臨床医療テキストの構造化システム

荒牧英治† 三浦康秀‡ 外池昌嗣‡ 大熊智子‡ 杉原大悟‡ 増市博‡ 大江和彦+
†東京大学 知の構造化センター
‡富士ゼロックス (株) 研究技術開発本部
+ 東京大学 医学部附属病院

eiji.aramaki@gmail.com
{Yasuhide.Miura, masatsugu.tonoike, daigo.sugihara,ohkuma.tomoko,
hiroshi.masuichi}@fujixerox.co.jp
kohe@hcc.h.u-tokyo.ac.jp

はじめに

平成 13 年度に政府が発表した「保健医療分野の情報化にむけてのグランドデザイン」にて，電子カルテシステムの普及が課題の一つとして掲げられて以降，我が国では急速に電子カルテが普及し，その結果，大量の臨床データが電子化された状態で蓄積されつつある．このデータを構造化されたデータとして利用できれば，過去に類をみない大規模な臨床研究が実現可能であり，大きな期待がよせられている．しかし，カルテ中の一部の情報は自然言語で記述されており，カルテデータの全てを構造化されたデータとして利用するためには，自然言語処理技術が必須となる．

このような背景から，2007 年より我々はカルテから患者情報を抽出するシステムを研究／開発している．本稿では，カルテの一種である退院サマリ（退院時に記述される患者の経過を要約した文書）を対象とし，特に投薬とそれによって生じた副作用に関する記述を抽出するシステムを研究／開発したので報告する．

開発したシステムは医療ドメインに特化した極めて実際的なものであるが，システムを構成するモジュールは，(M1) 固有表現認識，(M2) 表記ゆれ吸収，(M3) 事実性判定および (M4) 関係抽出という言語処理にとって馴染み深い基礎研究となっている．

言い換えれば，本プロジェクトは言語処理の基礎研究を組み合わせることで，どこまで実用的な医療システムを作れるかの挑戦と捉えることもできる．本研究の事例が今後の言語処理システムのデザインに有用な知見となると考え，報告する．

提案アプローチと対象文章

提案システムは入力となる退院サマリ文章を順次 4 つの言語処理モジュールが処理していくことで副作用を抽出する．各モジュールの入出力例を図 1 に示す．扱う対象が医療テキストであるが，図に示されるように，各処理は一般的な言語処理技術の組み合わせである．

本研究で扱う退院サマリの入院後経過セクションとは時間軸にそって，入院後の患者状態，処置内容，検査結果が記載される．例を表 1 に示す．表に示されるように，文としての構造をもつものの専門用語が多く，また，サマリの名が示すように，すべての情報が臨床的に重要な意味を持っている．

本稿では各モジュールの概要，手法，訓練データ（材料）及び精度を以降の章で概観する．

(M1) 固有表現認識モジュール

【概要】テキスト中の医療表現を特定する．ここでいう医療表現とは，医師との議論を行い下表のカテゴリを用いた．

TIMEX3	時間表現．汎用的な時間表現アノテーションの枠組みである TimeML (TIMEX3)[15] のサブセットを用いている．
S	副作用表現
M-NAME	医薬品名
M-NUM	医薬品の投与量および単位
TEST	検査群
T-NAME	具体的な検査名
T-NUM	検査の値および単位
ACTION	退院，入院，転院など
A-LOC	ACTION の行われる場所
P	病理所見
CHANGE	変化に関する表現
R	治療処置表現．治療，または治療行為を表す表現.（例）ギプス固定
D	疾患／症状表現: 疾患を表す表現.
DEID	匿名化すべき表現． 医師名，施設名など．
B	部位表現: 場所が特定できる部位を表す表現.（例）視神経

【手法】このタスクは固有表現認識の一種であるので，系列ラベリング問題として解く手法を採用した [12]．電子カルテ文章に対しては，Support Vector Machine (SVM) [19] による手法 [17, 16]，Conditional Random Field (CRF) による手法 [1] があるが，本研究では後者の手法を用いた．

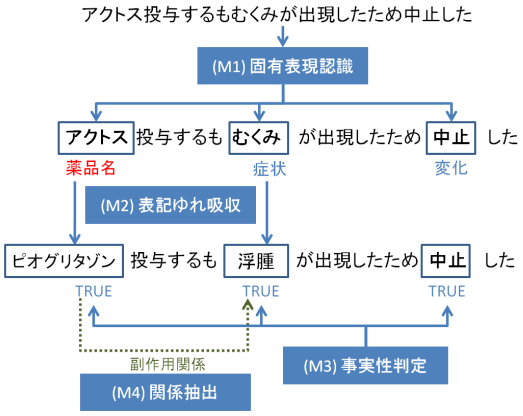

図 1: システムの構成.

表 1: 退院サマリの例.

****年*月健診で FBS***，HbA1c9.7 %であった。当院当科紹介受診し，糖尿病と診断された。食事・運動・内服 (ダオニール (1.25)1 T) で対応していた。HbA1c6 %台を推移。****年末より運動量低下し，BS コントロール徐々に悪化した。****年から HbA1c9.5 %と増悪。ダオニール (1.25)4 Tへ増加するも，HbA1c11 %台と増悪した。アクトス (30) 0.5〜1 T追加。以後 HbA1c 5 %台となったため，****年*月ダオニール (1.25)1 Tへ減量した。ダオニール 1.25 mg＋アクトス 15 mgにて HbA1c6 %台を維持していたが，翌年，生活が不規則となり，同年*月 HbA1c10 %台と増悪したため，アクトス (30)1 Tへ増量。以後 HbA1c6 %を推移していた。****年*月 HbA1c12.4 %と増悪し，入院勧めるも仕事上の都合から本人拒否されたため，食事を 1 制限とし，ダオニール 2.5 mgに増量の上，外来で治療継続した。

テキスト中の一部の数値／日付は「*」とした.

ただし，形態素解析エラーに対して頑健にするため，文字単位でラベル (IOB2 ラベル) を与えた．学習の素性には，医療用辞書を追加した形態素解析器（ChaSen）*の解析結果を用いた.

【材料】下のように各医療表現カテゴリを xml タグとして約 600 の文章に対してアノテーションを行った.

```
<TIMEX3>2001 年</TIMEX3>、<R> <D> <B>舌</B>癌</D>手術</R>時の<B>腹直筋採取部位</B>に<D>疼痛</D> を自覚し<A-LOC>当院消化器内科</A-LOC>に<ACTION>通院</ACTION>を開始。
```

【実験】実験は，交差検定 (10 フォールド) にて Precision(P), Recall(R), F-measure(F) を調査し，表 2 の精度を得た．多くのカテゴリで 70〜80 %の再現率を得ており，今後，コーパスを拡張することを考えると有望な精度だと考えられる．ただし，部位などバリエーションが多い一部のタグについては，低い精度にとどまり，外部リソースの利用が必要であろう．手法／実験の詳細については文献 [21] をあたられたい.

*http://chasen-legacy.sourceforge.jp/

表 2: 固有表現認識の精度.

	P(%)	R(%)	F
A-LOC	89.1	87.9	88.5
ACTION	94.6	91.0	92.8
B	68.8	59.3	63.7
CHANGE	84.6	74.8	79.4
DEID	82.7	77.1	79.8
D	85.5	80.2	82.8
M-NAME	86.9	81.3	84.0
M-NUM	95.9	93.8	94.8
R	84.5	76.3	80.2
T-NAME	84.7	76.6	80.5
T-NUM	80.7	76.3	78.4

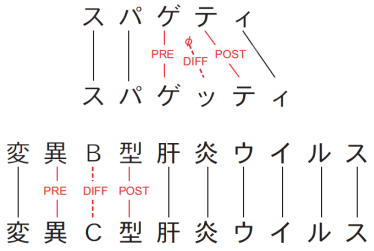

図 2: 表記ゆれ判別に用いる素性の例（2 つの語の差分文字 (下図の DIFF) や周辺文字（下図の PRE と POST)）.

(M2) 表記ゆれ吸収モジュール

【概要】(い) 固有表現認識モジュールにて特定された表現と（ろ）各種用語集（病名マスター [8]，医薬品マスター†，MedDRA/J‡など）に収載された用語を比較し，(い) が（ろ）の表記ゆれであるかどうかを判別する.

【手法】表記ゆれ解消はよく研究されている分野であり，McCallum 等 [11] の編集操作を CRF で学習する手法，Tsuruoka ら [18] の文字列置換ルールの汎用性・曖昧性を計算する手法，Aramaki 等 [2] や Okazaki 等 [14] の表記ゆれかどうかを機械学習で識別する手法などがある.

本システムでは，Aramaki 等 [2] の 2 つの語の差分文字 (図 2 の DIFF) や周辺文字 (図 2 の PRE と POST) を素性として，SVM にて表記揺れを識別する手法を用いた.

【材料】下のように表記ゆれであった表現の場合に，value タグとして標準的表現をアノテートした.

```
<M-NAME value=ダオニールダオニール</M-NAME>1.25 mg＋<M-NAME value=ピオグリタゾン>アクトス</M-NAME>15 mgにて...
```

【実験】実験の結果，表 3 の精度を得た．編集距離ベースの類似度による手法と比べ大きく精度を向上させており，提案手法の有効性を示している．手法実験の精度については文献 [2] をあたられたい.

†http://www.medis.or.jp/
‡http://www.sjp.jp/

表 3: 表記ゆれ吸収の精度.

	P(%)	R(%)	F
編集距離 [10]	91.2	36.3	51.3
PROPOSED	81.7	82.7	82.2

表 4: 事実性判定の精度

	#	P(%)	R(%)	F
NEGATION	441	84.1	77.3	80.6
PURPOSE	346	91.3	63.8	75.1
S/O	242	90.7	72.3	80.5
POSSIBLE	36	83.3	40.5	54.5
OTHER	32	76.6	29.3	42.4

(M3) 事実性判定モジュール

【概要】各医療イベントには事実か事実でなかったかを示す**モダリティ**を付与した．ここでいうモダリティとは，医療分野に特に頻出する下記のものを扱った．

NEGATION	「.. は認められず」
PURPOSE	「... 予定で」「目的で」「のため」「の方針となる」
S/O	「.. の疑いがある」「.. の可能性があり」「S/O」
POSSIBLE	「の必要性があり」「適応あり」
OTHER	「患者が.. を希望したため」

【手法】先行研究では，人手でパターンを記述したり [4, 5, 6, 13] 機械学習でパターンを学習 [7, 9] している．また，否定のスコープをラベリングするアプローチも提案されている [20].

本研究では，先行研究で主流である機械学習でパターンを学習するアプローチをとったが，単なる表層のパターンではなく，加えて係り受け構造上でのパターンも扱った．すなわち，係り受け構造上でイベント表現周辺（文節数 5）に含まれる形態素を素性として，各モダリティ毎に，そのイベントがモダリティを持つか否かを SVM にて識別している．

【材料】下表のように各医療表現毎に，モダリティを modality 属性としてアノテーションした．

> ****年*月　HbA1c12.4　%と増悪し，<ACTION modality=OTHER>入院</ACTION>勧めるも仕事上の都合から本人拒否されたため...　糖尿病の<R modality=PURPOSE>教育血糖コントロール</R>目的で<ACTION>入院</ACTION>となった。

【実験】Precision 89.4%, Recall 82.5%, F-measure 85.5% を得た．各モダテリティごとの精度は表 4 になる．# はアノテーションした数である．表に見られるように，頻出するモダリティについては高い精度を持つ傾向があり，低精度なモダリティについても，今後コーパスを拡充することで対応できる可能性がある．詳しくは文献 [3] をあたられたい．

表 5: 関係抽出の精度

	P(%)	R(%)	F
A: PTN	29.8	37.5	32.6
B: A+DEPT	39.1	36.5	37.7
C: B+NER	37.5	44.4	40.3

(M4) 関係抽出モジュール

【概要】特定された医療用語間の関係を特定する．現在は，(1) 時間表現とイベント表現の間の関係，(2) 薬剤とそれが引き起こす副作用という 2 つの関係を扱っている．

【手法】任意の 2 つの医療表現について，それらが関係をもつかどうかを SVM にて識別した．SVM の素性としては，係り受け関係上での距離（文字数 or 形態素数）や形態素列，他の医療表現などを用いた．

【材料】コーパス上で relation 属性として関係を表現した．例えば，下例では，薬品である「アクトス」とその副作用である「浮腫」が同じ番号の relation を持つことで，これらが薬剤-副作用関係を持つことが表現されている．

> <M-NAME relation=1>アクトス</M-NAME>投与するも<S relation=1>浮腫</S>が出現したため中止した.

【実験】医療源間の表層的な形態素パターンだけを用いた手法よりも (A)，係り受け構造上での形態素パターン (B) や，医療表現認識結果を用いた (C) の方が精度がよく，最高で Precision 37.5%, Recall 44.3%, F-measure 40.3%を得た（表 5）．詳しくは文献 [22] をあたられたい．

まとめ

本稿では，カルテの一種である退院サマリから副作用情報を抽出するシステムと，それを構成する各モジュールについて概観を述べた．

各モジュールの手法／精度を述べたが，実際のシステムの精度は (M1)-(M4) のすべての精度をかけあわせたものとなる．今後，実用的な環境でどのような適合率，再現率のバランスを求めるかなど，検討すべき課題は多く残されている．

紙面の都合，各モジュールについては簡便な記述のみにとどまったが，詳細については，引用された各論文を，また，システムの動作に関しては，ウェブ上の試用バージョンを参照されたい（図 3）[§]．

最後に，退院サマリをはじめ，カルテ文章は，医師間で閲覧される高度に専門的な文章である．一般的に，このような専門的な文章を処理するにあたって，ドメイン固有の知識をインプリメントする方式が安易で，少なくとも研究開発初期は効率がよいと考えられる．しかし，我々は，自然言語処理の基礎研究をベースに，ドメインへの依存性を可能な限り避けながら，研究／開発を行っている．本プロジェクトによって，医療のような実用的要請の高い分野においても，自然言語処理の基礎研究の有効性を示唆できればと望んでいる．

[§]http://luululu.com/mednlp/

図 3: 入力文章（図上）と構造化結果（下）．

参考文献

[1] Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe. Automatic deidentification by using sentence features and label consistency, 2006.

[2] Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe. Orthographic disambiguation incorporating transliterated probability. In *Proceedings of International Joint Conference on Natural Language Processing (IJCNLP2008)*, pp. 48–55, 2008.

[3] Eiji Aramaki, Yasuhide Miura, Masatsugu Tonoike, Tomoko Ohkuma, Hiroshi Mashuichi, and Kazuhiko Ohe. Text2table: Medical text summarization system based on named entity recognition and modality identification. In *Proceedings of the Human Language Technology conference and the North American chapter of the Association for Computational Linguistics (HLT-NAACL2003) Workshop on BioNLP*, pp. 185–192, 2009.

[4] Wendy Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce Buchanan. Evaluation of negation phrases in narrative clinical reports. In *Proceedings of AMIA Symp*, pp. 105–109, 2001.

[5] Wendy Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, Vol. 5, pp. 301–10, 2001.

[6] Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R, Bergstrom, and Dietlind L. Wahner-Roedler. *A controlled trial of automated classification of negation from clinical notes.* BMC Medical Informatics and Decision Making 5:13, 2005.

[7] Ilya M. Goldin and Wendy Chapman. Learning to detect negation with not in medical texts. In *In Workshop at the 26th ACM SIGIR Conference*, 2003.

[8] K. Hatano and K. Ohe. Information retrieval system for japanese standard disease-code master using xml web service. In *American Medical Informatics Association (AMIA) Symposium*, pp. 597–602, 2003.

[9] Yang Huang and Henry J. Lowe. A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association*, Vol. 14, No. 3, pp. 304–311, 2007.

[10] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk SSSR*, Vol. 163, No. 4, pp. 845–848, 1965.

[11] A. McCallum, K. Bellare, and F. Pereira. A conditional random field for discriminatively-trained finite-state string edit distance. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 388–395, 2005.

[12] A. McCallum and W. Li. Early results for named entity recognition with conditional random fields. In *Proceedings of The Seventh Conference on Natural Language Learning (CoNLL)*, pp. 188–191, 2003.

[13] Pradeep G. Mutalik, Aniruddha Deshpande, and Prakash 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*, Vol. 8, No. 6, pp. 598–609, 2001.

[14] Naoaki Okazaki, Yoshimasa Tsuruoka, Sophia Ananiadou, and Jun'ichi Tsujii. A discriminative candidate generator for string transformations. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, pp. 447–456, 2008.

[15] J. Pustejovsky, J.M. Castano, R. Ingria, R. Sauri, R.J. Gaizauskas, A. Setzer, G. Katz, and D.R. Radev. *New Directions in Question Answering: Timeml: Robust specification of event and temporal expressions in text.* AAAI Press, 2003.

[16] Tawanda Sibanda, Tian He, Peter Szolovits, and Ozlem Uzuner. Syntactically-informed semantic category recognizer for discharge summaries. In *Proceedings of the Fall Symposium of the American Medical Informatics Association (AMIA 2006)*, pp. 11–15, 2006.

[17] Sibanda Tawanda and Uzuner Ozlem. Role of local context in automatic deidentification of ungrammatical, fragmented text. In *Proceedings of the Human Language Technology conference and the North American chapter of the Association for Computational Linguistics (HLT-NAACL2006)*, pp. 65–73, 2006.

[18] Y. Tsuruoka, J. McNaught, and S. Ananiadou. Normalizing biomedical terms by minimizing ambiguity and variability. *BMC Bioinformatics*, Vol. 9, No. 3, 2008.

[19] Vladimir Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, 1999.

[20] Veronika Vincze, Gyorgy Szarvas, Richard Farkas, Gyorgy Mora, and Janos Csirik. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, Vol. 9, No. 11, 2008.

[21] 荒牧英治, 三浦康秀, 外池昌嗣, 大熊智子, 増市博, 大江和彦:. 退院サマリ文章可視化システムの構築. 言語処理学会 第 15 回年次大会, pp. 348–351, 2009.

[22] 三浦康秀, 荒牧英治, 外池昌嗣, 大熊智子, 杉原大悟, 増市博, 大江和彦:. 電子カルテからの副作用関係の自動抽出. 言語処理学会 第 16 回年次大会, 2010.