

NTCIR MedNLP: 本邦初の医療分野の言語処理コンテスト

荒牧 英治^{*1} 大熊 智子^{*1}

^{*1}京都大学 デザイン学ユニット ^{*2}富士ゼロックス

NTCIR MedNLP: The First Japanese Medical NLP Shared Task

Eiji ARAMAKI^{*1} Tomoko OHKUMA^{*1}

^{*1}Unit of Design, Kyoto University ^{*2}Fuji Xerox

Recently, medical records are increasingly written on electronic media instead of on paper, thereby increasing the importance of information processing in medical fields. We have organized an NTCIR-10 pilot task for medical records. Our pilot task, MedNLP, comprises three tasks: (1) de-identification, (2) complaint and diagnosis, and (3) free. These tasks represent elemental technologies used to develop computational systems supporting widely diverse medical services. Development has yielded 22 systems for task (1), 15 systems for task (2), and 1 system for task (3). This report presents results of these systems, with discussion clarifying the issues to be resolved in medical NLP fields.

Keywords: Natural Language Processing, Shared Task, Named Entity Recognition, de-identification

1. はじめに

近年、急速に紙カルテから電子カルテへの移行がはじまっている。新規に開業する診療所で電子カルテの導入率は7割を上回るとされ、今後さらに電子カルテの普及が進んでいくことは確実である¹⁾。カルテの電子化に伴い、紙カルテの時代には事実上不可能であった大規模な医療情報の利活用が進むと期待されている。この活用先として、例えば次のような例が想定される。

- 1) 新たな医学的知見の抽出
- 2) 診療行為の結果評価
- 3) 類似症例の検索
- 4) まれな副作用や疾患の頻度の正確な把握

これらの実現のためには、病名や医薬品名などの専門用語やそれに対応するコードの標準化、データ記録形式の標準化などが必須となり、それを効率よく行うために言語処理技術の利用が注目されている。

この流れを受けて、英語圏では1960年代から医療分野の言語処理研究が盛んになっている。しかし、一方で他の分野と比べて進歩が遅いことも指摘されている²⁾。その理由として、主なものは次の3つである：

- 1) 入手できるコーパスが不足している
- 2) アノテーション済みのコーパスはさらに不足している
- 3) アノテーションされる場合もその方針が統一されていない

我が国においても、こうした問題は同様であり、日本語で書かれたアノテーション済みの文書を研究者が共有できる仕組みが望まれる。

以上のような背景から、私たちは研究利用が可能な日本語のアノテーション済み医療文書を構築し、これを用いた解析タスクと共に研究コミュニティに提供することを計画した。

こうすることで、解析技術の客観的な評価を行うことが可能となる。また、医療文書の解析技術の開発に興味はあるがコーパスを持っていない研究者、および解析技術を適用する場を持っている企業を集めて産学連携のコミュニティを形成することで、課題の共有と解析技術の発展・向上を図ることができ。

2. 関連研究

さまざまな分野において、実験材料を共有して解析手法の評価を行う、ということが行われている。こうした催しの呼び方はいろいろであるが(sharedtask, contest, competition, challengeevaluation, criticalassessmentなど)、ここでは「シェアドタスク」に統一する。

シェアドタスクに参加するグループには実験材料が配られるため、研究者がまず直面する実験材料の入手という壁が取り払われる。また、複数のグループで同じ実験材料を共有することで、手法ごとの解析精度の特徴を評価や議論が可能となる。

このような利点から、言語処理分野では、TREC³⁾をはじめとしてCoNLL⁴⁾やCLEF Initiative⁵⁾、国立情報学研究所によるNTCIR⁶⁾、生命科学分野のBioNLP-ST⁷⁾、BioCreative⁸⁾、CALBC⁹⁾など多くのシェアドタスクが開かれている。

医療分野でも、医療分野の言語処理シェアドタスクとして2006年から米国国立衛生研究所(NIH; National Institutes of Health)の主導によるi2b2 (Informatics for Integrating Biology and the Bedside)¹⁰⁾が開催されている[20]。また、TREC[9]は2011年開催のTREC2011からMedical Records Trackを開始した。これらのシェアドタスクは英語圏の文書解析技術の向上に貢献している。

しかし、日本語での医療文書のシェアドタスクはこれまでなく、MedNLPが初めてとなる。

3. タスクの概要

先に挙げた目的を達成するためには、我々は日本語の医療文書を用いた言語処理シェアドタスクをNTCIR-10[URL]のパイロット・タスクとして開催した。MedNLPパイロット・タスクでは医療文書コーパスを利用した次の3種類のタスクを設定した：

- 1) 匿名化タスク：医療文書を研究利用するためには、そこに患者や関係者の個人情報が含まれていないことが望ましい。そこで匿名化タスクとしては、医療文章に含まれる個人情報抽出するタスクを設けた。
- 2) 症状と診断タスク：文章から症状や診断病名を抽出することは、様々な応用場面でまず前

処理として必須の処理となっている。そこで医療文書に含まれる症状や診断病名を抽出するタスクを設けた。

- 自由タスク: 上記の2タスク以外に、与えられたデータを用いて何が出来るか、実用的で創造的なアイデアを募集するタスクとして自由タスクを用意した。

4. 材料

4.1 コーパスの概要

本パイロット・タスクのためのコーパスとして、医師によって書かれた患者の病歴要約(medical history)を用意した。

病歴要約は、たとえば入院していた患者が退院する際や患者を他の医師に紹介する際に、第三者がその症例を理解できることを目的として書かれるもので、自然言語で記述される(図1にその例を示す)。病歴要約には診察をした医師によって整理された情報が凝縮されている。

ただし、病歴要約にはその患者本人の年齢や健康情報をはじめとして、社会生活像、家庭の事情や家族の病歴といった個人情報が含まれる。また、医療従事者の個人情報が含まれることもある。よって、個人情報保護の観点からこれをそのまま研究に利用することはできない。そこで、疑似的な病歴要約を書き起こしたものを収集した。この際、医学の知識のない者が書いた病歴要約が実際の患者像を反映することは大変難しいと考えられる。そのため、作成にあたっては、医師免許を持った臨床医に依頼した。この際、医師に研究の主旨を説明し、研究利用への同意を得た。

4.2 アノテーション

コーパスに対し、次のような2タイプのタグを付与した(括弧の中の数字は開発用コーパス2,244文中の各タグ数):

- 個人情報タグ:
 - <a> 年齢 (age, 56)
 - <t> 日時 (time, 355)
 - <h> 病院名 (hospital, 75)
 - <l> 場所 (location, 2)
 - <p> 個人名 (person, 0)
 - <x> 性別 (sex, 4)
- 医療情報タグ:

<c> 症状と診断表現 (complaint&diagnosis, 1,922)

症状に診断表現に関しては、単なる症状抽出タスク以外にも、

その症状が実際にあったのか(positive)、または、認められなかったのか(例; 胸水なし(negation)),または、家族歴としての症状なのか(family)、単なる疑いであったのか(suspicion)を区別して抽出するサブタスクも設けた。

それぞれの出現頻度を以下に示す。

- positive finding in the patient (1,314)
- positive finding in the person's family (32)
- negative finding (504)
- suspicious finding (72)

個人情報タグとは、匿名化の対象となる情報(日時, 年齢, 性別, 地名・施設名, 個人名)である。医療情報タグ

とは、病名で症状である。

図1にアノテーション済みのコーパスの例を示す。

5. 開催概要と結果

はじめに、本パイロット・タスクの参加者にアノテーションが付与された開発用コーパス(2,244文)およびアノテーション・ガイドラインを公開した。

2ヶ月間のシステム開発期間を経てテスト用コーパス(1,121文)を配布し、1週間以内に各チームから自動アノテーション結果を回収した。

参加資格は特に設けず、大学、研究所、企業から広く参加者を募集し、また個人でもグループでも参加できるものとした。

登録が切までに国内外からのべ16チームの参加申し込みがあった。各チーム最大3システムのエントリーを許しており、15システム(6グループ)がタスク1に参加、22システム(11グループ)がタスク2へ参加、1システムがTask3に参加した。

各チームの手法やリソースを表1に示す。結果を図2(タスク1)と図3(タスク2)に示す。

更に詳細なカテゴリごとの精度を表2(タスク1)と表3(タスク2)に示す。

より詳細な結果については文献¹⁾を参照されたい。

図に示されるように、上位のシステムについては大きな差がなかった。これら上位システムは、すべてCRFなどの機械学習を用いており、手法の類似性により接近した精度が得られているものと思われる。

今後、外部リソースの違いや経験則による各システムの特長についてさらなる考察が望まれる。

また、上位システム間の差については、コーパスのサイズから統計的に同等であり、システムの優劣を示すものではないことを補足する。

6. おわりに

本稿では、本邦初の医療分野の言語処理のシェアドタスクについて述べた。

他の分野と異なり、材料であるコーパスが入手困難である医療分野において、シェアドタスクは、産学連携を加速するよい機会となりうる。

この試みを継続的に開催することでコミュニティが形成され、さらに開発は促進されるであろう。

今後の継続開催に向け、様々な方の協力を得ながら努力を続ける予定である。

謝辞

本研究は、JST戦略的創造研究推進事業(さきがけタイプ)「情報環境と人」および科研費補助金(若手研究A)による。本シェアドタスクの開催にご協力して頂いたNTCIR事務局および医師、アノテーター、参加者の皆様に感謝いたします。

参考文献

- [1] 株式会社シード・プランニング. 2011-2012年版電子カルテの市場動向調査. 2012.
- [2] Wendy Chapman, Prakash Nadkarni, Lynette Hirschman, Leonard D'Avolio, Guergana Savova, Ozlem Uzuner. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. JAmMed Inform Assoc, 18, 540-543, 2011.
- [3] TREC. <http://trec.nist.gov/>.
- [4] CoNLL. <http://conll.cemantix.org/>.

[5] CLEF. <http://www.clef-initiative.eu/>.
 [6] NTCIR. <http://research.nii.ac.jp/ntcir/>.
 [7] BioNLP-ST. <http://www.bionlp-st.org/>.
 [8] BioCreative. <http://www.biocreative.org/>.
 [9] CALBC. <http://www.calbc.eu/>.

[10] i2b2. <http://www.i2b2.org/>.
 [11] Mizuki Morita, Kano Yoshinobu, Ohkuma Tomoko, Miyabe Mai, Eiji Aramaki. Overview of the NTCIR-10 MedNLP task. Proceedings of NTCIR-10, 2013.

工場に勤めている<a>64歳の<x>男性</x>。<t>2025年8月2日(来院5日前)頃から</t><c>腹痛</c>が生じるとともに、<c>食欲不振</c>、<c>嘔気</c>・<c>嘔吐出現</c>した。体幹は温かいが、末梢は<c>湿潤冷汗</c>で<c>ショック状態</c>。明らかな<c modality="negation">運動麻痺</c>はみられず。<t>翌日</t>、<c>意識障害出現</c>し、<c>腎機能障害</c>の増悪を認めて徐々に<c>尿量低下</c>し、<t>8月9日18時10分</t>に<c>心肺停止</c>。<t>8月9日21時44分</t><c>死亡確認</c>。

図1 コーパスの例

表1 参加チームの手法とリソース

Group ID	Methods	Resources
NTTD	Word match, Semi-supervised learning	MEDIS Standard Masters (disease names)
LSDP	N/A	N/A
KobeU ulab	Structured perceptron Online learning	Japanese newspaper (disease names)
msiknowledge	CRF, Language Model	
UT-FX	CRF	MedDRA/J, MEDIS Standard Masters, Original corpus
HCRL	CRF, Word match	Japanese Wikipedia (disease names)
niph	Word match	Original dictionary
oka1	CRF, Word match	
NECLA	CRF	UMLS, LSD
cks01	CRF	MEDIS Standard Masters
SinicaNLP	Word match, Machine translation	Original dictionaries (in Chinese)

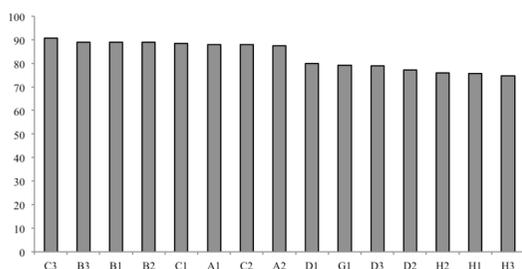


図2 各参加システムのタスク1の精度 (F値)

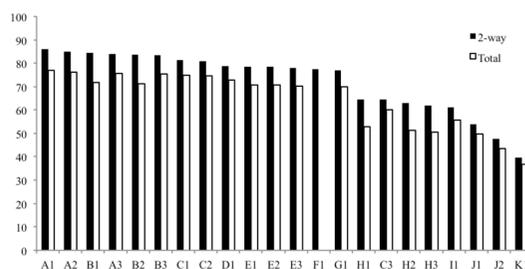


図3 各参加システムのタスク2の精度 (F値)
 表現の抽出 (2way) と症状の有無を考慮した精度 (Total)

表2 タスク1の詳細

	<a> age			<x> sex			<t> time			<h> hospital name		
	P	R	F	P	R	F	P	R	F	P	R	F
C3	90.32	87.5	88.89	100	100	100	87.16	91.49	89.27	97.30	94.74	96.00
B3	90.00	84.38	87.10	100	50.00	66.67	91.30	89.36	90.32	97.06	86.84	91.67
B1	93.33	87.5	90.32	100	100	100	90.65	89.36	90.00	89.47	89.47	89.47
B2	90.00	84.38	87.10	100	100	100	91.24	88.65	89.93	91.89	89.47	90.67
C1	96.67	90.62	93.55	100	50.00	66.67	91.18	87.94	89.53	93.55	76.32	84.06
A1	92.86	81.25	86.67	100	50.00	66.67	91.04	86.52	88.73	91.89	89.47	90.67
C2	96.67	90.62	93.55	100	50.00	66.67	89.13	87.23	88.17	96.77	78.95	86.96
A2	92.86	81.25	86.67	100	50.00	66.67	89.05	86.52	87.77	91.89	89.47	90.67
D1	92.31	75.00	82.76	100	50.00	66.67	82.84	78.72	80.73	96.15	65.79	78.12
G1	80.65	78.12	79.37	100	50.00	66.67	84.56	81.56	83.03	72.73	63.16	67.61
D3	88.89	75.00	81.36	100	50.00	66.67	83.08	76.60	79.70	96.15	65.79	78.12
D2	92.31	75.00	82.76	100	50.00	66.67	75.86	78.01	76.92	96.15	65.79	78.12
H2	83.87	81.25	82.54	100	100	100	73.79	75.89	74.83	77.78	73.68	75.68
H1	80.65	78.12	79.37	100	100	100	75.86	78.01	76.92	70.27	68.42	69.33
H3	83.87	81.25	82.54	100	100	100	73.79	75.89	74.83	70.27	68.42	69.33

表3 タスク2の詳細

	c-positive			c-family			c-negation			c-suspicion		
	P	R	F	P	R	F	P	R	F	P	R	F
A1	81.91	76.80	79.27	84.62	50.00	62.86	80.91	72.06	76.23	50.00	20.00	28.57
A2	79.02	77.12	78.06	83.33	45.45	58.82	80.18	72.06	75.91	57.14	26.67	36.36
B1	80.80	68.00	73.85	65.22	68.18	66.67	75.23	67.61	71.22	35.85	63.33	45.78
A3	79.16	75.36	77.21	76.92	45.45	57.14	82.41	72.06	76.89	63.64	23.33	34.15
B2	80.61	67.20	73.30	71.43	68.18	69.77	74.32	66.80	70.36	36.36	66.67	47.06
B3	80.92	73.28	76.91	82.35	63.64	71.79	84.50	68.42	75.62	50.00	30.00	37.50
C1	78.84	71.52	75.00	100	68.18	81.08	89.47	68.83	77.80	57.14	26.67	36.36
C2	79.61	71.20	75.17	100	68.18	81.08	90.16	66.80	76.74	57.14	26.67	36.36
D1	79.66	67.04	72.81	100	54.55	70.59	90.16	66.80	76.74	61.54	26.67	37.21
E1	73.04	70.24	71.62	77.27	77.27	77.27	71.26	71.26	71.26	42.22	63.33	50.67
E2	73.51	69.28	71.33	77.27	77.27	77.27	72.02	70.85	71.43	41.30	63.33	50.00
E3	74.39	68.80	71.49	77.27	77.27	77.27	71.37	67.61	69.44	41.30	63.33	50.00
F1	-	-	-	-	-	-	-	-	-	-	-	-
G1	72.87	67.04	69.83	66.67	36.36	47.06	82.35	68.02	74.50	55.00	36.67	44.00
H1	57.39	54.08	55.68	42.11	36.36	39.02	51.20	51.82	51.51	11.11	6.670	8.33
C3	68.21	54.24	60.43	100	63.64	77.78	69.79	54.25	61.05	36.84	46.67	41.18
H2	56.03	52.00	53.94	40.00	45.45	42.55	50.62	49.80	50.20	10.00	6.67	8.00
H3	54.01	51.68	52.82	40.00	45.45	42.55	50.21	49.39	49.80	10.00	6.67	8.00
I1	48.86	61.76	54.56	70.83	77.27	73.91	73.05	49.39	58.94	52.17	40.00	45.28
J1	49.53	50.72	50.12	78.95	68.18	73.17	53.61	42.11	47.17	35.71	50.00	41.67
J2	47.20	43.20	45.11	78.95	68.18	73.17	47.13	29.96	36.63	35.90	46.67	40.58
K1	54.70	35.36	42.95	100	18.18	30.77	50.00	10.93	17.94	44.44	13.33	20.51