

医師国家試験自動解答プログラムの治療薬問題への拡張

Extension of a question-answering program that automatically answers the medical licensing examination to drug related questions

水口 達矢^{*1}
Tatsuya Mizuguchi

伊藤 詩乃^{*1}
Shino Ito

佐藤 健吾^{*1}
Kengo Sato

榊原 康文^{*1}
Yasubumi Sakakibara

^{*1}慶應義塾大学
Keio University

Medical diagnostic support system is an automatic support system that prevents doctors from unknowingly mis-interpret medical results. Our final goal is to construct such a medical diagnostic support system that could improve the overall accuracy of medical diagnoses. As a pilot study, we built a program that automatically answers the medical licensing examination, in our previous study.

In the current research, we attempt to expand this program to answer drug related questions and improve the accuracy by vectorizing words.

1. はじめに

近年、医療データベースやレセプトデータ、電子カルテなどに代表される医療データの電子化が進んできた。当研究室では、この電子カルテの利用に着目し、電子カルテに記載された患者の情報を適切に読み取り、診断することのできる診断支援システムの開発を目指した。そのパイロット研究として、電子カルテと形式が類似している医師国家試験における臨床実地問題を、コンピュータにより自動解答するシステムの構築がされた [伊藤 14]。医師国家試験における臨床実地問題とは、問題文において患者の様々な情報（年齢・性別・症状等）や、検査結果が与えられ、それらの情報をもとに、選択肢の中から最も相応しいものを選択するという問題である。

先行研究 [伊藤 14] では臨床実地問題のうち、病名を診断する問題を自動解答するシステムが構築され、診断支援システムの土台を築いた。しかし、医師国家試験には病名を診断する問題以外にも、処方すべき薬品や、次に行うべき検査を選ぶ問題などが存在する。病名を診断だけでなく、処方薬を決定する能力は支援システムにとって重要である。したがって本研究では、既存のシステムを拡張し、処方薬を選ぶ問題を解答できるシステムの構築を目指す。また、本研究では、「ルールベース構築の自動化」と「単語のベクトル化」を行うことでシステムの改良を図った「ルールベースの構築の自動化」により、事前に薬品の情報をシステム内に入力しておく必要がなくなり、「単語のベクトル化」により、表記揺れや異表記の影響を回避できることが期待される。

2. 方法

2.1 解答対象

解答対象として第 80 回から第 109 回までの医師国家試験を用いた。医師国家試験とは、医師免許を取得するための国家試験である。この問題セットのうち、本研究では以下の問題を解答した。

1. 問題文において、患者の症状・検査結果などが示され、選択肢の中から適切な処方薬を解答する問題

2. 解答である薬品が、複数の問題中に 2 回以上出現するもの

このような条件に当てはまる問題は、64 問存在し、そのすべての問題に解答した。

2.2 プログラムの構成

問題解答プログラムの概要を以下に示す。図 1 に示した通り、問題解答システムは、問題文を読み込み数値や重要な単語（キーワード）を抽出する部分、選択肢からルールベースを参照する部分、word2vec[Mikolov 14] により単語をベクトル化する部分、スコア計算を行い解答を出力する部分、学習を行う部分の 5 つに分かれている。

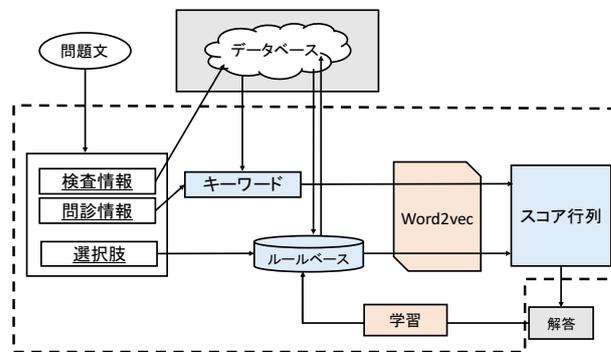


図 1: 問題解答システムの構成

2.3 キーワードの抽出

問題文を形態素解析器 MeCab[工藤 04] を用いて単語（形態素）に分割し、名詞または動詞である形態素の原型を、キーワードというリストに格納した。検査結果のような数値情報は、基準値を予め取得しておき、これと照らし合わせることで異常判定を行ったのち、キーワードに変換した。例えば、「尿酸 8.2mg/dl」と問題文に記載があると、「尿酸: 8.0」というように構造化して抽出する。この場合は基準値の 7.0 mg/dl を上回っているため「痛風」や「高尿酸血症」のような疑われる症状や病名がキーワードとして格納される。

2.4 ルールベースの作成

先行研究において、病名を判定する際のルールは、あらかじめマニュアルで入力されていた。そのため、システムに含まれて

連絡先: 慶應義塾大学理工学部生命情報学科

〒 223-8522 横浜市港北区日吉 3-14-1

E-mail: yasu@bio.keio.ac.jp

ない病名が選択肢に出現した際に解答することができない、という問題点が存在した。そこで本研究では、選択肢に薬品名が与えられると、インターネット上からその薬品の情報を自動取得し、薬品のルールとしてシステムに格納した。これにより、システムに含まれてない薬品が選択肢に出現した場合にも対応できるようになった。今回は『LIFE SCIENCE DICTIONARY』[金子 93]と『おくすり 110 番』[おくすり 110 番]の2つのサイトから薬品の情報を取得した。取得された情報は、形態素に分割し、名詞と動詞の形態素の原型を薬品のルールとして抽出した。この際、ルールの単語には、2.7 で説明する機械学習のために、それぞれ取得時に重み $w=1$ を設けた。例えば、薬品の情報として「うつ病を改善する薬」という文が得られると、ルールとして「 $w_{うつ病}=1$ 」「 $w_{改善}=1$ 」「 $w_{薬}=1$ 」が取得されることになる。ここで、薬品名とルール、および重みを1つの構造体に格納したものをルールベースと呼ぶ。

2.5 スコア計算

2.3 で問題文から抽出されたキーワードと、2.4 で取得した各薬品のルールを比較することにより、薬品ごとにスコアを決定した。先行研究ではキーワードとルールで共通して存在する単語の数を病名のスコアとしていた。しかしこの手法は単語が完全に一致した数をスコアに加算するため、表記揺れや同義語、似た意味を持つ単語動詞の関係を考慮することができないという問題点があった。そのため本研究では、word2vec[Mikolov 14]と呼ばれるニューラルネットワークモデルを使用してキーワードとルールの単語集合をすべてベクトル化し、その集合間の距離からスコアを計算した。word2vec で作成した分散表現には、似た意味を持つ単語ベクトル同士ほど、コサイン類似度が高くなるという傾向があるため、これにより表記揺れや同義語、似た意味などの情報を踏まえたスコア計算が可能となる。

問題文のキーワードを K 、薬品のルールを R 、そのルールの数を i 、あるルール r の重みを w_r とし、スコア計算を以下の式で与えた。

$$score = \sum_{k \in K} \max_{r \in R} w_r \cos(r, k) \quad \dots (3.1)$$

キーワードごとに、要素の最大値をキーワードスコアとし、すべてのキーワードのスコアの合計を薬品のスコアとした。キーワードごとに「近い」単語を取ってくることで、表記ゆれした単語や似た意味の単語を探索している。

2.6 解答の出力

すべての選択肢にスコアが付けられると、問題文に記載されている解答すべき薬品数に応じて、スコアが高い薬品を解答として出力した。例えば、問題文に「2つ選べ」と記載されていた場合、スコアが高い薬品から2つを選択し、解答とした。

2.7 学習

システムが出力した解答と正解を比較し、不正解であった場合には、正解に近づくように2.4 で付けたルールベースの重み(初期値1)を更新した。重みを更新することで、式(3.1)において w_r 、及び薬品のスコアが更新される。これによりスコアを正解に近づけることを学習と呼ぶ。重みの更新式は、システムが誤解答した選択肢のルールの重みを w_{FP} 、解答できなかった正解の選択肢のルールの重みを w_{FN} とし、以下の式で与えた。

$$w_{FP} = w_{FP} - \dots, w_{FN} = w_{FN} + \dots (3.2)$$

はハイパーパラメータであり、本研究では $\alpha = 0.1 \sim 1.5$ の15種類を用いた。

3. 結果

解答結果を表1にまとめる。本研究では、学習を繰り返し、トレーニングデータの正答率が下がったときに更新を中止し、テストデータを解答する、という方法で正答率を求めた。検証法として Leave-One-Out 交差検証を用いた。単語をベクトル化し、コサイン類似度によってスコアを求めた本手法では、病名診断システムで使用された、キーワードとルールベースの完全一致数をスコアとする従来手法と比較して、10%以上の正答率の向上がみられた。この要因として、単語の一致・不一致だけでは抽出できなかった単語間の関係を、ベクトル化によって扱えるようになったことがあげられる。

表 1: 正答率の比較

	正答率
従来手法(学習後)	43.7 %
本手法(学習前)	39.1 %
本手法(学習後)	57.8 %

4. おわりに

医師国家試験のうち、処方薬を選択する問題を自動解答できるシステムを構築した。ルールベース構築の自動化により、短時間で、大量の情報を取得することが可能になり、システムの汎化性能が向上した。このシステムを用いて、第80回から第109回医師国家試験の処方薬問題64問について解答した結果、新たに提案した単語のベクトル化を用いた手法において、57.8%の正答率が得られ、従来手法を上回った。今後の課題としては、画像の情報を用いた医師国家試験解答システムとの統合[川村 15]や、スコア計算手法の更なる検討、データ不足の改善などがあげられる。

参考文献

- [伊藤 14] 伊藤詩乃: 医師国家試験を自動解答するプログラムの構築, 人工知能学会論文誌, Vol. 31, No. 6 (2014).
- [おくすり 110 番] <http://www.jah.ne.jp/kako/>.
- [金子 93] 金子周司, 鶴川義弘, 大武博, 河本健, 竹内浩昭, 竹腰正隆, 藤田信之: Life science dictionary, <http://lsd-project.jp/cgi-bin/lsdproj/ejlookup04.pl>.
- [川村 15] 川村健史, 伊藤詩乃, 榎原康文: 検査画像情報を取り入れた医師国家試験自動解答プログラムの構築, 第2回医用人工知能研究会 (SIG-AIMED), 2016.
- [工藤 04] 工藤拓, 山本薫, 松本裕治: Conditional random fields を用いた日本語形態素解析. 情処学 NL 研報, pp. 161-13 (2004).
- [Mikolov 14] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig: Linguistic regularities in continuous space word representations. In HLT-NAACL (2013).