E-045

中医学における分散表現を用いた情報検索手法

Search method using word embedding in Traditional Chinese Medicine

太田遥人 [1]　関隆志 [2]　髙橋晶子 [3]　力武克彰 [1]

Haruto Ota　Takashi Seki　Akiko Takahashi　Yoshiaki Rikitake

1. 背景

近年,国際疾病分類第 11 版（ICD-11）において新たに伝統医学の疾病分類が追加される[1]など,中医学は補完医療としての需要が高まっている.中医学では, 診察により「証」と呼ばれる患者の状態を診断する. その過程で医師は,図 1 に示すような「病態図」を作成できれば, より正確・詳細・短時間に診断することができる. 病態図とは, 患者の問診から得られた症状と証の因果関係を図形と矢印で表現したものである.しかし,病態図の作成には証と症状を結び付けるための多くの因果関係の知識と経験が必要であり, 診療経験の少ない医師にとっては, 病態図の作成を支援する仕組みが求められている.

病態図作成の支援として, 医師の知識を補完するために症状から証を検索できるシステムを構築することが有用である. しかし, 実際に検索システムを構築する際には, 表記揺れや類義語が障害となり, すべての因果関係を網羅することは困難である. そのため, 因果関係をモデル化することが必要である.

因果関係のモデル化に関しては, 近年分散表現を用いた研究が行われている. 分散表現とは, 単語を高次元のベクトルで表現したモデルのことである. 分散表現を用いることによりモデル内に存在する全ての単語を数値によって比較することが可能となり, 証と症状の因果関係のモデルの構築が期待できる.

2. 目的

以上の背景から, 本研究では, 医師の診断を支援するために, 分散表現を用いた情報検索システムを構築することを目的とする.

以下, 証の名称, 証の概要, 証が表れる体の部位, 証に現れる症状, 治療法, 方剤例をまとめた, 診断に必要な証に関する情報のことを, 「証情報」を呼ぶ.

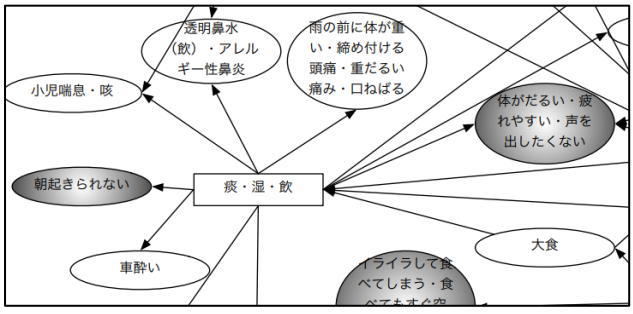

図 1. 病態図

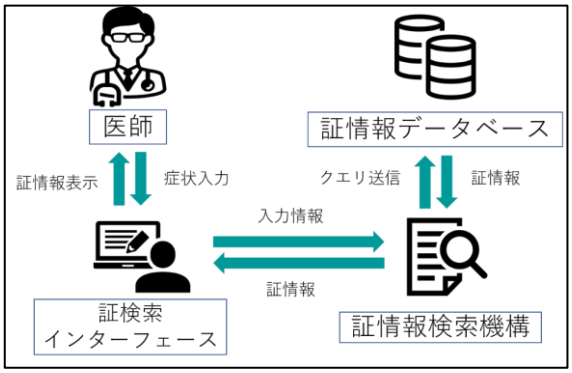

図 2. システム構成図

3. 証診断支援システム

本研究では, 分散表現を用いた情報検索手法を実現するために, 証診断支援システムを構築する. システムの動作を図 2 に示す. 本システムは, 証情報データベース・証情報検索機構・証検索インターフェースの 3 つからなる. システム構成を図 2 に示す. 各機構の役割は以下の通りである.

証情報データベース：中医学書[3]から抽出した証情報を格納する.
証情報検索機構： 分散表現を用いた情報検索手法を用いて, 入力された症状を証情報データベースから検索する.
証検索インターフェース：情報検索機構を利用するための UI となる. 医師は患者の訴える症状を証検索インターフェースに入力すると, 証検索検索機構で検索された証の名称が表示される.

ここで, 本システムの中心となる, 証を含む分散表現のモデル作成と, そのモデルを利用した検索手法の 2 つを説明する.

3.1 証の分散表現の獲得手法の概要

分散表現には, 単語を扱う単語分散表現と文書を扱う文書分散表現がある. 本研究で扱う証と症状はどちらも単語であるため, 単語分散表現を扱える Word2Vec を用いる.

Word2Vec は, テキストデータからニューラルネットワークを用いて単語をベクトルとして表現する. Word2Vec のベクトル空間では, 似た意味を持つ単語は近いベクトルとして表される.

Word2Vec の研究はさかんであり, 既に学習済のモデルが公開されている. 公開されている学習済みモデルは学習に大量のテキストデータを用いており, 多くの単語に対応で

1. 仙台高等専門学校, National Institute of Technology, Sendai College
2. フジ虎ノ門整形外科病院, Fuji Toranomon Orthopedic Hospital
3. 東北大学, Tohoku University

きることが期待できる. しかし, 語彙は一般的な単語にとどまっており, 中医学に関する単語は含まれていない.

そのため, 学習済みのモデルに対して, 中医学の単語である証をベクトル化したものを追加することでモデルの作成を行う.

モデル作成の流れは以下の通りである.

(1) 中医学文献から, ある証と, その証の症状が記述された文章を抽出する.
(2) 症状が記述された文章から, 名詞を抽出する
(3) 抽出した名詞のベクトル表現を, 学習済みのモデルから取得する.
(4) 取得したベクトル表現の和を取り, 証のベクトル表現と定義してモデルに登録する.
(5) (1)〜(4)を全ての証において繰り返し行う.

3.2 証の分散表現の獲得手法の概要

検索キーワードを症状とし, 検索対象を証とする.症状と証が意味的にどれほど類似しているかを分散表現のモデルを用いて比較する.

単語同士の類似度を測るには, 単語同士のベクトル表現のなす角がどれだけ近いかという, 「コサイン類似度」を用いる. コサイン類似度は 0〜1 で表され, 値が高いほど意味的に類似していることを示している.

検索手順は以下の通りである.

(1) ユーザーが検索キーワードである「症状」の単語を入力する
(2) 「症状」の単語ベクトルをモデル内から取得する.
(3) 「症状」の単語ベクトルと, モデル内の全ての証の単語ベクトル間のコサイン類似度を計算する.
(4) (3)で計算したコサイン類似度の高い順に, モデル内の全ての証に順位をつける
(5) 順位付けした結果をユーザーに提示する

4. 証を含む分散表現のモデル構築

モデルの構築には, Word2Vec の学習済みモデルである日本語 Wikipedia エンティティベクトル[2]を用いた. さらに, モデルの精度向上を図るためレトロフィッティング[4]を行った. この手法は,類語辞典を用いて, モデル内の類語同士のベクトルを近づけることにより精度の向上を図るものである. 類語辞典には日本語 WordNet を用いた.

追加する証として, 証情報から抽出した 75 個の証を用いた. これらの証を, 本提案手法を用いてモデルに追加した.

5. 証情報検索機構の動作実験

3 で構築したモデルと, 本検索手法を用いて症状から証の検索を行った. 入力する症状は「発熱」とした. また, 検索結果の証の症状として, 「発熱」が含まれているかどうかの比較を行った.

5.1 実験結果

実験結果を表 1 に示す. 実験結果は,本検索手法の検索結果全 75 件中, 上位 10 件を抽出したものである.

5.2 実験結果の考察

今回の実験結果には, 症状に「発熱」の記述がない証も見受けられた(例：順位 6, 風熱頭痛). この証には, 「発熱」という記述が直接記述されていなかったものの, 他の証と比べ, 分散表現において「発熱」と類似しているという結果となった. この理由としては, 症状の文章に含まれる「熱」という単語が「発熱」のコサイン類似度と近いため, 上位にヒットしたと考えられる. そのため, 必ずしも文字列に一致する証のみが有用な情報であるとは限らないことが分かった.

また, 症状に「発熱」の記述がない証の中で, 「発熱」と同義語と考えられる単語も存在していないにもかかわらず. 他の証と比べ, 分散表現において「発熱」と類似している

表 1. 実験結果

順位	証の名称	「発熱」が症状として含まれている	「発熱」に関係する単語が症状として含まれている
1	風熱犯衛	○	○(発熱・熱感)
2	水飲阻滞	×	×
3	風線犯衛兼裏実	○	○(発熱・熱感)
4	陰暑	○	○(発熱・熱感)
5	表実証	○	○(発熱)
6	風熱頭痛	×	○(熱が高い)
7	痰熱上擾	×	×
8	痰熱上擾	×	×
9	風寒襲表兼陰血虚損	○	○(発熱・熱感)
10	暑熱傷気	○	○(発熱)

という結果となる証も見受けられた. (例：順位 2,水飲阻滞) この理由としては, 症状の文章に含まれる「吐き気」「めまい」という単語がモデル内の「発熱」のコサイン類似度と近いため, 上位にヒットしたと考えられる.

この結果から, 分散表現に基づく因果関係は構築できていると考えられる. しかし, 実際の診療に用いることができるかどうかに関しては今回の実験では測定できていない. そのため, 医師に本検索システムの結果を提示し, 実際の証と症状の因果関係との比較を依頼する必要があると考えられる.

6. おわりに

本研究は, 中医学において医師の診断を支援するために, 証に関する情報の検索システムを構築することを目的とし, 分散表現を用いた情報検索手法の提案, 証診断支援システムの構築を行うものである.

今後の予定としては, 検索に対応できる証を増やすために, 証情報データベースに証を追加していくと共に, 実際の証と症状の因果関係と, 分散表現内での証と症状の因果関係の比較を行っていく必要があると考えられる. また, 今回動作を確認した証情報検索機構に加え, 証検索インターフェースの実装を行い, 証診断支援システムを構築する.

参考文献

[1] WHO, “ICD-11”, 国際疾病分類第 11 版, https://icd.who.int/en, (2021) (最終アクセス日:2021-06-15)
[2] 鈴木正敏, “日本語 Wikipedia エンティティベクトル”, http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ (2018-09-24) (最終アクセス日 2021-06-15)
[3] 森 雄材,「漢方・中医学臨床マニュアル症状から診断・処方へ」, 医歯薬出版株式会社, ISBN 4263201957(2004)
[4] Manaal Faruqui, et at. ”Retrofitting Word Vectors to Semantic Lexicons”, 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp1606-1615, (2015)