

一般口演

一般口演18

医療データ分析6（機械学習・テキストマイニング1）

2018年11月25日(日) 09:00 ～ 10:30 D会場 (4F 413+414)

[4-D-1-6] Multi-Column Convolutional Neural Networks を用いた Web 文書に出現する患者の症状表現抽出、症状疾患関係認識

○和田 聖哉¹, 飯田 龍², 鳥澤 健太郎², 武田 理宏¹, 真鍋 史朗¹, 松村 泰志¹ (1.大阪大学大学院医学系研究科 医療情報学, 2.情報通信研究機構)

【背景・目的】 本研究では、患者に症状を入力させ、その症状に基づいて疾患名を推定する診断支援システムの構築に必要な症状表現、疾患名と対応する症状表現の自動獲得手法を提案する。これまでに作成された医療用語辞書には Comejisyo（相良ら、2008）、万病辞書（荒牧ら、2017）があるが、これらは医療従事者用に作成されたものであり、患者が使用する一般的な症状表現に関して網羅性が低い。本研究では患者が日常的に使用する症状表現(所見等も含む)の候補を Web 文書から抽出し、その候補に対して①症状表現を表す表現か否かを分類する症状表現抽出、②疾患名とそれに起因する症状表現の対か否かを分類する疾患症状関係認識の2種類の問題を独立に解くことで症状表現、疾患と対応する症状表現の対の自動収集を行う。【方法】 ICD-10対応標準病名マスター等から25,916件の疾患名リストを作成し、質問応答システム WISDOM X（Mizunoら、2016）に、「[疾患名]が何を引き起こす」といった質問を入力して症状表現候補とその候補が含まれる文を獲得する。次に名詞意味クラス（Kazamaら、2008）を用い、症状表現の含まれる可能性の高い候補のみを選別して約12万件の症状候補を得た。うち5万件をランダムに抽出し、症状表現抽出と疾患症状関係認識のそれぞれについて人手でアノテーションを行い学習・評価用データを作成した(7名のアノテータが従事)。さらに、Multi-Column Convolutional Neural Networks（Cireganら、2012）を用いた症状表現抽出器と疾患症状関係認識器を構築し、自動検出・自動認識の性能評価を行った。【結果】 本手法の症状表現抽出、疾患症状関係認識の性能はそれぞれ F1 値で 0.937、0.876 であった。新規の症状表現の抽出に関する評価結果は F1 値で 0.894 あり、本手法が既存辞書に含まれない症状表現についても高い性能で表現を抽出できることを示している。

Multi-Column Convolutional Neural Networks を用いた Web 文書に出現する患者の症状表現抽出、症状疾患関係認識

和田 聖哉^{*1}、飯田 龍^{*2}、鳥澤 健太郎^{*2}、武田 理宏^{*1}、真鍋 史朗^{*1}、松村 泰志^{*1}

^{*1} 大阪大学大学院医学系研究科 医療情報学、^{*2} 情報通信研究機構

Symptom Extraction and Disease-Symptom Relation Recognition from Web Texts with Multi-Column Convolutional Neural Networks

Shoya Wada^{*1}, Ryu Iida^{*2}, Kentaro Torisawa^{*2}, Toshihiro Takeda^{*1}, Shiro Manabe^{*1}, Yasushi Matsumura^{*1}

^{*1} Department of Medical Informatics, Osaka University Graduate School of Medicine,

^{*2} National Institute of Information and Communications Technology

We propose two methods: one for extracting symptom (or findings) while another for recognizing the relationships between diseases and their symptoms. We make a disease list which contains 25,916 disease names from ICD-10, the question criteria 2018 about doctor national examination in Japan and so on. We retrieve sentences which include candidate noun phrases for symptoms using the factoid QA module of WISDOM X with question such as “what does the disease cause?” We apply a dictionary of semantic noun classes to the candidate noun phrases, select a set of noun phrases that are likely to be symptoms and extract about 120,000 sentences that include such noun phrases. Then, we randomly sample 50,000 sentences from the extracted sentences. Seven annotators manually judged whether a given candidate noun phrase is a proper symptom or not. In addition, they also judged whether a given symptom is regarded as a proper symptom of a target disease or not. We create two multi-column convolutional neural networks to extract symptoms and recognize the disease-symptom relationship and conduct empirical evaluation using the annotated data. Experimental results show that the F-score of our two methods are 0.932 and 0.880, respectively. Especially in case of extracting new symptoms, which are even not contained Comejisyo and MANBYO Dictionary, the F-score is 0.883. Our methods can extract symptoms that are not included in existing dictionaries with high performance.

Keywords: Medical entity recognition, Relation extraction, Natural language processing, Deep learning

1. 緒論

医療情報記録の電子化ならびに人工知能研究の発展に伴い、自動診断を始めとした診断支援システムへの期待が高まっている。診断支援システムのためには、症状名（もしくは症状名の組み合わせ）が入力された際に入力した内容を照合するための症状表現一覧と、それに対応する疾患名の一覧を事前に獲得しておく必要がある。これらを収集するためのリソースとして考えられるものは、医学教科書、論文や電子カルテ文章など様々である。しかしこれらをリソースとした場合には次の問題が生じる。

- その大部分は自然言語文であり、情報抽出の処理が必要となる。
- 教科書に記載される表現は専門性が高く、一般の患者が入力する症状の表現と異なる可能性がある。

前者に対しては、自然言語処理技術を用いて、係り受け構文解析や深層学習を含めた機械学習により疾患に關係する表現（疾患名、治療、副作用、症状、所見など）の固有表現抽出や、関係抽出が行われている¹⁻⁶⁾。人手で整備された医学用語ソーラスとしては、アメリカ国立医学図書館によって定められた MeSH (Medical Subject Headings)⁷⁾や UMLS (Unified Medical Language System)⁸⁾、ヒト疾患に關する表現型語彙について体系的に整備された HPO (Human Phenotype Ontology)⁹⁾が代表的であるが、一般の患者が使用する症状表現を含んでいない。また、日本語のリソースでは Comejisyo¹⁰⁾や万病辞書¹¹⁾が利用可能であるが、いずれも専門性の高い文書から表現が取得されており、上記 ii の問題を解決するには不十分である。

この問題を解決するために、本研究では Web ページに出現する多様な症状表現の獲得、また、疾患名とその疾患名に

關係する症状表現の対の獲得を行う。

2. 目的

患者が日常的に使用する症状表現（症状、所見等を含む）の候補を Web 文書から抽出する。その候補に対し、症状表現か否かを分類する症状表現抽出と、疾患名とそれに起因する症状表現の対か否かを分類する疾患症状関係認識の2種類の問題を独立に解くことで症状表現、疾患と対応する症状表現の対の自動収集を行う（表1：赤字…疾患名、青字…症状表現候補。症状表現抽出では①「鈍痛」、②「嘔吐」を抽出する。疾患症状関係認識では①「偏頭痛-鈍痛」は正しい関係として出力し、②「蕁麻疹-嘔吐」は不適切な関係として出力する）。本研究では、深層学習を利用した予測モデルの構築とその性能評価を行う。

表1 自動収集した症状候補を含む文

①	それから、 偏頭痛 は、残っている親知らずの側（右）で、現在下あごの神経まで 鈍痛 がある状態です。
②	発汗や顔が急に腫れる、全身にひどい 蕁麻疹 が出る、吐き気や 嘔吐 、声が出にくい、息苦しいなどの症状が続き、やがてショック状態に陥る激しい全身反応を意味します。

3. 方法

この節ではまず本研究で収集対象とする「症状表現」について定義する。次に効率的に症状表現に關する文を収集し、対象データを作成する手法と、与えられた文に対して正しいか否かの情報を人手で付加する作業（アノテーションと呼ぶ）について説明する（図 1）。最後に本研究で用いる深層学習について述べる。

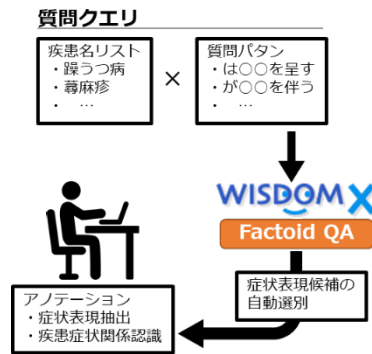


図1 本研究で提案する症状候補、疾患症状対候補の獲得手順

3.1 「症状表現」の定義

症状 (symptoms)とは主に患者から発せられる表現であり、一方で所見 (signs, findings)とは主に医療従事者によって客観的に記される表現である。どちらも疾患の特徴について言及したものであり、それらが疾患に関する文章内に出現する場合、構文的に症状と所見を区別することは難しい。また、診断システムを運用する場合、ユーザはその表現が症状か所見かを区別せずに入力すると考えられるため、本研究では、症状と所見を併せて「症状表現」と扱った。

また、疾患を説明する文章では、症状・所見とともに合併症が並列表記される場合がある。知識がなければ疾患名と症状・所見の判定が困難な表現の存在 (例えば「麻痺」ならば症状・所見だが、「顔面神経麻痺」は疾患名)に関して、その分類を予め決定する必要がある。

以上から、本研究では生体が疾患に罹患したことにより生じる変化を全て「症状表現」として獲得する対象とした。すなわち「症状表現」には症状や所見 (身体所見、検査所見を含む)、疾患名 (合併症)が含まれる。医療文書には治療方法や副作用の表現も出現するが、疾患により生じる生体の変化ではないことから、本研究ではこれらを収集対象としない。

3.2 対象データの作成

WISDOM X (<https://wisdom-nict.jp/>)は情報通信研究機構が開発、一般公開しているシステムであり、自然言語処理技術を用いて約 40 億件の Web 文書に対して種々の言語解析を行い、回答を抽出する大規模情報分析システムである¹²⁾。本研究では WISDOM X 内に実装されているファクトイド型 QA モジュール (例えば、「急性心筋梗塞は何を引き起こす?」に対して「心停止」「不整脈」「心室細動」といった回答を返す)を利用する。入力する質問は「(疾患名) + (質問パタン)」で定義される。質問クエリ作成のため、それぞれの一覧を準備する。

疾患名リスト作成のため、平成 30 年版医師国家試験出題基準¹³⁾、ICD10 対応標準病名マスター V 4.02¹⁴⁾から疾患名を取得した (それぞれ 2,119 個、24,553 個)。また、疾患名表現を拡張するために、医師国家試験出題基準必修領域に掲載されている疾患名 186 個を、ALAGIN 文脈類似語データベース¹⁵⁾に適用してそれぞれの類似語上位 100 語を獲得した。人手で判定を行い、1,776 個の疾患名を取得した。これらを結合後に重複除去処理を行い、25,916 の疾患名からなるリストを作成した (表 2)。

WISDOM X には、名詞のみを入力するとユーザに質問クエリを提案する機能がある。その機能を利用し、以下の手順で使用する質問パタンを選択した。

- i. 疾患名リストを WISDOM X に入力
- ii. 高頻度に出現する質問を抽出
- iii. 回答に症状表現が多く出現する質問パタンを判定
これにより、67 の質問パタンを設定した。質問パタンの例を表 3 に示す。

表2 疾患名リストに含まれる疾患名の例

平成 30 年版医師国家試験 出題基準	Crouzon 病、くも膜下出血、蕁麻疹、骨粗鬆症、…
ICD10 対応標準病名 マスター	声門癌、食道損傷、腹部てんかん、声帯結節症、…
ALAGIN 文脈類似語 データベース	じんま疹、リウマチ、骨粗しょう症、胃がん、…

表3 質問パタンの例

因果関係	は何が起る は○○を起こす が何を引き起こす …
症状に頻用される表現	は○○を伴う は何を呈す は○○で発症する …

準備した疾患名リストと質問パタンから質問クエリを作成し、WISDOM X へ入力する。得られた回答から効率的に症状表現候補を含む文を自動選別するために、述語の言語パタンによる選択と名詞意味クラスを適用する。

WISDOM X の質問応答システムでは、述語とそれに係る 1 つもしくは 2 つの名詞句からなる言語パタンがインデックス化され、検索時に用いられている。<A が B をもたらす>のような、述語とそれに係る項が 2 つの言語パタンをバイナリパタン、また項が 1 つの言語パタンをユナリパタンと呼ぶ。ユナリパタンは回答としてヒットしやすい反面、項が一つしかないため、回答精度は低いとされる。このため、本研究ではバイナリパタンに限定してデータを作成する。WISDOM X では、質問から抽出したバイナリパタン (例:A が B をもたらす) に関して、あらかじめ構築された含意パタンデータベースで、同義、もしくは含意しているバイナリパタン (例:A が B を引き起こす) を求め、そのバイナリパタンでも検索を行う。これによって、「急性心筋梗塞は何をもたらす?」という質問を入力しても、「急性心筋梗塞が*を引き起こす。」という文からも回答が得られる。これによって柔軟な質問応答が可能になる。総質問数 1,736,372 (25,916 疾患名×67 質問パタン) のクエリを WISDOM X に入力し、バイナリパタンと疾患名・症状表現候補を含む回答 3,894,833 事例を得た。症状表現候補の異なり数は 309,141 であった。次に、症状表現候補が作成した疾患名リストに含まれるものをそこから除外した。最後に、風間ら¹⁶⁾の単語意味分類手法を大規模 Web 文書 (約 6 億文書) に適用して作成された名詞意味クラス辞書を事前に分析し、特に症状表現が含まれる 10 クラスを特定した。この 10 クラスに含まれる表現のみを抽出した。これにより、疾患名数 7,423/症状表現候補の異なり数 32,550/疾患名・症状表現候補対の組み合わせ数 127,352 例のアノテーション候補を取得した。

アノテーション候補から 50,000 例をランダムサンプリングし、アノテーションを実施する。アノテーション基準を以下に記す。

【症状表現抽出】

症状表現候補が 3.1 で定義した症状表現であるか否かを分類する。判定は正例/負例の 2 値で行う。

症状表現候補のみでは症状表現になりえないが、身体部

位表現が症状表現候補に係ることで疾患による変化、すなわち症状を表す場合(例えば、「皮膚への害」、「肌への影響」などの下線部)や、症状表現候補が変化を表す表現で、疾患名等による修飾で症状表現となり得る場合(「改善」、「悪化」、「増大」、「縮小」、「不変」など)には、正例と判定する。

アノータが症状表現候補の表現自体を知らない場合には、Web 検索等を使用して調べることを許可する。

【疾患症状関係認識】

症状表現抽出で正例と分類された例に対して実施する。症状表現候補が対象疾患名により引き起こされたものであれば、正例と判定する。ここで、症状表現候補が対象疾患名に起因するか否かの判断は、与えられた文を読んで判断することに留意する。すなわち、対象疾患名と症状表現候補の関係を評価する上では医学的に正しいか否かの確認を、Web 検索等を用いて行う必要はない。文に書かれた内容のみでアノータが判断する(例えば、「アトピーは自律神経症状を起こします。」という文が与えられた場合、医学的信頼性は問わずに正例とする)。

症状表現抽出タスク、疾患症状関係認識タスクのアノテーション例を表4に示す。

表4 アノテーション実施例

疾患・症状に関係する文	症状抽出	関係認識
ショック、心不全、特定の薬または慢性腎不全による低血圧。	O	O
末期癌の父と6歳の娘私の父は末期癌で余命いくばくもありま	X	
頭蓋内石灰化、肝脾腫および肝障害、黄疸、小頭症、精神運動発達遅滞、痙攣、網膜炎などの症状を起こす。	O	X
激しい胸の痛みや冷や汗、意識の喪失などが起こる急性心筋梗塞。	O	O

赤字: 対象疾患名、青字: 症状表現候補。o: True、x: False

アノテーションは7名の非医療従事者により実施される。1つの事例につき3名のアノータが症状表現抽出タスクと疾患症状関係認識タスクを行う。これを、対象データ50,000例で実施後、以下の基準で集計する。

【症状表現抽出】

3人中2人以上が正例と判定する場合、症状表現として用いられる表現(True)と分類する。

【疾患症状関係認識】

症状表現抽出でTrueと分類した文に対して評価を行う。すなわち、必ず2-3名のアノテーション結果が存在する。3人中2人以上、もしくは2人中1人以上が正例と判定する場合、対象疾患名に起因する症状表現としてTrueに分類する。

症状表現抽出でFalseと分類した文に対しては、疾患症状関係認識の分類を全てFalseとする。

集計の結果、症状表現抽出タスクの正例は50,000例中26,757例(53.5%)、疾患症状関係認識タスクの正例は16,223例(32.4%)であった。Fleiss' kappa値はそれぞれ0.819、0.766であり、アノータ間で高い一致率を確認した。

アノテーション済みの文例50,000件を、train, valid, dev, testとして、表5に示す個数に分割した。

train データで学習を行った。各方法の最適なepochはvalid データを用いて決定した。最適なハイパーパラメータの選択には、dev データを用いて、最もF値の高いモデルに決定した。最適なハイパーパラメータを決定後、それらを用いた

各方法について、test データで最終的な精度評価を行った。また、未知の表現に対する精度を評価するために、train, valid に出現する症状表現候補12,523パターン(正例/負例を問わない)と、既存の医学用語辞書(Comejisyo¹⁰⁾、万病辞書¹¹⁾に含まれる表現210,848パターンに一致する症状表現候補をtest データから除外したデータセットunkを用意した。症状表現抽出タスクの正例は1,056例中596例(56.4%)、疾患症状関係認識タスクの正例は380例(36.0%)であった(表5)。

表5 データセット: 症状表現抽出/疾患症状関係認識

	総事例数	症状表現抽出 正例数	疾患症状関係認識 正例数
train	35,000	18,763	11,356
valid	5,000	2,620	1,595
dev	5,000	2,667	1,682
test	5,000	2,707	1,590
unk	1,056	596	380

3.3 畳み込みニューラルネットワークを用いた 症状表現抽出、疾患症状関係認識

Multi-column convolutional neural networks (MCNN)はCiresanらにより画像分類手法として発表された¹⁷⁾。自然言語処理分野では、Dongらが質問応答に、Zengらが関係抽出にMCNNを利用した^{18, 19)}。本研究のMCNNアーキテクチャはChopraらの研究²⁰⁾から着想を得ており、それをMCNNに拡張して出力の類似度をsoftmaxに置換している。同様の構造を有するMCNNアーキテクチャを利用した研究には、ゼロ照応解析²¹⁾、因果関係認識^{22, 23)}がある。MCNNは因果関係認識の原因句と結果句の関係の認識、ゼロ照応解析の省略と補完表現候補の関係の解析を近傍文脈の情報も利用することで高い性能で解析できることがわかっており、(疾患名が文脈中に出現している状態での)症状表現の抽出や疾患症状関係認識でも有効であると考え、MCNNを利用した。

MCNNの各カラムは畳み込み層、プーリング層、全結合層が1層ずつ存在する畳み込みニューラルネットワークで構成される。入力として与えられる文中の疾患名と症状表現候補に関してどのように各カラムに入力すると高い性能が得られるかは自明ではない。これらの入力方法に関していくつかの方法を提案し、どの方法が最終的に良い検出・認識の結果を得るかの調査を行う。

3.3.1 各カラムへの入力内容

バイナリパタンの文章は対象疾患名Disと症状表現候補Symの出現位置により、5つのブロックに分割される(図2a: DisとSym以外のブロックを、左から順にa, b, cとする)。Symの前後にある表現でDisとの関係性が重要であり、この位置関係に注目した入力内容を設定する。各ブロックの内容とその出現位置情報を元に、以下を定義する(図2b)。

- BASE: Dis+Sym (2カラム)
- NOFIXED: a+b+c (3カラム)
- FIXED: b+c+a or a+b+c (3カラム)

本研究で使用するMCNNの概要を図3に示す。入力単語列は事前に学習済みの分散表現モデル(2015/01/18の日本語版Wikipedia記事から抽出した全35,975,219文を、Negative Samplingを用いたSkip-gramモデル²⁴⁾で構築: 500次元・1,658,487語)により、(単語数)×500から成る行列に変換される。変換後の行列は各カラムの畳み込みニューラルネットワークに入力される。畳み込み層では、フィルタ数は{100,

300}、フィルタ幅は{(3, 4, 5), (4, 5, 6)}の組み合わせを試みる。プーリング層では max-pooling を行い、その結果を全て結合し最終層とする。これにより、最終層の全ノード数は 600～4,500 になる(カラム数{2, 5}×フィルタ幅数 3×フィルタ数{100, 300})。

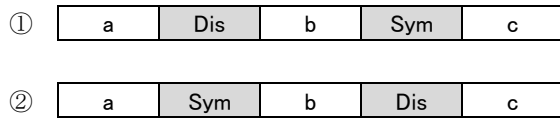
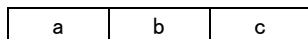


図 2a バイナリパタンの構成

i. BASE

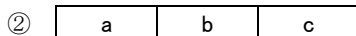
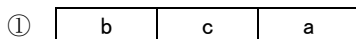


ii. NOFIXED



Dis と Sym の位置に関わらず、左から順に並べる。

iii. FIXED



Sym の直前、直後にあるブロックをそれぞれカラム 1、カラム 2 にする。残りをカラム 3 にする。

図 2b BASE, NOFIXED, FIXED の設定

学習にはミニバッチ確率的勾配降下法を用いる。過学習を避けるために、valid データを用いて最適な epoch を決定する。早期終了と、50%の最終層ドロップアウトを採用する。最終層に線形 softmax 関数を適用し、症状表現抽出もしくは疾患症状関係認識ラベル (True/ False) のクラス確率を生成する。ハイパーパラメータの決定には dev データを用いて行い、各予測モデルの精度を相互に比較する。

MCNN は PyTorch²⁵⁾で実装した。与えられた文の形態素解析には MeCab²⁶⁾を用いた。形態素解析用辞書には、MeCab とともに公開されている Juman 辞書を使用した。

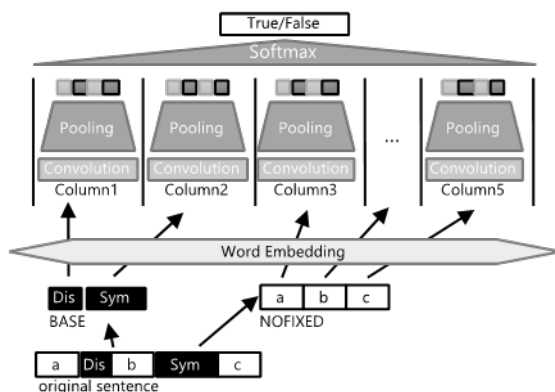


図 3 本研究の MCNN アーキテクチャ

4. 結果

4.1 症状表現抽出器／疾患症状関係認識器

入力パターンを組み合わせ、各々で最適なハイパーパラメータを、dev データを用いて決定し、相互に比較した。

予測モデルの比較対象として、全ての予測を正例と判定した場合の精度を確認した(これを NAIVE とする)。症状表現抽出器では、NAIVE の F 値は 0.706 であった。BASE で F 値 0.936 と上昇した。周囲の表現を入力しても、精度はそれ以上改善しなかった (BASE+NOFIXED/FIXED)。このモデルで

test データを評価したところ、Recall 0.932、Precision 0.931、F 値 0.932 であった(表 8)。

疾患症状関係認識器では、NAIVE の F 値は 0.516 であった。BASE で F 値 0.778 であった。一方、最も高い精度を得た入力パターン BASE+FIXED の F 値は 0.883 であった。Sym との出現位置関係を考慮しない場合 (BASE+NOFIXED)、F 値は 0.835 に低下した。最高精度を得た予測モデルによる test データの評価は、Recall 0.883、Precision 0.876、F 値 0.880 であった(表 8)。Precision-Recall curve を図 4 に示す。

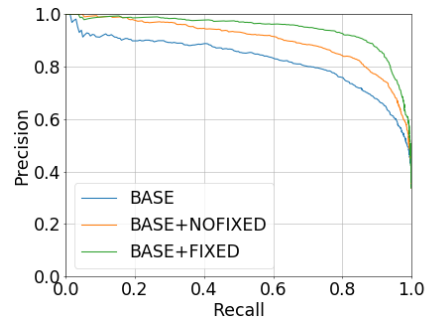


図 4 疾患症状関係認識器の Precision-Recall curve

4.2 未知の表現に対する予測モデルの評価

unk データに対し、4.1 で構築した予測モデルを適用した。症状表現抽出器と疾患症状関係認識器の F 値はそれぞれ 0.883、0.869 であり、全評価データの精度から少し低下した。(表 8)。各モデルで正しく取得できた例を表 6 に示す。

表 6 未知語データセットから正確に取得できた表現例

症状表現抽出	疾患名	
十二指腸嵌頓、苔癬様型皮疹、急性皮質壊死、かきかたストレス、くしゃみ・鼻水・咳、最近口角炎、口臭、排尿機能障がい、むかつき感、…		
疾患症状関係認識	疾患名	症状表現
	深在性真菌症	抗菌薬不応性発熱
	大腿骨頸部骨折	ロコモティブシンドローム
	前庭神経炎	持続的振振
	肺気腫	心臓衰弱
	前立腺肥大	男性おしっこ頻尿

5. 考察

本研究で構築した予測モデルは、高精度で症状表現の抽出と、対象疾患名と症状表現の関係を認識することが出来た。

症状表現抽出において、BASE (Dis+Sym)とそれ以外の文章情報を入力したものとでは、その予測精度に変化はなかった。今回構築したモデルは、症状表現か否かを判定する上では対象疾患名と症状表現候補の情報のみでも、高い精度で判定出来ることを示している。その一因として、事前に学習させた単語分散表現モデルの利用が挙げられる。医学表現に用いられる単語が、分散表現モデル上で既にその関係性を評価されていたことが、本研究では有利に働いたと考えられる。

疾患症状関係認識において、BASE と BASE+NOFIXED では F 値はそれぞれ 0.772 と 0.819 であった。この結果から、対象疾患名と症状表現候補の周囲表現を MCNN の入力に含めることで、疾患症状関係認識の精度が上昇することがわかった。また、BASE+FIXED の F 値は 0.880 と BASE+NOFIXED よりも高値を示した。これは、周囲の表現に関して、症状表現候補との前後位置情報を追加することで、精度が上昇することを示している。以上から、症状表現候補の周囲表

現とその出現位置情報は疾患症状関係認識には必要な情報であると言える。本研究で高い精度が得られた要因には、アノテーションで指示した、取得する疾患症状関係の判断基準と、取得した文のパターンにあると考えた。前者に関しては、疾患名と症状表現関係の判定を構文から読み取れる範囲の因果関係に制限したことで、正例における周囲表現の特徴を捉えやすくなった可能性がある。後者では、取得した文例中に症例報告のタイトルに使われるものが散見されたことが一因と考えた。症例報告のタイトルは、「Sym を呈した Dis の一例」や「Sym を伴った Dis」「Sym で発症した Dis」など、注目を集めるために定番のフレーズに集約される傾向がある。また、規定文字数に収めるために Dis と Sym が短いフレーズ内で出現するので、バイナリパターンとしても認識されやすい。ゆえに、対象疾患と症状表現の因果関係を説明する文の頻出パターンとして、症例報告タイトル文の存在が、その高い精度に寄与した可能性がある。

次に誤判定事例を検討する。疾患症状関係認識器で予測モデルが誤った事例 (FP: False Positive/ FN: False Negative) をサンプリングし、その判定を再評価した (表 7)。その一部には、アノテーション結果の誤りのために False Positive や False Negative に分類されたものが存在した (True-True/ False-False)。表で示した事例を整理すると「うつ病-日内変動」の関係は正しく、「被害妄想-フラッシュバック」の関係は誤りであることがわかるが、難読な文である。他にも、自動翻訳文や症例報告タイトルなどでは、与えられる文が日本語として不自然になりがちであり、アノテータ間で判定が分かれやすい事例となっていた。本研究では、関係評価に際して Web 検索や個人の知識は用いず行うよう指示をしていた。このような事例を解決するには、対象疾患名と症状表現候補で Web 検索を行った結果をアノテーション作業時に自動的に提示し、判定の参考にするような補助を行うことなどが改善案として検討される。

表7 疾患症状関係認識における
予測モデル誤判定事例の再評価

		再判定後の分類	
		True	False
予測モデルの判定	True	うつ病は脳内物質の一種のセロトニンやノルアドレナリンが減少し、憂うつ、悲しい、何の希望もない、落ち込んでいるというような感情で、午前中にひどく、午後から夕方にかけて改善してくるという日内変動がある。	・【林寛之】日常診療のピットフォール偽痛風による発熱をきたした認知症の1症例【笠松哲司／寺澤敏秀／三原正朋／鶴尾美穂、瀬尾浩二／小谷和彦】ただいま後期研修中！
	False	視神経炎(乳頭炎、球後視神経炎)では中心暗点やラケット状暗点(石津暗点)が見られる。	でも実はそれは全て幻覚で、むかしひき逃げしたことを忘れフラッシュバックで実は自分が被害妄想を感じてるだけだった。

赤字: 対象疾患名、青字: 症状表現候補

予測モデルの判定が True/ False のものがそれぞれ、元々の False Positive/ False Negative 事例に対応する。

構築した予測モデルの新規症状表現および疾患症状関係の取得能力について検討した。学習データに含まれない症状表現候補が出現する事例で精度評価を実施し比較した。

同時に新規表現獲得能を評価するために、既存の医学用語辞書に含まれない表現という条件を追加して行った (表 8、未知語のデータ (unk))。F 値は全評価データ (test) に比較すると少し低下するが、それでも症状表現抽出と疾患症状関係認識において高い精度を維持することが確認出来た。

今回収集した症状表現を利用する際には、得られた新規表現には並列表記 (「くしゃみ・鼻水・咳」、「失語症 & 失認症」…) や不要な文字列を含む症状表現 (「最近口角炎」、「13アテトーゼ型脳性麻痺」…) が混在していることが課題となる (表 6)。未知語データ unk から正確に取得できた症状表現 488 例中、並列表現は 28 例、不要な文字列を含む症状表現は 111 例あった。これを人手で修正し再集計したところ、症状表現の異なり数は 524 例で、このうち学習データと既存の医学用語辞書に含まれない表現は 378 例 (72.1%) あった。以上から、得られた結果に対して修正を行えば、新規の症状表現が獲得できることを示せた。その内容は Web 文書を情報源としており、医学用語辞書にはない多様な症状表現や疾患症状関係を収集出来る可能性がある。

本研究では、症状表現候補のみでは症状表現になりえないが、症状表現候補にかかる修飾表現で症状を表す事例 (「～の(状態)」、「～の(症状)」、「～の(喪失)」…) も取得した。これを修飾表現も含めた形で症状表現辞書を整備すれば、例えば「赤く腫れた状態」、「ムカムカする症状」などが症状表現として利用可能になる。そのためには、構文解析や機械学習などにより、今回収集した名詞(句)にかかる修飾表現を収集する必要がある。

本手法で収集した疾患症状関係認識の限界として、医学的信頼性が不十分であることが挙げられる。この原因の一つは、情報源が Web 文書であり、医学書からの情報よりは信頼性が低いことである。もう一つは、疾患症状関係認識では症例報告のタイトルから取得した内容が一定数混在するため、その疾患に伴う症状表現としては稀なものが含まれる可能性があることである。視点を変えると、医学書に記載されていない表現の取得が可能と見ることも出来るが、一般に対象疾患との関連性の低い症状表現に関しては、その扱いには慎重になる必要がある。これを解決するには、医学書から症状表現抽出と疾患症状関係認識を行い、深層学習技術も利用しつつ、本研究で収集した疾患、症状名の候補とそれらの同義性、意味的な類似性を判定し (例: 「声枯れ」「しゃがれ声」= 「嗄声」)、場合によっては人手による確認を経て、信頼の置けるデータのみをデータベースに格納する等の新規技術や方法論を開発していく必要がある。

6. 結論

本研究では、深層学習により Web 文書から症状表現抽出ならびに疾患症状関係認識を実現する手法を開発した。

獲得した症状表現には既存の医学用語辞書には存在しない表現が含まれており、多様な表現を獲得しうる可能性を示せた。一方で、修正の必要な並列表現等が存在することや Web 文書の医学的信頼性が問題になることから、利用する際にはその内容について再評価を要する。人手によるさらなる整備については出力のスコアに基づいて効率的に行える可能性があり、今後の課題として検討する。

本研究により一般の患者が入力する症状表現とそれに関係する疾患名を関連付けた辞書を自動的に整理するシステムが確立できれば、患者が入力した症状に基づいて疾患名を推定する診断支援システムの構築の第一歩になるものと期待する。

参考文献

- 1) Chowdhury FM, Lavelli A. Disease Mention Recognition with Specific Features. BioNLP 2010 : 83-90.
- 2) Martin L, Battistelli D, Charoïs T. Symptom recognition issue. BioNLP 2014 : 107-11.
- 3) 矢野憲, 伊藤薫, 若宮翔子, 荒牧英治. 深層学習による医療テキストからの固有表現抽出器の開発とその性能評価. 人工知能学会全国大会 (JSAI), 2017.
[https://www.jstage.jst.go.jp/article/pjsai/JSAI2017/0/JSAI2017_2J2OS16a4/_article/-char/ja/ (cited 2018-Aug-23)]
- 4) Hassan M, Makkaoui O, Coulet A, Toussaint Y. Extracting Disease-Symptom Relationships by Learning Syntactic Patterns from Dependency Graphs. BioNLP 2015 : 71-80.
- 5) Muzaffar AW, Azam F, Qamar U. A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set. Computational and Mathematical Methods in Medicine 2015 : 1-12.
- 6) Lv X, Guan Y, Yang J, Wu J. Clinical Relation Extraction with Deep Learning. International Journal of Hybrid Information Technology 2016 ; vol. 9, no. 7 : 237-48.
- 7) The National Library of Medicine (NLM). Medical Subject Headings (MeSH).
[https://www.nlm.nih.gov/mesh/ (cited 2018-Aug-23)].
- 8) Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research 2004 ; 32 (suppl_1) : D267-70.
- 9) Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Research 2017 ; 45 : D865-76.
- 10) Comejisyo. ComeJisyo プロジェクト日本語トップページ - OSDN.
[https://ja.osdn.net/projects/comedic/ (cited 2018-May-29)]
- 11) 万病辞書. MEDNLP 医療言語処理グループ 万病辞書.
[http://mednlp.jp/DIC/ (cited 2018-May-29)].
- 12) Mizuno J, Tanaka M, Ohtake K, et al. WISDOM X, DISAANA and D-SUMM: Large-scale NLP Systems for Analyzing Textual Big Data. COLING 2016 : 263-7.
- 13) 厚生労働省. 平成30年版医師国家試験出題基準について.
[https://www.mhlw.go.jp/stf/shingi2/0000128981.html (cited 2017-July-3)]
- 14) MEDIS. ICD10対応標準病名マスター.
[https://www2.medis.or.jp/stdcd/byomei/ (cited 2017-July-10)]
- 15) 情報通信研究機構. ALAGIN 言語資源・音声資源サイト.
[https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html (cited 2017-July-3)]
- 16) Kazama J, Torisawa K. Inducing gazetteers for named entity recognition by large-scale clustering of dependency relations. Proceeding of ACL-08 2008 : 407-15.
- 17) Cireşan D, Meier U, Schmidhuber J. Multi-column Deep Neural Networks for Image Classification. In Proceedings of Computer Vision and Pattern Recognition 2012 : 3642-9.
- 18) Dong L, Wei F, Zhou M, Xu K. Question Answering over Freebase with Multi-Column Convolutional Neural Networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing 2015 : 260-9.
- 19) Zeng D, Liu Z, Chen Y, Zhao J. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing 2015 : 1753-62.
- 20) Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of Computer Vision and Pattern Recognition Conference 2005 : 539-46.
- 21) Iida R, Torisawa K, Oh JH, Kruengkrai Canasai, Kloetzer J. Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing 2016 : 1244-54.
- 22) Oh JH, Torisawa K, Kruengkrai C, Iida R, Kloetzer J. Multi-Column Convolutional Neural Networks with Causality-Attention for Why-Question Answering. In Proceedings of WSDM 2017 : 415-24.
- 23) Kruengkrai C, Torisawa K, Hashimoto C, Kloetzer J, Oh JH, Tanaka M. Improving Event Causality Recognition with Multiple Background Knowledge Sources Using Multi-Column Convolutional Neural Networks. In Proceedings of AAAI-17 2017 : 3466-73.
- 24) Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In Proceedings of Advances in NIPS 2013 : 3111-9.
- 25) the PyTorch core team. PyTorch.
[https://pytorch.org/ (cited 2018-Jan-26)]
- 26) Kudo T, Yamamoto K, Matsumoto Y. Applying conditional random fields to Japanese morphological analysis. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing 2004 : 230-7.

表8 各データセットに対する症状表現抽出器、疾患症状関係認識器の精度評価

		dev データ				全評価データ (test)				未知語のデータ (unk)			
		Recall	Precision	F-score	Avg.P	Recall	Precision	F-score	Avg.P	Recall	Precision	F-score	Avg.P
症状表現抽出	NAIVE	1.000	0.546	0.706	-	1.000	0.556	0.714	-	1.000	0.576	0.731	-
	BASE	0.945	0.927	0.936	0.985	0.932	0.931	0.932	0.984	0.871	0.895	0.883	0.964
	BASE+NOFIXED	0.925	0.945	0.935	0.985	0.914	0.949	0.931	0.986	0.849	0.917	0.882	0.969
	BASE+FIXED	0.929	0.939	0.934	0.986	0.920	0.947	0.933	0.985	0.867	0.927	0.896	0.969
疾患症状関係認識	NAIVE	1.000	0.348	0.516	-	1.000	0.332	0.498	-	1.000	0.370	0.540	-
	BASE	0.798	0.758	0.778	0.826	0.799	0.747	0.772	0.839	0.761	0.732	0.746	0.833
	BASE+NOFIXED	0.854	0.817	0.835	0.901	0.846	0.794	0.819	0.898	0.808	0.789	0.798	0.892
	BASE+FIXED	0.874	0.893	0.883	0.943	0.883	0.876	0.880	0.943	0.866	0.873	0.869	0.939