

# 医療テキスト解析のための事実性判定と融合した病名表現認識器

矢野 憲          若宮 翔子          荒牧 英治

奈良先端科学技術大学院大学  
 {yanoken, wakamiya, aramaki}@is.naist.jp

## 1. はじめに

電子カルテをはじめとする医療テキストには、患者情報があまねく記録されていると考えられるものの計算機で扱いが困難な非文法的かつ断片化した表現が多く含まれており、利活用は困難であった。このため、自然言語処理を用いた医療テキスト処理が注目されている[1]。すでに、英語言語での医療文書の解析には、MAYO Clinicの語彙サーバー、MedLEE [2]、cTAKE [3]など、いくつかのツールが利用可能になっており、広く利用されている。一方で、日本語テキストを扱う標準的な言語処理ツールは未だ存在しない。

本研究では、日本語テキストを処理するために、日本語による医療用自然言語処理ツールMedEx/Jを提案する。これは以下の2つの処理を行う。

- **事象認識 (ER)** : 医療テキストにおける病名および疾患名を識別する。これは、一般的な固有表現認識タスクと類似した処理である。以降、この処理を**ER (Entity Recognition)**とも呼ぶ。
- **陽性/陰性 (P/N) 分類** : テキストには、予測、疑い、可能性、否定など様々なモダリティの事象が含まれている。これらは大別すると、現在の疾患(陽性所見)と、それ以外の所見(陰性所見)に分類できる。これらを区別する必要がある。以降、この処理を**P/N分類**と呼ぶ。

医療テキストには多様な陰性所見が存在するため、経験則に基づくアプローチでのP/N分類が困難であり、機械学習を用いた手法が数多く提案されている[4-7]。これまでの手法に比べて、提案手法は、以下の2つの重要な違いがある。

- 文字ベースの処理
- 2つの処理 (ERとP/N分類) の融合

以下がそれぞれの処理の内容である。

- (1) **文字ベースの処理** : 従来のNLPの研究では、単語(形態素)を最小単位とみなすものが多い。しかし、医療テキストは、長い複雑な複合名詞(「傍大動脈リンパ節郭清」など)や一般には用いられない専門的なひらがな名詞(例えば、「びまん性」)が多く出現し、しばしば形態素解析の

### (a) 入力テキスト

腫瘍は肝細胞癌ではなく肝の孤立性形質細胞腫と診断された。

### (b) 文字ベース系列ラベリング

腫	瘍	は	肝	細胞	癌	で	は	な	く	孤	立	性	形	質	細胞	腫	と	診	断	さ	れ	た
O	O	O	B-	I-	I-	I-	O	O	O	B-	I-	I-	I-	I-	I-	I-	O	O	O	O	O	O
			N	N	N					P	P	P	P	P	P							

### (c) 単語ベース系列ラベリング

腫	瘍	は	肝	細胞	癌	で	は	な	く	孤	立	性	形	質	細胞	腫	と	診	断	さ	れ	た
O	O	B-N	I-N	I-N	O	O	O	B-P	I-P	I-P	I-P					O	O			O	O	O

### (d) タグ付き出力テキスト

腫瘍は<N>肝細胞癌</N>ではなく肝の<P>孤立性形質細胞腫</P>と診断された。

図1: 系列ラベリングにおけるシーケンス表現

誤りにつながる。したがって、本研究では、形態素解析処理を行わない文字ベースの方法を採用する。

- (2) **2つの処理の融合** : 通常、事象のP/N分類は、ERの後に適用される。しかし、P/N分類に必要な情報はERで必要な情報と重複する部分も多い。例えば、「～が認められる」「～が認められない」は、ともに病名出現の大きな手がかりであるとともに、P/N分類の手がかりにもなる。このため、本研究では、ERとP/N分類の2タスクを1つに融合する。

提案手法により形態素解析などの追加システムなしでMedEx/Jを構築することができる。また、1つのシーケンシャルなラベリングのもとで2つの処理を行うため、処理速度や実装のしやすさの点でも利点を持つ。

実験結果では、提案手法の精度は、(1) ER ( $F_{\beta=1} = 0.93$ ) および (2) P/N分類 ( $F_{\beta=1} = 0.91$ ) とともに高い精度を示し、提案手法の有効性を確認した。

## 2. 医療用自然言語処理ツール MedEx/J の構築

### 2.1 コーパス

本研究では、NTCIRの共有タスクデータ[8]と互換性の

ある医療テキストデータセットを用いる。これらのデータでは、症状および診断に関係する表現に、“<p></p>”（陽性の場合。以降、<p>タグと呼ぶ）または“<n></n>”（陰性の場合。以降、<n>タグと呼ぶ）が付与されている。本研究では、468 症例（計 9,286 文）の日本語テキストを用いる。

## 2.2 文字ベースのラベリング

本研究で提案する MedEx/Jの入力は文であり、出力は<p>タグや<n>タグが付いた文である。この出力は、(1) ERと (2) P/N分類の2つの処理の結果であり、処理(1)の結果は、タグ付けされた範囲として、処理(2)の結果は、タグのタイプ (<p>タグまたは<n>タグ)として表される。

ERは、一般の固有表現認識の手法にならない入力テキストシーケンス上で開始 (B), 内側 (I) および外側 (O) という3種類のラベルを推定する系列ラベリングタスクとして定義する。ただし、ラベル付けの単位は、文字である。図1に提案手法(文字ベース)と既存手法(単語ベース)の系列ラベリングの違いを示す。

通常、単語ベースのアプローチでは、品詞や原型などの単語単位の情報を特徴として使用しているが、文字ベースのアプローチではそれらの特徴を使用できない。そのかわりに、ひらがなやカタカナ、漢字といった文字種の情報を用いる。

図2は、文字ベースのアプローチ(図2(a))と単語ベースのアプローチ(図2(b))にて、それぞれ使用される素性関数を生成するためのテンプレートを示している。文字ベースのCRFテンプレートにある「Char」と「Char type」はそれぞれ表面文字と文字タイプを表す。文字タイプは漢字、ひらがな、カタカナ、英数字である。「N-gram」は、ユニグラム、バイグラムまたはトライグラムのうちのどのnグラムを使用するかを指定する。各設定のウィンドウサイズは、予備実験にて調整した。

## 2.3 融合法: ER+P/N 分類

通常、P/N分類はERの結果に適用される。これに対し、提案手法では、陽性と陰性をそれぞれ独立に扱うことにより、これらの2つの問題を同時に解決する。

コーパス中に現れる陰性の述語例を以下に示す。

- <N>悪性腫瘍</N>が**疑われた**
- <N>発熱</N>は**認めなかった**
- <N>肺結核</N>は**否定的だった**

陰性は、否定的または疑わしい述語句に関連し、患者とは無関係、または、一般的な文脈で現れる。系列ラベリングにおいて、陽性は「B-P」および「I-P」、陰性

Feature Type	N-gram	Window Size (unit : character)
Char	1, 2-gram	-2,-1,0,1,2,3,4,5
Char Type	1, 2-gram	-2,-1,0,1,2,3,4,5

(a) 文字ベース素性

Feature Type	N-gram	Window Size (unit : word)
Word	1,2,3-gram	-2,-1,0,1,2
Word[-1]	1,2,3-gram	-2,-1,0,1,2
Word[-2:-1]	1,2,3-gram	-2,-1,0,1,2
Yomi	1,2-gram	-2,-1,0,1,2
POS	1,2-gram	-2,-1,0,1,2

(b) 単語ベース素性

単語ベース系列ラベリングでは5種類の特徴を使用する。表中の、‘Word’, ‘Word[-1]’ and ‘Word[-2:-1]’ はそれぞれ表層単語、前単語、前2単語を示す。‘Yomi’は表層単語のひらがな読みで‘POS’は表層単語の形態素を示す。

図 2: 素性関数テンプレート

表 2: ER 性能

	適合率	再現率	F-尺度
文字ベース	0.912	0.954	0.933
単語ベース	0.854	0.827	0.841

表 3: ER + P/N 分類 性能

	P/N	適合率	再現率	F-尺度
文字ベース	<P>tag	0.902	0.952	0.926
	<N>tag	0.908	0.884	0.896
単語ベース	<P>tag	0.820	0.810	0.815
	<N>tag	0.724	0.603	0.659

は「B-N」および「I-N」のラベルを使用してタグ付けの範囲を定義する。

## 3. 評価

### 3.1 コーパスと設定

構築されたCRFのモデルを評価するために、2.1節で述べたコーパスとは別に、<p>タグと<n>タグを付与した日本語退院サマリ 100症例(計10,266文)を用いた。モデルの学習時のパラメータにはデフォルト値を用いた。単語ベースと文字ベースの2つの手法を実装し、性能を調査した。

### 3.2 評価方法

評価は、(1) ERおよび (2) ER + P/N分類 (融合タスク) の2つの処理で別々に行った。ERは事象抽出のみを行う処理であり、ER+P/N分類は、事象抽出および陽性または陰性の2つのタイプのモダリティを判定する処理である。

### 3.3 評価指標

性能は、既存研究の手法に基づいて、適合率、再現率およびF-尺度 ( $\beta=1$ ) を用いて評価した。抽出された事象が教師データファイル内の対応する正解事象と完全に一致する場合にのみ、正しいとみなした。評価方法は、CoNLL-2000共有タスクと同じである。評価に使用したPerlスクリプトは、CoNLL-2000のWebサイト<sup>1</sup>から入手できる。

表2は、文字ベースおよび単語ベースの方法でのERの性能評価の結果である。提案した文字ベースの方法は、すべての指標において単語ベースの方法よりも優れていた。

表3は、ER+P/N分類融合法の性能評価の結果である。この処理においても、文字ベースの方法がすべての指標で単語ベースの方法よりも優れていた。

ここで、<N>タグの精度は、<P>タグのそれと比較して低く、P/N分類がむしろ困難な課題であることを示唆している。P/N分類は時に、アノテーションの際にも難しい場合があり、この結果は妥当である。

## 4. 考察

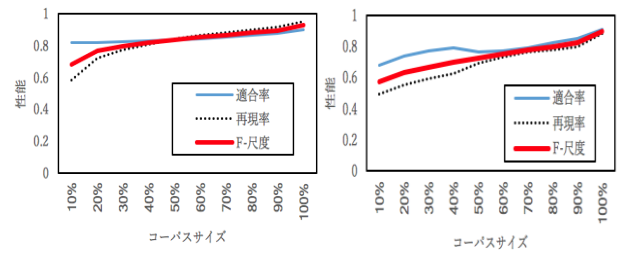
本章では、コーパスサイズの性能への効果、事象のカバー率、処理時間および今後の課題について考察する。

### 4.1 コーパスサイズの効果

学習に必要なデータ数を調べるために、文字ベースでのER+P/N分類融合法の性能を、コーパスサイズを10%ステップで10%から100%に変更しながら分析した。図3(a)と図3(b)に、<p>タグと<n>タグの抽出性能の変化をそれぞれ示す。コーパスサイズが大きくなるにつれて、それぞれの性能が徐々に向上しているのが分かる。再現率は、適合率よりもコーパスのサイズの影響を受けやすい傾向があるといえる。また、コーパス全体を使用しても性能が飽和状態に達しないこと、従って、コーパスのサイズを増やすことによってさらに性能の向上の余地があることを確認できた。

### 4.2 事象のカバー率：トークンレベルでの分析

トークンレベルでの分析として、頻度上位n位までの事象の種別およびそのトークンのカバー率を調査した



(a) 陽性事象<P>タグ (b) 陰性事象<N>タグ

図3: 性能とコーパスサイズの関係

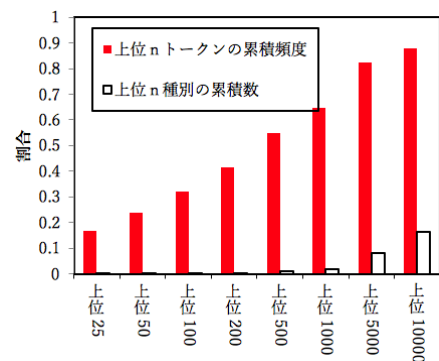


図4: 頻度が上位25から10000位までの陽性事象のトークン (赤) とその種別 (白) のカバー率

トークンとは、抽出された個々の事象であり、カバー率とは、コーパス中の全てのトークンまたは事象の種別の頻度の累計における、上位 n 位までのトークンまたは事象の種別の頻度の累計により求められる割合である。図4より、上位500位までの事象がトークンの約50%を占める。一方、事象の種別に関しては上位10,000位までを対象としても全体の2割弱のカバー率であった。つまり、典型的なロングテールの形を示しており、低頻度の多種類の事象が医療テキストに現れることを示している。

### 4.3 処理時間

処理速度は実用的なシステムを構築するうえで、重要な指標である。実際に、大学病院などの大規模な病院では毎日約3,000人もの患者が診療を受け、およそ6万もの文書が生成されている。これらの文書を処理するのにかかる時間を見積もった。処理時間は以下のスペックを有する計算機を用いて調査した。

- CPU: コア i7 6800K 6core/12thread 3.4GHz (ターボブースト 3.6GHz)
- メモリ: 32GB (8GB×4) DDR4-2133クワッドチャネル

<sup>1</sup> <http://www.cnts.ua.ac.be/conll2000/chunking/>

結果として、医療テキスト1,000件の処理時間は、モデルの学習処理では2分44秒、分析処理では2秒であった。1日で生成される全ての文書に対する処理時間に換算すると120(=2×60,000 /1,000)秒であり、ほとんどの病院にとって、実用に耐えられる処理時間と考えられる。

#### 4.4 今後の課題

今後の課題として、抽出された単語をICD-10やSNOMED-CT [9]などの概念にマッピングするという重要な課題が残っている。また、表記ゆれ対策のために、オーソグラフィック曖昧性除去技術を組み込んだモジュールの開発も検討している。もう1つの残りの課題は、再帰型ニューラルネットワーク (RNN) などの深層学習を用いた新たな手法の開発によりさらなる精度の向上が期待される。最後に、医療テキスト解析のニーズを満たすためには、エラー分析などによる改善の余地が残されている。

## 5. おわりに

本研究では (1) 事象抽出と (2) P/N分類の2つの処理からなる日本語による医療テキスト解析ツール MedEx/Jを開発した。提案した文字ベースのアプローチは、2つの処理を1つの系列ラベリングとして実行する。このアプローチには2つの重要な利点がある。1) 形態素解析が不要なため、事前処理が簡素化された点と、2) 単語ベースの方法よりもはるかに単純な特徴セットを使用しているが、単語ベースの方法よりも性能が優れている点 (ER ( $F_{\beta=1}$ ) = 0.93), ER+P/N分類 ( $F_{\beta=1}$ ) = 0.91)) である。

得られた結果は、提案したアプローチが医療テキストの解析できわめて有効であることを示唆している。さらに、シンプルな実装により、病院への設置の簡便性、メンテナンスコスト、アプリケーションの開発などの実用的な観点からも大きな利点があると考えられる。MedEx/Jは<http://sociocom.jp/parser.html>から入手可能である。

## 謝辞

この研究の一部は日本学術振興会補助金番号JP16H06395 および 16H06399, ならびに厚生労働科学研究費補助金番号 28030301 ならびに JST, ACT-I によって支援された。

## 参考文献

[1] W.W. Chapman, P.M. Nadkarni, L. Hirschman, L.W. D'Avolio, G.K. Savova, and O. Uzuner, Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative

solutions, *J Am Med Inform Assoc.* 18 (2011) 540-543.

[2] C. Friedman, S.B. Johnson, B. Forman, and J. Starren, Architectural requirements for a multipurpose natural language processor in the clinical environment, *Proc Annu Symp Comput Appl Med Care* (1995), 347-351

[3] cTAKES, <https://ctakes.apache.org/>

[4] J. Lafferty, A. McCallum, and F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proceedings of the International Conference on Machine Learning (ICML2001)*, 2001, pp. 282-289.

[5] Y. Huang and H.J. Lowe, A novel hybrid approach to automated negation detection in clinical radiology reports, *J Am Med Inform Assoc* 14 (2007) 304-311.

[6] P.G. Mutalik, A. Deshpande, and P.M. Nadkarni, Use of general-purpose negation detection to augment concept indexing of medical documents: A quantitative study using the UMLS, *Journal of the American Medical Informatics Association* 8 (2001) 598-609.

[7] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan, Evaluation of negation phrases in narrative clinical reports, *Proc AMIA Symp*, 2001, pp. 105-109.

[8] M. Morita, Y. Kano, T. Ohkuma, M. M., and E. Aramaki, Overview of the NTCIR-10 MedNLP task, *Proceedings of NTCIR-10*, 2013.

[9] Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe, Orthographic Disambiguation Incorporating Transliterated Probability, *International Joint Conference on Natural Language Processing (IJCNLP2008)*, 2008, pp. 48-55.