
一般口演B

[OB13] 一般口演 B

2020年6月6日(土) 14:50 ~ 15:20 第2会場 (Zoom)

[OB13-01] 診療記録で事前学習した BERT による疼痛表現の抽出 Pain Expressions Extraction with pre-training BERT based on Clinical Note

*柴田 大作¹、河添 悦昌²、嶋本 公德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)

*Daisaku Shibata¹, Yoshimasa Kawazoe², Kiminori Shimamoto², Emiko Shinohara², Eiji Aramaki¹ (1. Nara Institute of Science and Technology, 2. Graduate School of Medicine, The University Tokyo)

診療記録で事前学習した BERT による疼痛表現の抽出

柴田 大作^{*1}, 河添 悦昌^{*2}, 嶋本 公德^{*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

Pain Expressions Extraction with pre-training BERT based on Clinical Note

Daisaku Shibata^{*1}, Yoshimasa Kawazoe^{*2}, Kiminori Shimamoto^{*2},

Emiko Shinohara^{*2}, Eiji Aramaki^{*2}

^{*1} Nara Institute of Science and Technology

^{*2} Graduate School of Medicine, The University Tokyo

【背景】患者の疼痛に関する情報は診療記録にフリーテキストとして記載されるため抽出が困難である。これを精度良く行うことができれば、医療者間での情報共有に役立つことが期待される。【目的】診療記録からの疼痛表現の抽出を、疼痛表現の事象認識と事実性判定のタスクとして定義し、機械学習による抽出精度を評価する。【方法】東大病院の診療記録 20,000 文を対象とし、疼痛表現を手でアノテーションしたものを実験材料とした。機械学習は Bi-LSTM CRF を使用し、単語の分散表現は診療記録で事前学習した BERT から取得した。日本語 Wikipedia で事前学習した BERT、fastText との比較も行った。【結果】診療記録で事前学習した BERT の F 値が平均で 56.4 と最も高かった。【考察】診療記録からの疼痛表現の抽出において、文脈を考慮する BERT の有効性と事前学習に用いる文書の選択の重要性が示唆された。

キーワード 自然言語処理、BERT、情報抽出、診療記録、疼痛表現

1. はじめに

患者に生じている疼痛に関する情報の多くは診療記録などにフリーテキストとして記載されるため抽出が困難であり、課題の一つとなっている。近年、自然言語処理において、大規模なテキストデータから単語の分散表現を事前学習し、それを情報抽出などのタスクに使用することで精度が向上することが報告されている。中でも BERT (Bidirectional Encoder Representations from Transformers) は多くのタスクで高い精度を達成しており、その有効性が注目されている。日本語で利用可能な BERT は、京都大学より公開されているモデル (以下、K-BERT) や東北大学より公開されているモデルなどがあるが、これらは日本語 Wikipedia により学習されたモデルであり、ドメインが異なる診療記録などの医療文書に使用することは適切ではない可能性がある。そこで本研究では、日本語診療記録で事前学習を行った University of Tokyo Hospital BERT (以下、UTH-BERT) を用いた診療記録からの疼痛表現の抽出を行う。また UTH-BERT と K-BERT、従来手法である fastText の性能を比較し、事前学習に用いるデータの重要性について調査する。

2. 方法

1) タスク設定

疼痛表現の抽出を疼痛表現の事象認識と事実性判定を行うタスクとして定義した。Fig.1 のように事象認識は文に出現する疼痛表現 (痛いや腹痛など) を抽出し、事実性判定は疼痛表現の意味合い (肯定や否定など) を分類するものである。

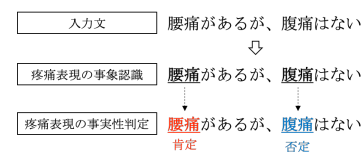


Fig.1 疼痛表現の抽出

2) 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 月に作成された診療記録 297,536 件から疼痛に関連する単語を含む診療記録を 164,444 件取得し、これらの記録を読点で文分割 (1,778,251 文) し、その中からランダムに 20,000 文抽出したものを実験材料として使用した。

3) 正解ラベルとなるタグの分類基準

疼痛表現は患者に生じている、否定されている疼痛だけでなく、生じているがコントロールされており問題となっていない疼痛、過去に生じた疼痛

や治療などから疼痛の存在が推論される場合など多様な事実性があることがわかった。そのため本研究では以下の4つの基準により正解ラベルとなるタグを決定した。

- 疼痛の有無 (3通り)
疼痛が肯定されている (事実として明確に記載されていないが、推論によって疼痛があると判断される場合も含む)、否定されている、記載なしの3通りを分類する。
- 疼痛の発生した時間区分 (3通り)
疼痛が肯定されている場合、それが発生した時間が未来、現在、過去のいずれであるかを3通りに分類する。
- コントロールの有無 (2通り)
疼痛が肯定されている場合、それがコントロールされているか否かを分類する。
- 事実か医療知識による推論か (2通り)
疼痛が肯定されている場合、それが事実として記載されているのか、医療知識に基づいて疼痛が推論されるのかを分類する。

4種類の基準の組み合わせは36通りであるが、疼痛の有無が否定もしくは記載なしの場合はそれ以外の基準を考慮しない。そのため正解ラベルとなるタグの総数は14通りとなる。また、一般的知識として記載される疼痛など患者に生じているわけではない疼痛表現に対しては<other>タグを設け、タグは合計で15種類とした。

4) アノテーション

疼痛表現に対し、正解ラベルのタグを手で付与した。例えば、「腹痛があるわけではない」という文は、腹痛が否定されているので、疼痛表現である「腹痛」に否定を示す<N>タグを付与し、「<N>腹痛</N>があるわけではない」とする。

5) 実験

疼痛表現の抽出を固有表現抽出タスクとし、モデルには Bi-directional Long Short-Term Memory Conditional Random Fieldsを使用した。単語の分散表現は UTH-BERT、K-BERT と診療記録から学習した fastText から取得した。評価は5分割交差検証における F 値の平均値で行なった。F 値は入力文中の1つの疼痛表現に対して付与されたタグの位置が正しく、かつタグの種類

を正確に分類できた場合を正解とし、それ以外を不正解として算出した。UTH-BERT は診療記録などのカルテ記載約1.2億文を、K-BERT は日本語 Wikipedia1,800 万文を学習データとし、事前学習を行ったモデルである。

3. 結果

実験結果を Table.1 に示す。実験において、出現頻度の低かったタグは除外した。

Table.1 実験結果

タグ	意味	UTH-BERT	K-BERT	fastText	平均タグ数
P1	事実のコントロール	63.7	55.4	49.3	45
P2	推論のコントロール	20.8	6.5	10.7	8
P3	事実の疼痛	81.9	76.0	73.8	160
P4	推論の疼痛	39.2	25.9	22.7	36
N	否定の疼痛	77.2	71.1	72.2	45
other	その他	55.3	49.1	46.5	13
平均値		56.4	47.3	45.9	-

4. 考察

全てのタグにおいて UTH-BERT を用いたモデルの F 値が高いことが確認された。UTH-BERT を用いることで、従来手法の fastText と比較して平均で 10.5 ポイント、K-BERT と比較して 9.1 ポイントの精度の改善が確認された。K-BERT は日本語 Wikipedia で事前学習しており、本研究の実験材料とは出現する単語、文体が異なるため UTH-BERT に比べ、精度が低下したと考えられる。また、fastText は文脈を考慮することができないため、文脈が重要である事実性判定で BERT よりも精度が低いことは合理的と考えられた。本研究で行った、診療記録における疼痛表現の事象認識と事実性判定のタスクでは、事前学習に用いる文書の種類が精度に影響を与える結果となり、事前学習に用いる文書の重要性が確認された。

5. 結語

UTH-BERT、K-BERT と fastText による疼痛表現の抽出を行った。その結果、UTH-BERT を用いたモデルの F 値が最も高く、BERT による従来手法からの精度改善の可能性と BERT の事前学習に用いるテキストの重要性が示唆された。これを用いたアプリケーションを開発することで、医療者への業務支援が期待される。

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the NAACL. 2019; 4171-4186.