ポスター
# ポスター12
# 病院情報システム４・電子カルテ
2018年11月24日(土) 10:00 〜 11:00 K会場(ポスター、HyperDemo) (2F 多目的ホール)

[3-K-2-3] 頭痛問診情報の対話型収集システムのプロトタイプ開発
○久保 晴輝[1], 井田 有亮[2], 大江 和彦[1,2] （1.東京大学大学院医学系研究科, 2.東京大学医学部附属病院）

初診患者の主訴について問診によって詳細な臨床情報を収集することは初期診断にとって重要である。診断支援システムによりこれをある程度自動化すれば、非対面診療や診察前の事前問診における情報収集が効率良く行えるが、音声認識と発話機能を備えた対話型の問診情報システムはまだ実用化できていない。そこで、本研究では患者の訴えのみが主な情報源となる痛みとして、頭痛を対象とし、頭痛に関する発言を対話により聞き出して、その詳細な臨床情報の収集を行うアプリケーションシステムの開発を目的とした。頭痛に関する臨床情報の属性を性状、部位、発症及び持続時間、増悪傾向、既往、痛みの表現、契機及び随伴症状の7種に区分し、それぞれについて会話における患者表現を含む文を作成し、さらに研究協力者からも列挙してもらう方法で表現データセット(534件)を作成した。次にこの表現データセットに対して BIO法によりアノテーションを行なった。このアノテーションされたデータセットを使用し、機械学習の手法の一種である Conditional Random Fields(CRF)を用いて、患者の頭痛に関する会話表現から前記7種の頭痛属性情報を取得するアルゴリズムを開発した。開発された学習済みアルゴリズムを評価するため、学習に用いなかった会話表現を入力として前記7種の頭痛属性情報をどの程度取得できるかを K=5の K-fold法により評価した。その結果精度0.72、再現率0.74となり、改善が必要な結果となった。学習データセットの件数が圧倒的に不足していると考えられるため、クラウドソーシングの手法により規模の大きい学習データセットを作成して性能を向上させたいと考えている。こうして一定程度性能を向上させた後に、対話機能を持つ人型ロボットやスマートスピーカー、音声認識機能を持つスマートホンやタブレットPCに実装することで遠隔医療で活用できる可能性がある。また胸痛、腰痛、腹痛などといった様々な痛み症状の情報収集に発展させることを考えている。

頭痛問診情報の対話型収集システムのプロトタイプ開発

久保 晴輝[*1]、井田 有亮[*2]、大江 和彦[*1,2]
*1 東京大学大学院医学系研究科、*2 東京大学医学部附属病院

Development of a prototype of dialogue-based information collection system for headache diagnosis

Kubo Haruki[*1], Yusuke Ida[*2], Kazuhiko Ohe[*1,2]
*1 Graduate School of Medicine, The University of Tokyo, *2 The University of Tokyo Hospital

Although automation of collection of clinical information based on the interview of patients' complaints using diagnostic support system is expected to make information collection more efficient, no such dialogue-based information collection system equipped with vocal recognition ability and speech function has been developed. Here, we aimed to establish a system for the collection of clinical information based on dialogues regarding the complaints of headache.

We divided clinical information regarding headache symptoms into 7 categories and created sentences that include patients' complaints, to produce a dataset (534 issues). We next annotated this dataset using IOB tags. Using this annotated dataset and conditional random fields, we newly developed an algorism to obtain information regarding the above-mentioned headache symptoms from the patients' dialogues. To evaluate the developed algorism, we used k-fold cross-validation(k=5) to examine whether it can obtain information from the patients' dialogues that were not used for machine learning. The results indicated precision of 0.72 and recall of 0.74, which requires further improvements. We plan to improve the algorism by producing a large-scale learning dataset using crowdsourcing. We expect that it can be utilized in telemedicine in future by being equipped by human robot with interactive ability, smart speaker, smartphone and/or tablet PC with vocal recognition ability.

Keywords: computer-assisted diagnosis, natural language processing, headache

1. 緒論

日本の外来診療において、待ち時間が長いことは以前から問題として挙げられている。[1)]外来診療は受付の後問診、検査、診断、処置といった流れで進むが、その中でも問診による情報収集が多くの時間を占めている。[2)3)]問診はその後の臨床推論を行うために必要な臨床情報を収集するための段階であり、特に初診患者においてはその主訴について詳細に聞き出す必要があるため、多くの時間を必要としている。[4)]

問診に要する時間を短縮するために、来院前や医師による問診を行う前に、問診票を記入させるなどの対策を取っている医療機関も存在する。[5)]しかし、選択式の質問や症状の記載などだけでは収集が難しい情報がある。例として、痛みはその強度に関しては定量的な評価を行う手法が研究されてきているものの、痛みの性状や増悪傾向などは測定が難しく、対話形式の問診による臨床情報収集の重要性が高い。[6)7)]

対話形式の問診を自動で行う診断支援システムを構築することができれば、今後発展していくであろう非対面診療や、医療資源の有効活用、外来診療の待ち時間の削減などに役立てることができるだろう。しかし、現在音声認識と発話機能を備えた対話型の問診情報システムはまだ実用化できていない。

2. 目的

本研究では、患者の訴えのみが主な情報源となる痛みの中でも、緊急性の高い疾患が多く存在する頭痛を対象とし、頭痛に関する発言を対話により聞き出して、発言の中に存在する臨床情報を含む表現を抽出することで、詳細な臨床情報の収集を行うアプリケーションシステムのプロトタイプ開発を目的とした。

3. 方法とシステム概要

医師の協力により、頭痛に関する臨床情報の属性を性状、部位、発症および持続時間、増悪傾向、既往、痛みの表現、契機および随伴症状の７種に区分し、患者の発言からそれら臨床情報を含む表現を機械学習により抽出するシステムを作成した。

まずシステム開発および性能試験に用いるデータとして、頭痛に関する臨床情報を含む文を作成し、同様に研究協力者からも列挙してもらうことにより、頭痛に関する534件の文例から成るデータセットを作成した。対話を前提としたシステムの作成のため、文を作成する際には会話的表現を用いることとした。また、「資金繰りで頭が痛い」などの比喩表現で用いられる、疾患によらない頭痛に関する文は除外した。

次にこのデータセット中の各文例に対して形態素解析を行った（図 1）。解析には形態素解析ソフトである MeCab を使用し、解析用の辞書として mecab-ipadic-NEologd を用いた。[8)]

続いて、形態素解析を行った各文例に対して、IOB（Inside-Outside-Beginning）タグ形式を用いて形態素単位でアノテーションを行なった（図 1）。IOB タグ形式は、抽出を行いたい表現の始まりとなる形態素には B（Begin）タグを、表現の中間から終わりまでの形態素には I（Inside）タグをつけ、それ以外の形態素に対しては O（Outside）タグを付与する手法である。[9)]これを用いることで、文中から表現を抽出することを系列ラベリング問題として解くことができるようになり、機械学習による処理を可能とするものである。臨床情報を含む表現が存在した場合、その臨床情報が前述の頭痛に関する 7 種の属性（性状、部位、発症及び持続時間、増悪傾向、既往、痛みの表現、契機及び随伴症状）のどれに属するかを機械学習で判断できるようにするために、「B-性状」「I-部位」タグ

のように属性情報を含んだ形でのアノテーションを行った。臨床情報の属性が 7 種あり、各々に B タグと I タグが存在するため、タグの種類は O タグを加えて 15 種類となった。

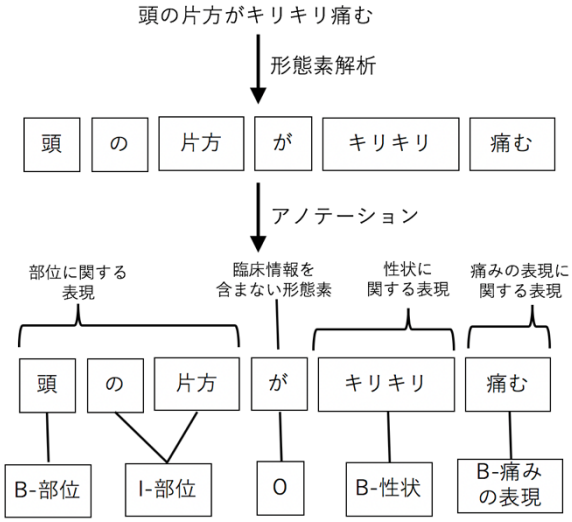

図 1 形態素解析とアノテーション

このアノテーションされたデータセットを学習データセットとして、系列ラベリング問題を解くための機械学習の手法の一種である Conditional Random Fields（CRF）を用いた学習済みアルゴリズムを作成した。[10]この学習済みアルゴリズムにアノテーションされていない状態の患者の頭痛に関する会話表現を入力することで、学習済みアルゴリズムが予測したタグによる形態素単位でのアノテーションが行われた状態の会話表現が出力される（図 2）。CRF の特徴量には、予測を行う対象の形態素を含んでその前 2 形態素から後ろ 2 形態素の表層系、形態素を構成している文字種、品詞および品詞細分類を使用した。

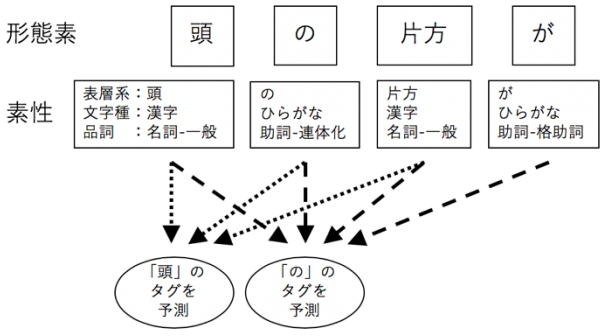

図 2 Conditional random fields による予測

作成した学習済みアルゴリズムを評価するため、学習に用いなかったアノテーション済みの文例を入力として、どの程度人間によるアノテーションに近い結果が出力できるかを K=5 の K-fold 法により評価した。評価には適合率（Precision）、再現率（Recall）と、（2×適合率×再現率）／（適合率＋再現率）で定義される F 値（f1-score）を使用した。

前述の学習済みアルゴリズム単体では対話は不可能であるため、対話デバイスとして SoftBank Robotics 社とアルデバラン社により共同開発された人形ロボットである Pepper を使用した。Pepper で問診における質問の発話と、患者の発言の音声認識およびテキスト化を行い、発言内容のテキストを学習済みアルゴリズムに処理させることで、会話による臨床情報の収集を実装した。また、処理を行った患者の発言のみでは臨床情報が不足している場合、不足している属性に関しての質問を行うことで、さらなる情報を収集するアルゴリズムを作成した。

4. 結果および評価

534 件の文例に対し、形態素単位でのアノテーションを行った結果を表 1 に示す。臨床情報を含む表現の形態素の数は合計で 5330 となった。B タグを付与された形態素は 1828、I タグが付与された形態素はおよそ 2 倍の 3502 となった。属性ごとでも I タグの付与された形態素の数は B タグの二倍以上である場合が多いが、部位と痛みの表現の属性に関しては B タグと I タグで形態素数が近かった。属性ごとにデータセットに存在する形態素数に差があり、増悪傾向、既往、発症及び持続時間に属する形態素は比較的少ない結果となった。

表 2 に、K-fold 法により得られた属性ごとの精度を示す。属性全体での精度の平均値は、適合率が 0.72、再現率が 0.74、F 値が 0.73 となった。属性ごとに精度に偏りがあり、増悪傾向、既往の属性の精度はかなり低い結果となった。一方で、部位と痛みの表現の属性は比較的精度が高く、0.8 程度の F 値となった。どの属性に関しても、適合率と再現率に大きな差は存在しなかった。

表 1 属性ごとの形態素数

属性名	B タグ	I タグ	合計
増悪傾向	68	221	289
既往	61	118	179
発症及び持続時間	131	265	396
性状	413	941	1354
部位	433	468	901
契機及び随伴症状	228	911	1139
痛みの表現	494	578	1072
合計	1828	3502	5330

5. 考察

ほとんどの属性において、I タグをつけられた形態素数が B タグをつけられた形態素数の 2 倍程度であることから、臨床情報を含む表現は平均的に 3 つ程度の形態素からなっていると推測される。一方で、部位、痛みの表現の属性は B タグと I タグの形態素の数が近いことから、臨床情報を含む表現は 2 形態素程度で構成されていると思われる。

表 2 属性ごとの精度

属性名	適合率	再現率	F 値
増悪傾向			
B タグ	0.60	0.46	0.52
I タグ	0.42	0.42	0.42
既往			
B タグ	0.59	0.36	0.45
I タグ	0.40	0.32	0.36
発症及び持続時間			
B タグ	0.73	0.59	0.65
I タグ	0.62	0.62	0.62
性状			
B タグ	0.78	0.73	0.75
I タグ	0.76	0.83	0.79
部位			
B タグ	0.87	0.87	0.87
I タグ	0.84	0.81	0.83
契機及び随伴症状			
B タグ	0.62	0.59	0.61
I タグ	0.56	0.68	0.61
痛みの表現			
B タグ	0.96	0.94	0.95
I タグ	0.70	0.80	0.75
合計	0.72	0.74	0.73

精度の平均値は適合率 0.75、再現率 0.73、F 値が 0.74 となり、改善が必要な値となった。特に精度が低かった増悪傾向と既往に関しては、データセット中に存在する形態素数が少ないことが大きな影響を及ぼしていると思われる。精度が高かった部位、痛みの表現の属性に関しては、表現を構成する形態素数が少ないことと、データセット中に存在する形態素数が多いこと、そして似たような表現が異なる文例で繰り返し出現することが要因となって精度が高くなっているものと思われる。精度を向上させる方法は複数考えられるが、まずはデータ数を増やすために、クラウドソーシングによって頭痛に関する会話表現を用いた文例を大量に収集し、精度を上げることを考えている。

上記で論じている精度は学習済みアルゴリズム単独の精度であり、音声認識や質問といった対話の要素を含んでいないものである。実際に対話を行った場合、患者の発言中に言い間違いやフィラーなどといったノイズが入ってくるほか、音声認識の失敗によるノイズが加わることが想定されるため、対話システム全体としての精度はさらに下がることが予想される。また、不足している情報に対してどのように質問を行うかや、どのように自動問診の結果を出力していくかに関しても考慮する必要がある。システム全体としての性能評価や改善を行っていく必要があるだろう。

将来的にこのシステム全体の性能が向上させられた場合、対話機能を持つ人型ロボットやスマートスピーカー、音声認識機能を持つスマートホンやタブレット PC に実装することで遠隔医療で活用できる可能性がある。また、臨床情報の基本構造が同じであり、属性が流用できるとすれば、胸痛、腰痛、腹痛などといった様々な痛み症状の情報収集に発展させられる可能性があるだろう。

6. 結論

本研究では、頭痛に関する患者の発言を対象として、対話により発言を引き出し、その発言から臨床情報を含む表現を抽出するアプリケーションシステムのプロトタイプを作成した。システム中で使用する、テキスト化された発言から臨床情報を抽出する機械学習のアルゴリズムの精度評価を行ったところ、適合率が 0.72、再現率が 0.74、F 値が 0.73 となり、改善が必要な結果となった。また、システム全体としての精度評価と改良も行っていく必要がある。

7. 文献

1) 厚生労働省大臣官房統計情報部. 受療行動調査の概要. 厚生労働省. 2010. [http://www.mhlw.go.jp/toukei/saikin/hw/jyuryo/09/index.html (cited 2018-09-05)]

2) Steven A. Cole. メディカルインタビュー―三つの機能モデルによるアプローチ 第 2 版. メディカル・サイエンス・インターナショナル, 2003.

3) H. Harold Friedman 著, 日野原重明訳. PO 臨床診断マニュアル 第 7 版. メディカル・サイエンス・インターナショナル, 2002.

4) 徳永誠, 中根惟武. 日本外来患者の受診状況ごとに検討した待ち時間調査. 医療マネジメント学会雑誌, 2006 ; 7 ; 3 : 434-437.

5) 稲熊良仁, 岡山雅信, 古城隆雄ら. 総合診療科の初診外来問診票で何が問われているか. 日本プライマリ・ケア連合学会誌 2012 ; 35 ; 1 : 12-16.

6) Miyashita M, Matoba K, Sasahara T, et al. Reliability and validity of Japanese version STAS（STAS-J）. Palliat Support Care 2004; 2: 379-85

7) 小川節郎. 痛みをモニターする. 日本臨床麻酔学会誌, 2011 ; 31 ; 3 : 496-500.

8) 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. 情報処理学会研究報告自然言語処理（NL）, 2004 ; 47 : 89-96

9) Lance A. Ramshaw, Mitchell P. Marcus. Text Chunking using Transformation-Based Learning. ACL Third Workshop on Very Large Corpora, 1995 : 82-94

10) J. Lafferty, A. McCallum, and F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In Proc. of ICML, 2001 : 282-289