

一般口演 | 知識工学

一般口演10 知識工学

2019年11月23日(土) 09:00 ~ 11:00 C会場 (国際会議場 2階国際会議室)

[3-C-1-04] 診療記録からの疼痛に関する事実性判定

○柴田 大作¹、河添 悦昌²、篠原 恵美子²、荒牧 英治¹（1. 奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域 ソーシャル・コンピューティング研究室, 2. 東京大学大学院医学系研究科 医療AI開発学講座）

キーワード：Natural Language Processing, Medical Informatics, Pain

【背景】医療において、疼痛のマネジメントや医療者間での情報共有は重要である。発生した疼痛とその性状が構造化情報として入力されれば、疼痛ケアの質の評価や介入を要する症例の抽出等に役立つと考えられるが、その多くはフリーテキストとして記載されることが多いため、その利用が課題となる。

【目的】自然言語処理技術を用いることで、診療記録に疼痛の情報が含まれているかを識別するための手法を検討する。また、識別に必要な学習データのアノテーション方法についての検討も同時に行う。

【方法】東京大学医学部附属病院の診療・看護記録からランダムに約1,000件を抽出し、対象として使用した。対象記録にはさまざまな表現で疼痛に関する症状が記載されるため、大量に存在する記録から候補となる記録を絞り込むために分散表現を用いた。そして識別では、対象記録に疼痛の記載があるのかなのかといった2値分類だけでなく、疼痛が現在発生しているのか、過去に発生したのか、発生が否定されているのか、潜在的に起こりえるイベントとして記載されているのかという詳細な5クラスの事実性判定を行った。

【結果】病名などを事前学習した固有表現抽出器を用いて、特徴量として用いた時のマクロ F値が0.49と最も高かった。

【考察・まとめ】本研究では文書単位でアノテーション作業と識別実験を行ったが、1つの記録から複数のラベルが推測されるような場合は識別が困難であることが分かった。また、クラスの細分化を行ったことで、アノテーション作業の難易度が上がり、アノテーションにばらつきが発生することが確認された。そのためこれらについて今後の更なる検討が必要であると考えられる。

診療記録からの疼痛に関する事実性判定

柴田大作^{*1}、河添悦昌^{*2}、篠原恵美子^{*2}、荒牧英治^{*1}

*1 奈良先端科学技術大学院大学 先端科学技術研究科、*2 東京大学大学院 医学系研究科

Factuality Analysis of Pain Expressions in Clinical Notes

Daisaku Shibata^{*1}, Yoshimasa Kawazoe^{*2}, Emiko Shinohara^{*2}, Eiji Aramaki^{*1}

*1 Nara Institute of Science and Technology, *2 Graduate School of Medicine, The University of Tokyo

[Background] Sharing information for pain between medical staffs and pain management of patients are significant in medical domain. However, information for patient's pain is generally written as a free text in clinical notes. Therefore, it is extremely challenging to extract it from clinical notes. In this study, we attempt to extract pain expressions from clinical notes and conduct factuality analysis of them with Natural Language Processing technique.

[Material] Clinical notes are obtained from The University of Tokyo Hospital which were generated in July, 2016. These are divided into sentences based on space, indentation and comma. and 10,000 sentences are randomly selected from them. Medical experts annotate pain expressions in the data.

[Methods] We apply machine learning technique to extract pain expressions from the data. Our machine learning model use a Bi-LSTM CRF model that is composed of embedding layer, Bi-directional Long Short-Term Memory and Conditional Random Fields.

[Result] Our model achieved the maximal average F1 score of 52.48 (Maximum value is 87.12).

[Discussion] The experimental results show that our model could extract pain expressions with the practical accuracy.

Keywords: Natural Language Processing, Machine Learning, Medical Informatics, Pain

1. 緒論

疼痛のマネジメントを行う上で、医療者間で患者に生じている疼痛を情報共有することは重要である。患者に発生している疼痛とその性状が構造化された情報として入力されれば、その抽出は容易であるが、現状その多くは診療記録や看護記録にフリーテキストとして記載されるため、抽出することは困難である。

医療文書からの情報抽出やその分析を行う研究はこれまでも報告されている。荒牧ら¹⁾は日本内科学会の症例報告を用いて、文書から病名を自動で抽出する病名抽出器を開発した。これは病名を認識する事象認識とその病名の極性を分類する事実性判定(陽性・陰性分類)を同時に行うものである。一般的には事象認識の後に陽性・陰性の極性分類が行われるが、荒牧らはアノテーションの段階でタグに極性情報の属性を付与することでこれらを同時に実現した。また、文字ベースの Conditional Random Fields を用いた実験では陽性病名の抽出性能が F 値で 83.3、陰性病名の抽出性能が F 値で 55.4 であった。

Ma ら²⁾は退院サマリ、経過記録や看護記録などのカルテ記載から肥満、喫煙と飲酒の状態に関する情報を含む文を自動で識別する機械学習モデルを構築した。文を窓サイズを考慮した Bag of Words によりベクトル化しオーバーサンプリングを行った後に、サポートベクターマシンを用いて識別することで、高い性能を達成したことを報告している。またアノテーションでは、喫煙と飲酒の状態はその時制(現在の喫煙や過去の喫煙)まで含めたアノテーションを実施している。

Dang ら³⁾は有害薬物反応について報告している Tweet データと PubMed の概要データセットを用いて、それに関する言及を抽出する固有表現抽出器を開発した。Attention 機構を有する深層学習モデルにより End-to-End で抽出を行った結果、F 値がそれぞれ 0.844 と 0.906 であり、State-of-the-Art の性能を達成したことを報告した。

Shared Task として Informatics for Integrating Biology to the

Bedside (i2b2)の Smoking Challenge⁴⁾がある。これは退院サマリから患者の喫煙状態(最近の喫煙、過去の喫煙、時制不明の喫煙、喫煙をしない、不明)を識別するものである。2006年に開催されたタスクではあるが、医療文書のイベント分類方法や評価方法などが詳細に定められており、重要な知見が多いと考えられる。

Kagawa ら⁵⁾は診療記録から患者の罹患する病名を自動抽出する上で、異なる意図で記載される病名(鑑別疾患や家族歴など)の割合を調査した。これによれば診療記録に出現した病名が患者の罹患する病名であった割合は全体の 78%と少なく、事象認識だけでは不十分であり、事実性判定を行うことの重要性を指摘した。

本研究では診療記録や看護記録からの疼痛表現の事象認識と事実性判定を行う。これらの文書に出現する疼痛表現は、患者に現在生じている疼痛だけでなく、将来起こり得るリスクとしての疼痛やコントロールされており問題となっていない疼痛など異なる意味で使用される。そのため疼痛表現の事象認識だけでは不十分であり、その事実性まで判断する必要がある。また、事実性判定においては関連研究¹⁾²⁾よりさらに細かい粒度での事実性判定を行う。これらの実現には機械学習を適用するのが一般的であると考えられるが、明確な基準に基づいてアノテーションされた学習データが必要である。しかし我々の知る限り、診療記録や看護記録で出現する疼痛表現について調査した研究は報告されておらず、疼痛表現のアノテーションを実施した研究も同様に報告されていないため、検討の余地があると考えられる。

2. 目的

本研究では疼痛表現の事象認識と事実性判定の実現に向けた診療記録や看護記録へのアノテーション方法を提示し、正解ラベル付きのコーパスを作成する。このコーパスを利用した自然言語処理技術に基づく事象認識と事実性判定を行う抽出器を構築し精度を評価する。

3. 方法

3.1 実験材料

実験材料として東京大学医学部附属病院で 2016 年の 6 月と 7 月に作成された診療記録と看護記録 (以下、対象記録) を使用した。6 月に作成された対象記録をアノテーションに、7 月に作成された対象記録を分散表現の学習に使用した。

本研究は東京大学大学院医学系研究科の医学部倫理委員会により承認 (承認番号: 10791) を得て実施した。

3.1.1 実験材料の絞り込み

一般的に機械学習を適用するにはラベル付きの学習データが必要である。ラベルは基本的には人手で付与されることが多く、金銭的・時間的なコストが大きい作業である。6 月に作成された対象記録の合計は 297,536 件であったが、一般的にこれら全てに対して人手でラベル付けをすることは困難である。そのため対象記録の絞り込みを行った。その手順を以下に示す。

Step1: 基本となる検索クエリの決定

特定の単語 (以下、検索クエリ) を含む対象記録のみを抽出することで絞り込みを行う。最初に基本となる検索クエリを既存の患者表現辞書¹から取得した。これはクラウドソーシングにより患者が使用する表現を収集したリソースである。このリソースから本研究で対象とする症状である疼痛の際に患者が使用する表現を取得し、日本語形態素解析器 Juman++²による形態素解析を行い、品詞が名詞、形容詞と副詞である単語を 29 単語取得し、基本となる検索クエリとした。検索クエリの例を図 1 に示す。

ズキズキ、痛風、手、苦痛、激痛、痛苦、痛い、痛み

図 1 患者表現辞書から取得した疼痛の患者表現の例

Step2: 検索クエリの拡張

Step1 において基本となる検索クエリを決定したが、単語数が少ないため検索クエリとして不十分である可能性がある。そのため分散表現を用いて類似語を取得することで検索クエリの拡張を行った。

分散表現は単語を密なベクトルで表現する手法の 1 つで、類似する文脈で使用される単語同士の分散表現はその距離が近くなる特徴を有している。この特徴を利用し、基本となる検索クエリと類似する単語を取得し、これらを拡張された検索クエリとした。

基本となる検索クエリと類似する単語を分散表現から 20 単語ずつ取得した。この作業を基本となる検索クエリ全てに対して実行し、重複を削除した後、拡張された検索クエリとした。この結果、基本となる検索クエリと拡張された検索クエリ (これらをまとめて検索クエリとする) の合計は 435 単語となった。図 2 に拡張方法の例を示す。

分散表現の学習には Facebook 社のオープンソースライブラリである fastText³を用いた。学習データには 7 月に作成された対象記録 208,506 件を文ごとに分割し、Juman++により単語分割を行ったものを使用した。またパラメータである分散表現の次元数は 100、窓サイズは 5 とし、それ以外はデフォルト値を適用した。

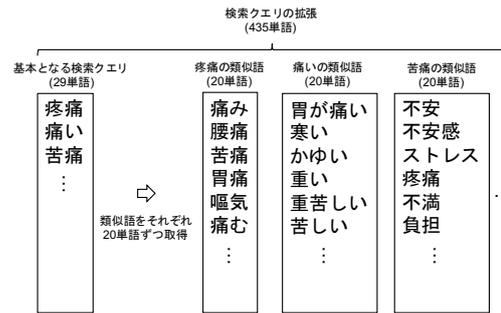


図 2 基本となる検索クエリの拡張

Step3: 検索クエリによる対象記録の抽出

検索クエリを含む対象記録を文字列一致により抽出した結果、対象記録 (6 月分) が 297,536 件から 164,444 件となった。これらを読点と改行で分割し、文書単位から文単位に変換することで 1,778,251 件の文を得た。そしてその中から 10,000 件をランダムサンプリングしたものをアノテーションの対象として使用した。

3.2 アノテーション方法

本研究におけるアノテーションは文に出現する疼痛表現に正解ラベルとなるタグを付与することを意味する。それを行う上で、正解ラベルの分類基準とラベル付けを行う手順が必要となる。これらに関しては、以下 3.2.1 節と 3.3 節で述べる。一方、タグ付けの方法に関して、事象認識と事実性判定を同時に実現するため、荒牧ら⁴のタグ付けの方法を参考とした。荒牧ら⁴との違いは、図 3 に示すように、文に対してタグ付けを行うのではなく、文を事前に分割し単語の単位でタグ付けを行った。その理由として、分割せずにタグ付けを行うと、後の処理で形態素解析を行った際に、単語の一部に疼痛表現が含まれ、3.5 節で示す IOB2 タグに変換することが困難になる場合があるためである。

・右鼠経の<P>痛み</P>がおさまった
・右鼠経の<P>痛み</P>がおさまった

図 3 正解ラベルのタグ付け方法: 上段が荒牧らの方法、下段が我々の方法である。空白は単語境界を示し、<P></P>に囲まれた部分が疼痛表現を示す。

3.2.1 正解ラベルの分類基準

3.1.1 節の Step3 で取得したアノテーションの対象記録とは別に 5,000 件の文を抽出し、正解ラベルの分類基準について検討した。この結果、疼痛表現は患者に生じている、もしくは生じていた疼痛の中にもコントロールされた疼痛があり、これらを分けて抽出することが有用であると考えられた。また、これら事実としての疼痛以外に、医療者が行った処置やケアプランとしての文脈から、疼痛の存在が推定される場合や、将来の発生が予期される疼痛があることもわかった。このような観点から、本研究では、以下の 4 種類の属性値の組み合わせで正解ラベルの分類をすることとした。以下に各属性値の分類基準を示す。

1. 疼痛の有無

患者への疼痛が肯定されている (明確に肯定されていないが、推論される場合も含む)、否定されているもしくは記載なしかを分類する。また患者以外の疼痛である場合はアノテーション対象外とし、学習データから除外する (3.2.2 節)。

¹ <http://sociocom.jp/~data/2019-pde/>

² <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

2. 疼痛の発生した時間区分

疼痛が肯定されている場合、その発生した時間が現在、過去もしくは未来かを分類する。未来は将来生じる可能性のある疼痛、過去は疼痛が消失していることが明らかである場合や既往歴の場合、それら以外の時間は現在とする

3. コントロールの有無

疼痛が生じている場合、それがコントロールされており問題となっていないのか、不明なのかを分類する。疼痛が自制内やコントロールされていることの記載がある場合はコントロールされているとする。また、緩和、軽減や増強なしといった表現も同様に扱う。

4. 事実としての疼痛か医療知識に基づく推論が必要か

疼痛が生じている場合、それが事実として記載されているのか、もしくは医療知識に基づいて推論されるのかを分類する。推論に基づいて判断される疼痛表現の場合、医療知識がない学習器では記載されている内容から疼痛が生じていることが判断できない可能性があるため、学習器の誤り分析をする上でこのような区分を設けることは有用であると考えた。

事実か推論かを分類する基準は、疼痛のコントロールが不明の場合、疼痛が生じている明確な記載がないが、何かしらの処置や治療が行われており、それが疼痛の処置や治療として一般的に行われる場合には推論とし、それ以外なら事実とする。この場合は、医薬品名や病名、処置や治療表現に対してタグを付与する。また、疼痛のコントロールが自制内の場合、処置や治療に基づいて疼痛が軽減、消失した結果として否定されている場合は、推論によって疼痛がコントロールされているとし、それ以外は事実によって疼痛がコントロールされているとする。

3.2.2 除外基準

以下の基準に該当する対象記録はアノテーションが困難であるため、学習データから除外した。

- アノテーションに他の文を参照する必要がある文
- 内服のみの記述、十分な文脈がなく、アノテーションが困難な文
- アノテーションに影響する誤字を含む文
- 処方薬の用法に疼痛表現が出てくる場合
- その他アノテーションが困難な文

3.3 正解ラベル付けの手順

4種類の属性値の組み合わせは36通り(疼痛の有無が3通り、時制が3通り、コントロールの有無が2通りと倫理が2通り)となるが、疼痛の有無が「否定」もしくは「記載なし」の場合には、それ以外の属性の値を考慮する必要はない。そのため正解ラベルとなるタグの総数は14通り(表1)となった。

また、対象記録には一般的知識としての疼痛など患者に生じているわけではない疼痛表現が見られるため、このような表現を分類するためのものとして<other>タグを設け、合計15種類とした。アノテーションの手順を示す樹形図を図4に、実際のアノテーションの例を表2に示す。

表1 属性値の組み合わせと対応するタグの種類: 1から4の数値は3.2.1節に対応している。

タグ	1		2		3		4		
	肯定	否定なし	未来	現在過去	有	無	事実	推論	
F1	✓		✓		✓		✓		
F2	✓				✓			✓	
F3	✓		✓			✓	✓		
F4	✓		✓			✓		✓	
P1	✓			✓			✓		
P2	✓			✓		✓		✓	
P3	✓			✓		✓	✓		
P4	✓			✓		✓		✓	
H1	✓			✓	✓		✓		
H2	✓			✓	✓			✓	
H3	✓			✓		✓	✓		
H4	✓			✓		✓		✓	
N	✓								
O		✓							

表2 対象記録へのアノテーションの例:

- はアノテーションにおいて出現しなかったことを示し、太字と下部線は疼痛表現を示す。

タグ	例文
F1	-
F2	-
F3	術後、 疼痛 や環境の変化により良眠が得られない可能性あり 術後 疼痛 やルートにより転倒リスク上昇するため注意必要 苦痛 症状出現時は医療者に伝えるよう説明し、早期対応できるようにする
F4	-
P1	ボース使用せず 疼痛 自制内 そんなに 痛 くはないの ももとの 左下 肢痛は軽快
P2	疼痛 に関してはコロナール効果あり 疼痛 あるが鎮痛剤は不要と コロナール内服し、その後 頭痛 の訴えなし
P3	チクチクお腹は 痛 いんだけどね その時も 肛門部 痛があった 痛み はNRS5/10 強い時
P4	ロキソニン処方だけでいけそう 寝る前に 痛み 止めください 疼痛 コントロールを図る
H1	-
H2	-
H3	mass(-)恥骨左側にあった 圧痛 消失 胸の 痛み はずっかりなくなりました 【現病歴】1ヶ月前くらいに右耳下を 痛 がっていた
H4	-
N	胸の方は 痛 くないですよ 現在 疼痛 や嘔気など苦痛症状なし 疼痛 があるわけではない
other	皮膚・皮膚光線レーザ科麻酔科・ 痛み センター) 痛 いのは苦手ですね 便が二日出なかつたら熱が出たりお腹が 痛 くなるって言われたの
除外	咽頭痛 (トラマール(100mg/A)0.25A+NS50ml50ml/h 疼痛 時頓用4時間以上あけて1日4回まで

3.4. 統計情報

表3に3章においてアノテーションを実施した学習データ(9,889件)のタグの分布を示す。実験においては出現頻度が低いタグにおいてはその学習が困難であると考え、除外した。そのため最終的に対象としたタグの種類は6種類(P1、P2、P3、P4、Nとother)である。また、学習データにおける1文あたりの平均単語数は22単語、標準偏差は51であった。

表3 学習データにおけるタグの種類と数

タグ	数	実験対象
F1, F2, F4, H1, H3, H4	0	
F3	12	
P1	113	✓
P2	27	✓
P3	410	✓
P4	96	✓
H3	15	
N	135	✓
other	21	✓

3.5. 実験

3.5.1. 固有表現抽出について

本研究では事象認識と事実性判定を固有表現抽出タスクとして扱う。固有表現抽出は入力文の各単語に対して付与されるタグを予測(図5)するものである。固有表現のタグ付けはIOB2形式(疼痛表現において開始をB、それ以外をI、疼痛表現以外は全てO)で実施した。

単語	左	下	肢	痛	を	訴	え	る
正解タグ	B	I	I	O	O	O	O	O

図5 疼痛表現における固有表現抽出の例

3.5.2. モデル

固有表現抽出にはBi-LSTM CRFを使用する。これは近年、高い精度を達成しているモデル⁷⁾⁸⁾⁹⁾であり、Bidirectional Long Short-Term Memory (Bi-LSTM)とConditional Random Fields (CRF)から構成される。本研究ではLampleらの研究⁷⁾に基づきモデルを構築した。そのモデルの概要を図6に、各層の詳細を3.5.3節から3.5.5節で述べる。

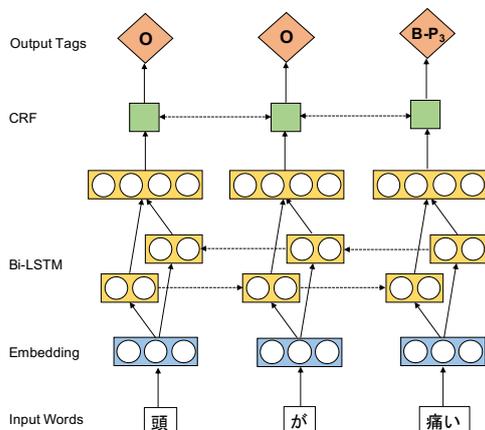


図6 Bi-LSTM-CRFのモデル概要

3.5.3. Embedding層

入力する単語系列を $w = (w_1, w_2, \dots, w_t, \dots, w_N)$ とし、 w_t は t 番目の単語とする。また単語系列に対応するタグ系列を $y = (y_1, y_2, \dots, y_N)$ とする。Embedding層から各単語の埋め込み表現を取得し、 $x = (x_1, x_2, \dots, x_N)$ とする。 x_t はそれぞれ D 次元のベクトルである。

Embedding層には3.1.1のStep2で用いた事前学習済みの分散表現(fastText)と近年の自然言語処理タスクにおいて高い精度を達成しているBidirectional Encoder Representations from Transformers (BERT)¹⁰⁾の2種類を使用した。BERTは京都大学の黒橋・河原研により配布³⁾されている事前学習済みモデル¹¹⁾を用いた。

3.5.4. Bi-LSTM層

Embedding層から取得した埋め込み表現 x をBi-directional Long Short-Term Memory (Bi-LSTM)に入力する。順方向のベクトル表現 \vec{h}_t と逆方向のベクトル表現 \overleftarrow{h}_t を取得し、これらを結合し、 h_t とする。Bi-LSTMを用いることで各単語の前後の文脈を考慮した表現を得ることができる。

$$\begin{aligned} \vec{h}_t &= \text{BiLSTM}(x_t) \\ \overleftarrow{h}_t &= \text{BiLSTM}(x_t) \\ h_t &= [\vec{h}_t; \overleftarrow{h}_t] \quad (; \text{は結合を示す}) \end{aligned}$$

3.5.4. CRF層

Conditional Random Fields (CRF)は出力タグ間の依存関係をモデル化⁷⁾⁸⁾⁹⁾することができる。Bi-LSTMのみでもタグ系列の予測は可能であるが、この場合は各単語に対して最も可能性が高いタグを出力する。一方、最終層にCRFを使用することで、タグ系列全体の遷移確率を考慮することができ、それにより x に対して尤もらしいタグ系列を得ることができる。学習時は以下の $\log P(y|X)$ を最大にするようにモデルのパラメータを学習する。

$$\text{Log } P(y|X) = s(X, y) - \log \text{add } s(X, \tilde{y})$$

$$s(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{t, y_i}$$

(\tilde{y} は系列 X における全ての可能なタグ系列、

A は遷移行列、 P は h_t を線形変換した行列を示す)

3.6 実験方法・評価方法

実験では入力文の最大単語数は126単語に設定し、最適化関数にAdamを使用し、Bi-LSTM層の隠層のサイズは100とした。またepoch数は10、ミニバッチのサイズは16で固定し、BERTのパラメータの学習率は $1e-5$ 、それ以外のパラメータは $1e-3$ とした。

評価では5分割交差検定を用いて、Precision、Recall、F値の平均値を算出する。Precision、Recall、F値は、1つの疼痛表現に対して完全に正しいタグを予測できた場合は正解、1つでも誤った予測をした場合は不正解とする。例えば「左 下肢 痛」のタグがB、I、Iであった時に、完全にB、I、Iと予測できれば正解とし、B、B、IやB、O、Oのように1つでも予測を誤った場合は不正解とする。

5. 結果

Embedding層にfastTextを使用した際の実験結果を表4に、BERTを使用した際の実験結果を表5に示す。

実験結果より、Precision、Recall、F値のいずれの平均値もBERTを使用した場合の方がfastTextを使用したときよりも高いことが確認された。

³⁾ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT> 日本語 Pretrained モデル

表 4 実験結果 (fastText)

	Precision	Recall	F 値
P1	53.92	37.91	43.96
P2	20.00	2.86	5.00
P3	67.35	83.00	74.26
P4	43.45	23.53	30.40
N	68.42	55.84	61.23
other	40.00	10.00	16.00
Average	48.86	35.52	38.48

表 5 実験結果 (BERT)

	Precision	Recall	F 値
P1	60.11	52.24	53.01
P2	11.67	5.36	7.27
P3	82.07	85.54	83.44
P4	40.96	39.62	40.02
N	80.74	95.10	87.12
other	80.00	35.00	44.00
Average	59.26	52.14	52.48

6. 考察

6.1 誤り分析

表 6 に表 5 の実験結果において予測を誤った文の例を示す。

No.1 は「疼痛はあまり強くない」とあり、完全な否定ではないことから正解タグは<P1: 事実としての自制内の疼痛>であるが、<N: 否定された疼痛>と予測した例である。これは「ない」という否定表現の影響が強く、「あまり」という表現を捉えることができなかつたためであると考えられる。

No.2 は内服後に頭痛の訴えがなくなっているため、処置により疼痛がコントロールされたと考えられる。そのため正解タグは<P2: 推論としての自制内の疼痛>であるが、<N: 否定された疼痛>と予測した例である。これは No.2 と同様に、後ろの否定表現に大きく影響されたためであると考えられる。

No.3 は「疼痛軽減しなかつた」とあるので、正解タグは<P3: 事実としての疼痛>であるが、<P1: 事実としての自制内の疼痛>と予測した例である。これは後ろの「軽減」という<P1: 事実としての自制内の疼痛>にて頻出する単語の影響を受けたことが原因であると考えられる。

No.4 は「腰痛改善エクササイズを視聴」という処置から腰痛があることが推論されるため<P4: 推論としての疼痛>であるが、<P1: 事実としての自制内の疼痛>と予測した例である。これは No.3 と同様に<P1: 事実としての自制内の疼痛>にて頻出する「改善」という単語の影響を受けたためであると考えられる。

No.5 は「痛いのは苦手」とあり、患者の一般的知識であるため正解タグは<other>であるが、<P3: 事実としての疼痛>と予測した例である。患者に生じている疼痛とするか、患者の一般的知識とするかの判断は人間でも困難であるため、それらを判断する知識を持たない機械学習では困難であると考えられる。

No.6 は「疼痛訴えなし」とあるので、正解タグは<N: 否定された疼痛>であるが、<P3: 事実としての疼痛>と予測した例である。これは形態素解析において、「なし」を「な」と「し」に分割してしまったこと、つまり形態素解析誤りが原因ではないかと思われる。形態素解析誤りは日本語の言語処理において避けることのできない課題であるため、今後の対応が必要で

ある。

6.2 分類性能に関する考察

BERT を用いた場合の F 値の平均値は fastText を用いた場合より 14 ポイント大きいことが確認された。BERT は fastText とは異なり、大規模なコーパスで事前学習しており、加えて文脈を考慮したベクトル表現を得ることができ、これが性能の差に寄与していると考えられる。

実験において、人間でアノテーションが比較的容易であったタグ (<P3: 事実としての疼痛>と<N: 否定された疼痛>)の性能は F 値で約 80 程度と一定の性能が得られている。しかし、二重否定などの複雑な文脈から予測する必要のあるタグ (<P1: 事実としての自制内の疼痛>)や人間でもアノテーション時に判断が分かれることが多かつたタグ、医療知識が必要なタグ (<P2: 推論としての自制内の疼痛>、<P4: 推論としての疼痛>や<other>)においては、F 値が低いことが確認された。このことから、人間で判断が難しい場合や知識が必要な場合は機械学習による予測は困難であることが考えられる。そのため、アノテーションにおいては作業員間で判断が分かれなように基準の更なる明確化が必要である。また医療知識をモデルに組み込む方法についても検討が必要である。

6. 結論

本研究では疼痛表現の事象認識と事実性判定の実現に向けた診療記録や看護記録へのアノテーション方法を提示し、正解ラベル付きのコーパスを作成した。このコーパスを用いて自然言語処理技術に基づく事象認識と事実性判定を行う抽出器を構築し、精度を評価した結果、アノテーションが容易な表現の抽出性能は高く、アノテーションが困難な表現の抽出性能は低いことが確認された。これは医療文書の実事性判定におけるアノテーション基準の重要を示しており、アノテーション基準の更なる明確化が必要であることが考えられる。

今後は、アノテーションをさらに進め、学習データを増やし、実験対象とならなかつたタグの実事性判定の性能について検討を行う。また、疼痛が生じた人体部位の情報についてもアノテーションを行い、疼痛とそれが生じた人体部位を同定する関係抽出タスクについても取り組む予定である。また本研究では検索クエリを含む診療記録や看護記録のみを対象としているため、検索クエリに含まれない疼痛表現がどの程度存在し、これらがアノテーション基準や実験結果などにどのような影響を与えるのかは定かではない。そのため、これについては今後、調査・検討を行う予定である。

謝辞

本研究は JST さきがけ (JPMJPR1654) の支援を受けたものである。

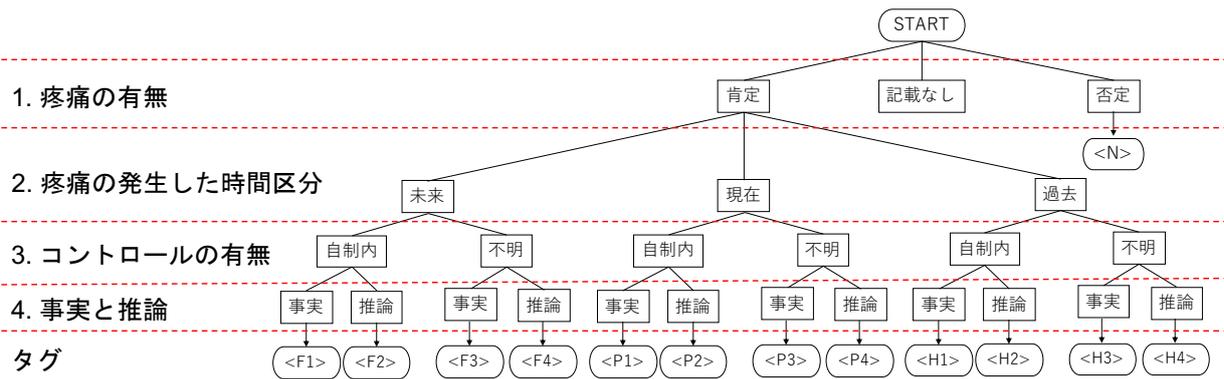


図4 アノテーション手順の樹形図

表6 予測結果の例: #はサブワードを示し、文中のタグは予測結果を示す。

No.	正解	予測
1	<P1>	表情からは<N>疼痛</N>はあまり強くないように見受けられる
2	<P2>	内##服しその後は<N>頭痛</N>の訴えなし
3	<P3>	後も<P1>腹##痛</P1>軽減しなかったとのこと
4	<P4>	VTRにて「<P1>腰痛</P1>改善エク##サ##サイ##ズ」について視聴し学習された
5	<other>	<P3>痛##い</P3>のは苦手##です
6	<N>	本##日起##居動作時の<P3>疼痛</P3>訴えなし(一部省略)

参考文献

- 1) 荒牧英治, 若宮翔子, 矢野憲, 永井宥之, 岡久太郎, 伊藤薫. 病名アノテーションが付与された医療テキストコーパスの構築. 自然言語処理「言語処理の応用システム」特集号(技術資料), 2017; 25(1), 119-152.
- 2) Xiaojun Ma, Emiko Shinohara, Hao Han, Masamichi Ishii, Takeshi Imai, Kazuhiko Ohe. Extracting Information on Lifestyle Issues from Clinical Narratives in EHR. 医療情報学, 2018; 37(6), 313-321.
- 3) P. Ding, X. Zhou, X. Zhang, J. Wang and Z. Lei, An Attentive Neural Sequence Labeling Model for Adverse Drug Reactions Mentions Extraction, in IEEE Access, 2018; vol. 6, pp. 73305-73315.
- 4) Ozlem Uzuner, Ira Goldstein, Yuan Luo, Isaac Kohane. Identifying patient smoking status from medical discharge records. Journal of the American Medical Informatics Association, 2008; Volume 15, 14-24.
- 5) Rina Kagawa, Emiko Shinohara, Takeshi Imai, Yoshimasa Kawazoe, Kazuhiko Ohe. Bias of Inaccurate Disease Mentions in Electronic Health Record-based Phenotyping. International Journal of Medical Informatics. 2019; Volume 124, 90-96.
- 6) Joulin Armand, Grave Edouard, Bojanowski Piotr, Mikolov Toma. Bag of Tricks for Efficient Text Classification. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017; Volume 2, 427-431.
- 7) Zhiheng Huang, Wei Xu, Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991.
- 8) Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016; 260-270.
- 9) Xuezhe Ma, Eduard Hovy. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics 2016; 1064-1074.
- 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. 2019; 4171-4186.
- 11) 柴田知秀, 河原大輔, 黒橋禎夫. BERTによる日本語構文解析の精度向上. 言語処理学会 第25回年次大会 2019; 205-208.