

伝統医学に基づく診断支援のための ベクトル検索システムの構築

伊藤元斗[†] 関隆志^{††} 高橋晶子[†] 力武克彰[†]

本稿では、伝統医学に基づく診断を支援することを目的とした、文埋め込みを用いる伝統医学文献のベクトル検索システムについて述べる。伝統医学では、証と呼ばれる患者の心身の状態を表す概念を取り扱う。証などの伝統医学に関する情報についての解説が記載されている伝統医学文献には表記ゆれの問題があり、文献からの情報を取得し伝統医学を活用しようとする医師の負担となっている。そこで、Sentence-BERT と呼ばれる機械学習ベースの文埋め込み手法を用いて、表記揺れの問題に対処しながら伝統医学文献の検索を行い、この課題の解決を目指した。検索性能の評価では、表記ゆれの問題を解消することと、文字列一致ベースの手法である TF-IDF と同等以上の性能を示すことを確認した。また、Sentence-BERT の文埋め込みを用いる、ベクトル検索機能を搭載したアプリケーションを構築し、伝統医学に基づく診断の支援に利用することが可能となった。

Development of a Vector Search System for Diagnostic Support in the Field of Traditional Medicine

Haruto Ito[†] Takashi Seki^{††} Akiko Takahashi[†] Yoshiaki Rikitake[†]

This paper describes a vector search system for Traditional Medical literature using text embeddings, aimed at supporting diagnoses based on Traditional Medicine. In Traditional Medicine, the concept of "patterns," which represent the mental and physical state of patients, is crucial. Traditional Medical literature, which contains explanations about patterns and other Traditional Medicine-related information, faces the issue of notation variations, burdening physicians attempting to utilize this knowledge. To address this challenge, we employed Sentence-BERT, a machine learning-based text embedding technique, to conduct searches within Traditional Medical literature while resolving the notation variation issue. The evaluation of search performance confirmed the resolution of notation variations and demonstrated equivalent or superior performance compared to the string matching-based method, TF-IDF. Furthermore, we developed an application equipped with a vector search feature using Sentence-BERT embeddings, making it possible to support diagnostics based on Traditional Medicine.

1. はじめに

1.1 研究の背景

2019年にWHOが定める国際疾病分類の第11版(ICD-11)^[1]が公表され、「第26章 補助チャプター 伝統医学の病態・モジュールI」として、伝統医学に関する章が追加された。ICD-11に追加された伝統医学、および本研究で取り扱う伝統医学は、東アジア伝統医学、東洋医学、漢方医学、伝統中医学などと呼ばれる、中国、韓国、日本で扱われてきた伝統医学のことを指す。^[2]

近年、高齢化が進む中、健康寿命が重要視されていると考えられる。健康寿命の延伸のためには、生活習慣病の予防等を適切に実施する必要がある。伝統医学特有の疾病を未然に防ぐ概念や方剤などの治療法に関する知見が、生活習慣病の予防や改善に役立つとされ、伝統医学の活用に期待が高まっている。^{[3][4]}

伝統医学では「証」という概念を取り扱う。これは、発熱や吐き気、頭痛などの患者の具体的な症状に加え、精神の状態、現在の状態に至るまでの原因等、患者の心身の状態を総合的に表したものである。^[5]

証やその他の伝統医学に関する基本的な情報や応用について、よく解説された伝統医学に関する日本語の文献は複数存在するが、同じような概念や事象を表す単語が複数存在することや、文献ごとの説明内容の違い等の問題（以後、表記ゆれの問題と省略する）が存在している。

伝統医学に基づく診断においては、医師の診察によって、一つから複数の証が特定され、最終的な診断や治療法の決定に用いられる。その過程では、ある証が他の証を引き起こすといった証同士の因果関係も考慮しながら、患者の状態と証の関係を適切に結びつける必要がある。伝統医学文献の表記ゆれの問題を自力で解決しつつ、適切な証を文献から探し、以上のような診断を実施するためには、多くの診療経験と知識が必要である。これが、診療経験の少ない医師の負担となり、伝統医学を活用することに対するハードルとなっている。

2. 関連研究

2.1 伝統医学に関連する情報システム

医師による伝統医学に基づいた診断を支援するための様々な試みが行われている。伝統医学の専門医が持つ知識や経験、伝統医学文献の記述をもとに収集した情報やその関連性を適切に電子化し、それらを活用した診断の支援を目指す研究が行わ

[†] 仙台高等専門学校
National Institute of Technology, Sendai College
^{††} フジ虎ノ門整形外科病院
Fuji-Toranomon Orthopedic Hospital

れている。

松本らにより、伝統医学に基づく自動診断を行うエキスパートシステムの構築が行われている^[6]。診断から治療の自動化を行うにあたり、複雑な東洋医学理論のうちの八綱と気血水、五臓の基本理論のみを取り扱う。計 15 個の基本的な証を組み合わせ、複雑な理論体系の証や患者の状態を適切に表現し得るとの考えに基づいたシステムである。システム内では、一般症候としての 99 の症状と、主要症状としての 64 の症状、129 の処方を取り扱っている。また、各症状、各処方について、A, B, C でランク付けされた 15 の証との関連性を用いる。患者の主訴や医師の診察から得られる症状から、関連する証をルールベースで決定し、証と関連する処方を推論するという手順をシステム上で実行している。

中川らにより、患者の証や症状の関係性を表す病態図を用いた診断を支援するための、診断支援システムのアーキテクチャの提案とシステムの構築が行われた^[7]。問診情報からの証候補の抽出機能、証同士の関係を様々な手法で分析する証構造分析機能、証の関係モデル図を作成する機能からなる診断支援システムを提案している。これらの機能により、診断時に候補としてあげられる複数の証同士の関係性を機械的に抽出・提示し、医師が伝統医学に基づく診断を行う際に使用する病態図を出力することで、医師の診断を支援することを目的としている。問診票に書かれた症状と関連を持つ証を証情報 DB から検索して証候補とし、証同士の因果関係の有無が格納されたデータベースと DEMATEL 法を用いた中心度と原因度の算出、クラスタ分析、ISM 分析による証の階層分けを行い、これらの分析結果をもとに証同士の因果関係を表した関係モデル図を作成している。

また、医療機関に蓄積される電子カルテの記載や、伝統医学文献の記載そのものを活用して診断支援の実現を目指す研究もなされている。

Chuanjie Xu らによりカルテの記載から診断結果のカテゴリを予測し、診断の補助を目指す研究が実施された^[8]。伝統中医学に基づく診断のカルテを用いて、「胸部麻痺・情動不安・めまい・心悸亢進・口渇」のカテゴリに医療記録の記載を分類するデータセットが作成された、また、中国語版の Wikipedia で事前学習済みの BERT と Bi-LSTM の構造を組み合わせたモデルを構築し、分類問題のデータセットを用いた学習を実施している。評価の結果、分類データセットにおいて 73.2% の正解率と、0.7468 の平均 F1 スコアが確認されている。

太田らにより、伝統医学文献に対する情報検索で、診断の支援を行うシステムが開発された^[9]。前述のとおり、伝統医学文献とそれをもとに構築された証情報のデータベースには、表記ゆれの問題が存在しており、この表記ゆれの問題を解決して情報の提供を行う検索手法が提案されている。この手法では、Word2Vec のモデルから得た単語埋め込みを用いて、検索文字列と検索対象の文埋め込みを作成し、文埋め込み同士のコサイン類似度が高い順に並び変えた証情報を検索結果としている。

ここでの埋め込み (Embedding) とは、単語や文の意味や特徴を任意の高次元のベクトルで表現したものであり、分散表現とも呼ばれる。専門医により作成された評価データセットを用いた検索性能の評価では、表記ゆれのある検索対象に対しては、高い性能を発揮したが、それ以外の検索対象については、比較対象で文字列一致ベースの手法である TF-IDF に及ばなかった。文脈や語順を考慮した、高品質な文埋め込みを獲得する手法を採用することで、表記ゆれの問題を解決しつつ、より性能が高い検索が実現可能だと考えられる。

2.2 BERT と Sentence-BERT

BERT^[10]は、機械学習ベースの自然言語処理技術である。学習には、数 GB 以上の大規模な文書を用いた事前学習と、解きたいタスクに応じたファインチューニングの 2 段階が存在する。文章を入力として受け取り、その文の全体の特徴をあらわす埋め込みと、その文を構成する各単語(正確にはトークン)の文中での意味合いを表す埋め込みを獲得することができる手法である。それらの埋め込みを、特定のタスクを解決するための構造に入力し、さらにそのタスクに関するファインチューニングを行うことで、特定のタスクに対してある程度高い性能を持つモデルを構築することができる。

Sentence-BERT^[11]は、事前学習済みの BERT のモデルを、文の埋め込みを獲得することに特化させる手法である。Sentence-BERT では、複数の文を入力として受け取る何等かのタスクにより、事前学習済みの BERT のモデルを文の埋め込みの取得に特化させる。Sentence-BERT で得られる文の埋め込みは、コサイン類似度で類似度の計算をすることが可能な特徴を持っており、本研究では、この文埋め込みを、文脈と語順を考慮した検索タスクに向く文埋め込みとして取り扱う。

3. 研究の目的

本研究では、医師による伝統医学に基づく診断の支援を目的とし、伝統医学文献に対する情報検索システムを構築する。

西洋医学で用いられる語彙や、医師が知りえる伝統医学特有の語彙を用いて検索を行い、それに対して関連する証や伝統医学用語に関する情報を表記ゆれの問題に対処しつつ提供する手法を確立し、その手法を用いた検索システムの構築を行うことで、医師の診断を支援することが期待できる。

先行研究の証の情報検索システム^[9]の検索性能の評価を踏まえ、より高品質な文埋め込みを獲得可能とする手法を提案し、伝統医学文献の表記ゆれの問題を解消しつつ、文字列一致ベースの手法の TF-IDF と同等の検索性能を持つベクトル検索システムの構築を目指す。

4. 伝統医学文献に対するベクトル検索

本研究は、検索性能が高く、表記ゆれの問題を解消した、伝統医学文献に対する検索システムを構築することを目的としている。Sentence-BERT のモデルを用いて獲得した文埋め込みを使用することで、性能が高く表記ゆれの問題を解決した伝統医学文献に対するベクトル検索を実現できると考えられる。

Sentence-BERT のモデルを学習するにあたり、本研究では NLI データセットを使用した学習手法を採用した。また、先行研究で伝統医学の専門医により作成された、証に対する検索の評価データセットを用いて検索性能の評価と比較を実施した。

4.1 NLI データセットの拡張

NLI(Natural Language Inference)データセットは、前提文と仮説文のペアと、それらの関係について entailment・neutral・contradiction(含意・中立・矛盾)の3つのうち適切なラベルが付与された、文ペアの分類に関するデータセットである。本研究では、JSNLI データセット^[12]と呼ばれる日本語の NLI データセットを用いる。日本語の文のペアに対してラベルが付与されたデータセットである。

これに対して、伝統医学文献由来の証情報 DB の中に保存されているデータを用いて NLI データセットの拡張を行うことで、伝統医学文献特有の表現を含む文に対する検索性能を向上させることができると考えられる。

証情報 DB は以下の表 1 のような構造を持つ。この証情報 DB を用いて、以下の表 2 のようなフォーマットで NLI データセットの拡張を行い、entailment のラベルが付与されたデータを 4608 件、neutral のラベルが付与されたデータを 30478 件生成した。contradiction のラベルが付与されたデータについては、一つ一つのデータに対する専門医の判断が必要になると考えられ、また、組み合わせの件数が膨大になると予想されることから、データの拡張を実施しなかった。

表 1 証情報 DB のデータ構造 (自動生成に関連する項目のみ示す)

項目名	内容	データ型
Name	証の名前	文字列型
Symptom	証の症状の説明文	文字列型
Symptoms	証の症状のリスト	文字列型のリスト

表 2 データ拡張の組み合わせ方の例 (表証の場合)

ラベル名 (件数)	生成時の組み合わせ方	ペアの生成例
entailment (4608 件)	Name, Symptom のペア	“表証の症状”, “悪寒, または……”
	Name, Symptoms のペア	“表証の症状”, “浮脈” “表証の症状”, “咳”
neutral (30478 件)	Symptom 内の 2 つの症状の組み合わせ	“浮脈”, “咳” “咳”, “頭痛”

4.2 Sentence-BERT の学習

Sentence-BERT のモデルを作成するためには、事前学習済みの BERT のモデルを用意する必要がある。本研究では、東北大学自然言語処理グループが開発・公開している、“cl-tohoku/bert-base-japanese-v3”(以下、東北大版 BERT と略記)^[13]を事前学習済みの BERT のモデルとして使用することにした。東北大版 BERT は、主に日本語 Wikipedia の文章から作成されたコーパスと、Web で収集された日本語の文章で作成されたコーパスを用いて事前学習が行われている。

伝統医学文献特有の未知語の出現による検索性能の低下を考慮し、単漢字や熟語から構成される 257 個の語彙を東北大版 BERT に対して追加した。また、事前学習時のコーパス中に含まれている伝統医学分野の情報量と品質を考慮し、本研究では約 3.7MB の伝統医学文献から得た文章で構成されたコーパスを使用して追加の事前学習を行い、BERT の伝統医学分野への適用を試みた。既知の文書に対する高性能な検索を行うために、コーパスには証情報 DB から得た約 200KB のデータを含めている。

Sentence-BERT のモデルの学習については、NLI データセットを用いた Multiple Negatives Ranking Loss^[14]による学習を実施した。これは、NLI データセットの分ペアの分類問題を解き学習を進める手法ではなく、NLI データセットを用いて作成した 3 つの文の組 (A, B, C) を入力として、A と B の文の埋め込みの類似度を高くし、A と C の文埋め込みの類似度を小さくするような学習を実施する手法である。

通常の Multiple Negatives Ranking Loss では、文 A を基準として、文 B には文 A との関係性に entailment のラベルが付与されたものが選択される。また、文 C には文 A との関係性に contradiction のラベルが付与されたものが選択される。今回、証情報 DB を用いて拡張した NLI データセットの内容には、entailment と neutral のラベルしか含まれていない。そこで、本研究では、文 C に選ぶ対象として、contradiction と neutral のいずれかの関係を持つものを選択し、学習を実施した。

東北大版 BERT のモデルに対する追加の事前学習と、追加の事前学習を行った後の BERT のモデルを用いた Sentence-BERT の学習について、以下の表 3 に各パラメータの値を示す。

表 3 各学習のパラメータ

学習段階	学習率	バッチサイズ	エポック数	ウォームアップステップの割合
BERT	5e-5	8	200	10%
Sentence-BERT	2e-5	64	1	10%

4.3 検索性能の評価

本研究で作成した Sentence-BERT のモデルによる検索と文字列一致ベースの検索における代表的な手法の TF-IDF, 先行研究で表記ゆれの問題を解消済みの Word2Vec^[9]を用いた検索の3つ手法について, 検索性能の評価・比較を実施した。

検索評価の評価に用いるデータセット

先行研究^[9]で伝統医学の専門医により作成された, 証情報の検索性能を評価するためのデータセットを用いる。このデータセットは, 75 種類の証を検索対象としており, 「いらいら・発熱・浮腫・出血・不眠」の5つの検索文字列に対応する各証の正解と不正解のデータが記載されている。

評価指標 Recall@k

Recall@k は, 全体の正解の中で, 検索結果の上位 k の中に含まれる正解の数を示す指標である。k の数は, ユーザーが閲覧する検索結果の数に対応していると考えることができ, k が小さい場合の Recall@k では, ユーザーが最初に閲覧する検索結果のうち, どれだけ適切な検索対象が含まれているかを評価することができる。本稿では, 検索システムのユーザーが最初に確認する件数を想定し, k=5 の場合を扱う。

Recall@k は以下の式(1)で計算する。

$$Recall@k = \frac{\text{上位 } k \text{ 件に含まれる正解の数}}{\text{全体の正解の数}} \quad (1)$$

評価指標 MAP (Mean Average Precision)

MAP (Mean Average Precision)は, 検索結果全体の精度を評価する指標である。検索文字列に関連する検索対象が適切なランク付けをされているのかを評価する指標である。ある一つの検索文字列に対する検索結果のみを評価する場合, AP (Average Precision)と呼ばれる指標を式(2)で計算することが可能である。すべての検索文字列について各々APを計算し, その平均をとると MAP となる。

AP は式(2), MAP は式(3)で計算される。|Q|は検索文字列の種類の数で本稿では |Q|=5, n は検索対象のドキュメント数で n=75 である。rel(k)は上から k 番目の結果が正解なら 1, 不正解なら 0 を返す関数である。

$$AP_i = \left(\sum_{k=1}^n \frac{\text{上位 } k \text{ 件に含まれる正解数}}{k} \times rel(k) \right) \div \text{全体の正解の数} \quad (2)$$

$$MAP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} AP_i \quad (3)$$

4.4 検索性能の評価結果

表 4 に Recall@5 の計算結果を, 表 5 に各症状の AP と MAP を示す。

Recall@5 から, Sentence-BERT の“発熱”・“浮腫”・“いらいら”の検索性能は, TF-IDF と同等で, “不眠”・“出血”については, TF-IDF 以上のスコアを示していることが確認できる。また, “出血”については表記ゆれのある検索のため, Word2Vec と同様に TF-IDF のスコアを上回っており, 機械学習ベースの埋め込み獲得手法が表記ゆれの問題を解消していることが確認できる。

MAP のみに注目すると, Sentence-BERT の性能が一番優れており, それに TF-IDF, Word2Vec が続く結果となった。各 AP の傾向は, Recall@5 と似たようなものになっている。

以上のことから, 本研究で作成した Sentence-BERT のモデルを用いた証情報に対する検索は, 表記ゆれの問題を解決しつつ, TF-IDF と同等かそれ以上の性能を示すことが確認された。

表 4 各手法, 各症状の Recall@5

モデル名	発熱	浮腫	出血	いらいら	不眠
Sentence-BERT	0.17	0.5	0.5	0.31	0.31
Word2Vec	0.17	0.2	0.5	0.25	0.19
TF-IDF	0.17	0.5	0.0	0.25	0.25

表 5 各手法, 各症状の AP と, MAP

モデル名	発熱	浮腫	出血	いらいら	不眠	MAP
Sentence-BERT	0.868	0.835	0.667	0.936	0.911	0.843
Word2Vec	0.682	0.364	0.518	0.631	0.393	0.518
TF-IDF	0.803	0.817	0.044	0.750	0.591	0.601

5. 検索システムの開発

これまでに説明した Sentence-BERT による検索手法、現時点で取り扱うことが可能な伝統医学文献から作成されたデータベースを搭載した、検索システムのアプリケーションの設計と実装について述べる。

5.1 要件・技術選定・設計

本研究で開発する検索システムは、大きく分けて、Sentence-BERT による文埋め込みを用いた検索機能、検索操作が可能な伝統医学文献、GUI の 3 つの要素を持つ必要がある。

システムのユーザーは、伝統医学に基づく診断を実施する医師、もしくは伝統医学文献を検索し収録内容を調べようとする任意の医師や研究者を想定している。ユーザーが日常的に利用している機器として、PC (Windows 10/11・macOS) とスマートフォン端末 (Android・iOS) を想定される。

様々な環境に対して統一的な UI と機能を提供するための選択肢として、Web アプリや、クロスプラットフォーム対応のフレームワークを用いたネイティブアプリの構築が挙げられる。Web アプリの開発・運用には、ソースコードに加えてインフラの継続的なメンテナンスが必要になることから、本研究では、".NET 8" の "MAUI"^[15] を用いたネイティブアプリによる実装を行う。これにより、Windows, macOS, Android, iOS の各 OS の環境の差異を吸収し、ほとんど同じ GUI を持つアプリケーションを提供することが可能となる。同じく ".NET 8" の上で動作するクラスライブラリの形でその他の機能を実装することにより、各 OS のアプリ内で同じ処理を実行することができる。

.NET 8 上で Sentence-BERT を動作させるために、Microsoft.ML.Tokenizers^[16] の枠組みに沿った東北大版 BERT の日本語トークナイザの実装と Sentence-BERT のモデルファイルの PyTorch 形式から ONNX 形式への変換を行った。これにより、.NET8 上で動作する日本語トークナイザによる文章の前処理と ONNX Runtime^[17] を用いた Sentence-BERT による文の埋め込みの取得を実現することができる。

伝統医学文献については、あらかじめ取得した文埋め込みと対応づける形で SQLite^[18] のデータベースファイルに格納する。これにより、検索時の Sentence-BERT のモデルを用いた推論の回数を 1 度のみに行うことができる。この際、SQLite は文埋め込みを適切に表すために必要な実数型の配列に対応していないことから、SQLite-VSS^[19] を参考に、任意の数値型のリストと Blob (byte 型配列) を、リトルエンディアンバイト表現を基準にして相互に変換する処理を実装する。また、SQL を用いたクエリの発行によるデータベースの操作は行わず、ORM (Object Relational Mapping) を用いたデータベースの操作を行う。

検索機能については、検索文字列に対して適切と思われる各証や各用語を列挙す

る通常の検索機能の他に、検索文字列と各コンテンツの説明文の一つ一つと類似度を計算し、類似している文自体を列挙する全文検索機能を備えることとした。また、文献中の各コンテンツの各項目の名前を複数選択し、検索対象とする記述を選択する検索オプションの設定機能を備える。

5.2 システム構成

以上を踏まえ設計したシステムの構成図を以下の図 1 に示す。構成図内のすべての要素を、アプリのインストール先の端末内で完結して動作させる。

UI 部分と、機能の主体となるクラスライブラリに実装した各 API を、ViewModel と呼ばれる層で仲介する。UI の各要素は、ViewModel のプロパティを参照し、その内容を随時受け取り、表示内容とする。一部プロパティは UI 部分からの入力された内容を受け付け、値が更新される。また、ユーザーからの操作を受け付け、適切に各コマンドを参照し、実行する。各コマンド内では、適切な Service API を呼び出し、処理を実行して結果を受け取り、ViewModel 内の各プロパティを更新する。

図 1 中の "文埋め込みの生成" が Sentence-BERT のモデルの推論を実行し、文埋め込みを取得する処理に相当し、データベース操作クラスが ORM を用いたデータベース操作に相当するものである。検索対象取得コマンドの実行、および検索オプションの初期化は ViewModel の初期化時のみに実行される。

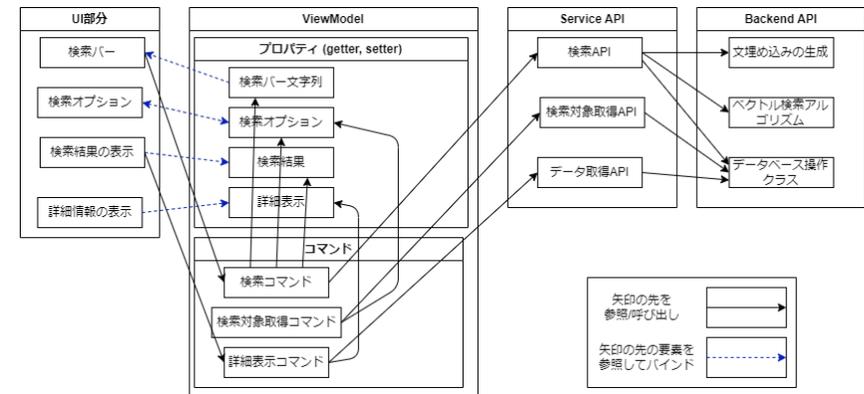


図 1 システム構成図

5.3 実装・動作確認

C#言語により、.NET 8 上で動作するクラスライブラリと MAUI を用いたアプリの実装を行った。また、表 6 に示す PC とスマートフォンにより、実装したアプリケーションの動作の確認を実施した。

アプリケーションは起動時から常に約 1.0GB のメモリ領域を占有する。CPU 推論による 1 回の検索にかかる時間は、PC で約 3 秒、スマートフォンで約 6 秒程度である。パフォーマンスのプロファイリングを行ったところ、PC では、Sentence-BERT による文埋め込みの取得で約 1.5 秒経過していることが確認された。スマートフォン上では文埋め込みの取得に約 4 秒以上の時間がかかっていることが予想される。実装した UI を図 2、各要素の説明を図 3、図 4、図 5 に示す。

表 6 動作確認を行った端末のスペック

ノート PC	
型番	FRNLKR510
OS	Windows 11 Pro (x64)
CPU	Intel Core i7-8550U (ベース動作周波数 1.8 GHz, 最大動作周波数 4.0 GHz, 4 コア 8 スレッド, TDP 15 W)
RAM	16 GB (DDR4-2400, 1200 MHz)
補助記憶装置	Crucial CT525MX300SSD
発売時期	2017 年 12 月
スマートフォン	
型番	Xperia 10 III
OS	Android 13 (aarch64)
SOC	Qualcomm Snapdragon 690 5G (CPU: 2.0 GHz+1.7GHz, 8 コア)
RAM	6 GB
発売時期	2021 年

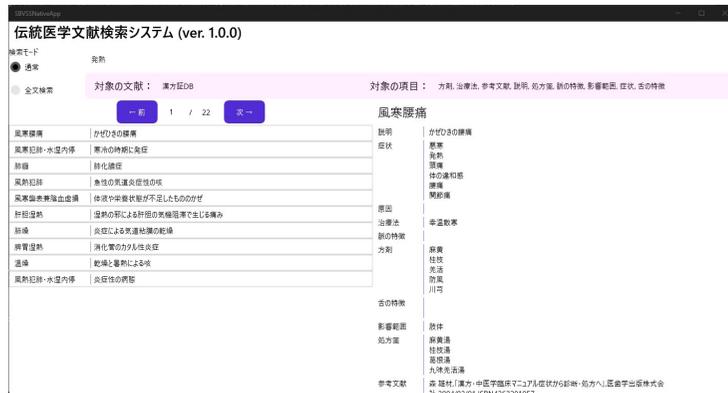


図 2 検索システムのアプリケーション (PC 版)

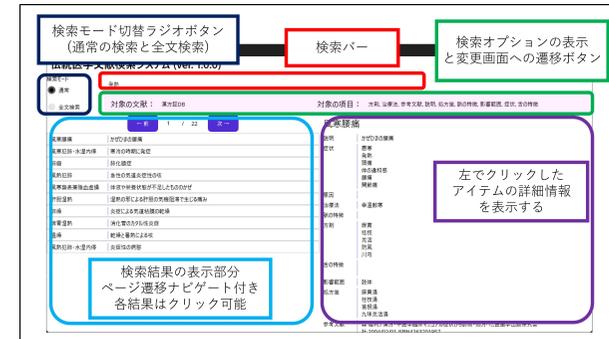


図 3 アプリ画面の各要素の説明

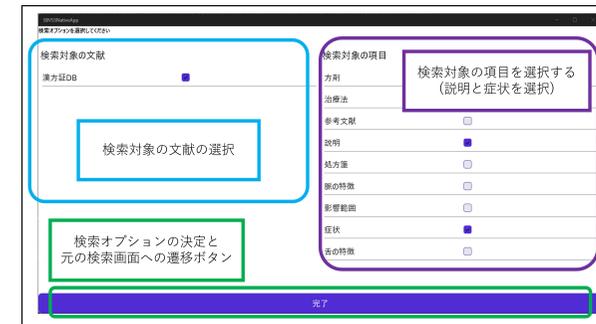


図 4 検索オプションの選択画面

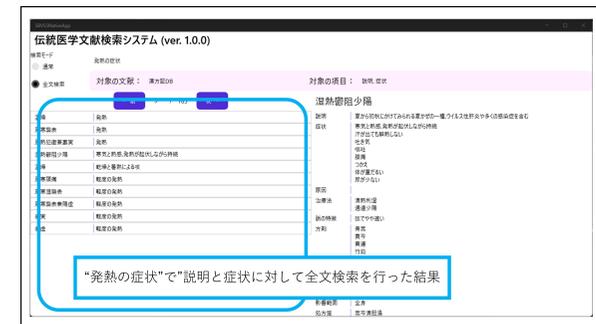


図 5 検索オプションで説明と症状を指定し、全文検索を行った結果

6. まとめ

本研究では、伝統医学に基づく診断の支援を行うことを目的として、Sentence-BERTによる検索手法の評価と、手法を組み込んだ検索システムの構築を実施した。

事前学習済みのBERTのモデルを伝統医学文献に適用させる手順を踏み、Sentence-BERTのモデルの作成を実施した。作成したモデルに関する検索性能の評価では、太田らのWord2Vecを用いた手法のように、表記ゆれの問題を解消しつつ、TF-IDFと同等かそれ以上の性能を発揮することを確認した。これにより、従来よりも高性能な伝統医学文献に対する検索が可能になると考えられる。

また、Sentence-BERTを用いた検索手法を容易に利用可能とすべく、検索システムの設計・実装を行った。検索システムは、クロスプラットフォームのネイティブアプリとして実装され、伝統医学文献自体と、伝統医学文献に対するSentence-BERTを用いた検索機能を搭載している。また、伝統医学文献のどの項目を検索対象とするのかを制御する機能も搭載され、より実用的な検索システムを構築することができた。これにより、本研究の目的である、伝統医学に基づく診断の支援が可能となった。

謝辞 本研究は JSPS 科学研究費補助金(JP23K11344)の助成を受けたものです。

参考文献

- [1] WHO: ICD-11, 入手先<<https://icd.who.int/en>> (参照 2024 年 2 月 9 日)
- [2] 和辻直, 齊藤宗則, 篠原昭二: ICD-11 伝統医学章の現状と課題, 第 34 回バイオメディカル・ファジィ・システム学会年次大会講演論文集, Vol.34, pp.2-9 (2021)
- [3] 丁宗鐵: 伝統医学における未病 -古くて新しいアプローチ-, 日本未病システム学会雑誌 Vol.6, No.1, pp.153-156 (2000)
- [4] 神奈川県: 未病の改善 - 神奈川県ホームページ, 入手先 <<https://www.pref.kanagawa.jp/docs/cz6/me-byokaizen/index.html>> (参照 2024 年 2 月 9 日)
- [5] 渡辺健治: 東洋医学における ICD-11 活用, 保健医療科学, Vol.67, No.5, pp.471-479 (2018)
- [6] 松本克彦, 西本隆, 和田茂士, 辻茂樹, 田中克己, 北村新三: 東洋医学エキスパートシステムの開発 -東洋医学の普遍化を目指して-, 日本良導絡自律神経学会雑誌, Vol.32, No.6-7, pp.132-135 (1987)
- [7] Nakagawa Ryo, Komuro Yuka, Takahashi Yui, Seki Takashi, Rikitake Yoshiaki, Takahashi Akiko: Design of a Diagnostic Support Method Utilizing Interrogation Information in Traditional Chinese Medicine, Advanced Information Networking and Applications. AINA 2021, Lecture Notes in Networks and Systems, Vol.227, pp.177-187 (2021)
- [8] Chuanjie Xu, Feng Yuan, Shouqiang Chen: Research on Assistant Diagnostic Method of TCM

Based on BERT, 2021 11th International Conference on Information Technology in Medicine and Education, pp.282-286 (2021)

- [9] 太田遥人, 関隆志, 高橋晶子, 力武克彰: 中医学のための単語埋め込みに基づく情報検索システムの研究, 情報処理学会第 84 回全国大会講演論文集, Vol.2022, No.1, pp.747-748 (2022)
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1, pp.4171-4186 (2019)
- [11] Nils Reimers, Iryna Gurevych: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, pp.3982-3992 (2019)
- [12] 吉越卓見, 河原大輔, 黒橋禎夫: 機械翻訳を用いた自然言語推論データセットの多言語化, 情報処理学会研究報告 第 244 回自然言語処理研究会, Vol.244, No.6, pp.1-8 (2020)
- [13] 東北大学自然言語処理研究グループ: cl-tohoku/bert-base-japanese-v3 - Hugging Face, 入手先<<https://huggingface.co/cl-tohoku/bert-base-japanese-v3>> (参照 2024 年 2 月 9 日)
- [14] Nils Reimers: Natural Language Inference — Sentence-Transformers documentation, 入手先 <<https://www.sbert.net/examples/training/nli/README.html>> (参照 2024 年 2 月 9 日)
- [15] Microsoft: .NET MAUI とは - .NET MAUI, 入手先 <<https://learn.microsoft.com/ja-jp/dotnet/maui/what-is-maui?view=net-maui-8.0>> (参照 2024 年 2 月 9 日)
- [16] Microsoft: Microsoft.ML.Tokenizers 名前空間, 入手先 <<https://learn.microsoft.com/ja-jp/dotnet/api/microsoft.ml.tokenizers?view=ml-dotnet-preview>> (参照 2024 年 2 月 9 日)
- [17] Microsoft: microsoft/onnxruntime: ONNX Runtime: cross-platform, high performance ML inferencing and training accelerator, 入手先<<https://github.com/microsoft/onnxruntime>> (参照 2024 年 2 月 9 日)
- [18] SQLite Consortium:SQLite Home Page, 入手先<<https://www.sqlite.org/index.html>> (参照 2024 年 2 月 9 日)
- [19] Alex Garcia, et.al: asg017/sqlite-vss: A SQLite extension for efficient vector search, based on Faiss!, 入手先<<https://github.com/asg017/sqlite-vss>> (参照 2024 年 2 月 9 日)