考文献

[1] 長寿科学振興財団: フレイル予防・対策—基礎研究から臨床、そして地域へ。長寿科学振興財団報告書。
https://www.tyvjyu.or.jp/kankoubutsu/gyoseki/pdf/R2_frailty_gyosekishu.pdf (2020).

[2] 東京大学未来ビジョン研究センター: フレイル予防を軸とした健康長寿・幸福長寿まちづくりの実現に向けた提言。
https://ifu.u-tokyo.ac.jp/wp/wp-content/uploads/2024/02/recommendation_for_society_gerontology_20240219.pdf (2024).

[3] 國枝由紀子ほか: 健康・医療相談サイトに寄せられた相談内容の分析—生活習慣関連相談を中心に—. 日本健康教育学会誌 21(3), 189-198 (2013).