HOKKAIDO UNIVERSITY

Title	患者が発信する情報活用のための薬剤効果抽出に関する研究
Author(s)	北嶋, 志保
Citation	北海道大学. 博士(情報科学) 甲第11762号
Issue Date	2015-03-25
DOI	10.14943/doctoral.k11762
Doc URL	http://hdl.handle.net/2115/58913
Type	theses (doctoral)
File Information	Kitajima_Shiho.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文

患者が発信する情報活用のための薬剤効果抽出に関する研究

北嶋志保

北海道大学大学院情報科学研究科

目次

要旨 iv

第 1 章　はじめに 1
1.1 研究の背景と目的 1
1.2 目指すシステムの構想 2
1.3 本研究の位置づけ 3
1.4 本論文の構成と概要 4
参考文献 6

第 2 章　関連研究 8
2.1 副作用の関連性推測システム 8
2.2 医療情報の入手方法 8
2.2.1 患者の記録を共有する試み 8
2.2.2 医薬品検索サービス 9
2.3 副作用記述箇所特定システム 9
2.4 商品や場所についての評価を抽出する研究 10
2.5 既存研究の問題点と提案システムについて 11
参考文献 12

第 3 章　対象とするブログとその傾向 14
3.1 対象とするブログ 14
3.2 TOBYO データベース 16
3.3 TOBYO スニペットの傾向 16
3.3.1 薬剤名とその評価の出現位置 16
3.3.2 手がかり語 17
3.3.3 対象と効果をつなぐ助詞 18
参考文献 19

第 4 章　提案システムの概要 20
4.1 解析文生成 21
4.2 三つ組要素存在文節の発見 21
4.3 三つ組要素生成 23
参考文献 24

第 5 章　評価実験 25
5.1 データ 25
5.2 評価に用いる式 25
5.3 比較対象 26
5.4 結果と考察 27
5.4.1 出力誤り分析 28
5.4.2 正解データ出力状況分析 29
参考文献 31

第 6 章　精度改善手法の提案 32
6.1 複数パターンを用いた再現率向上手法 32
6.2 辞書を用いた適合率向上手法 33
6.2.1　お薬 110 番辞書 (110 dictionary) 33
6.2.1 薬剤添付文書辞書 (MPi dictionary) 34
6.2.3 辞書の評価実験結果に対する考察 35
6.3 手がかり語リストの拡張 36
6.3.1 人手による追加 36
6.3.2 類語辞典を用いた追加 37
参考文献 38

第 7 章　アノテーションデータを用いた評価実験 39
7.1 スニペットから生成したデータを用いた評価実験 39
7.1.1 正解データ 39
7.1.1.1 アノテーション方法 39
7.1.2 結果 41
7.2 ブログ記事全文に対する評価実験 42
7.2.1 正解データ作成方法 42
7.2.2 実験結果とその考察 43

第 8 章　薬剤服用情報の出現位置,内容に関する考察 45
8.1 薬剤服用情報の出現特徴 45
8.1.1 スニペット分析結果 45
8.1.2 ブログ記事全文分析結果 49
8.2 薬剤服用情報の新規性 51
8.2.1 対象データ 51
8.2.2 結果と考察 52
参考文献 55

第 9 章　おわりに 56

謝辞 58
研究業績 59
付録 A 63
付録 B 65
付録 C 71

要旨

患者中心の医療が進められており，患者が治療に主体的に参加し，治療方法を選択することが求められている．しかし，病気や診断方法，治療方法が複雑であるため，突然病気を患った患者がこれらを正しく理解し，判断，決定することは容易ではない．患者の意思決定や治療に対する積極性を促し，QOL（生活の質）や健康状態に良い影響をもたらすためには，医療に関する必要な情報を取得し，理解，活用する力が患者自身に求められている．ところで，情報化社会の進展に伴い，誰でも容易にインターネット上に情報発信が可能となっている．そのため，インターネット上に存在する患者やその家族によって発信された情報が近年注目されている．実体験に基づく医療情報は患者の不安を軽減させたり，励みとなる可能があることに加えて、大規模かつ即時的な情報は，既存の調査を上回る可能性がある．しかし，人手で膨大なデータから，求める情報のみを取捨選択し，収集するには多大な労力が必要である．また，人手による検索では，悪い面ばかり無意識に注目し収集してしまうといった問題点が挙げられる．したがって，本研究は，インターネット上から得られる患者の実体験情報を自動的，網羅的に収集，提示することで，患者の判断材料や励みとなるシステムの構築を目的としている．

近年，自然言語処理技術を用い，医療カルテから薬剤名や病状を特定する研究が進められている．著者も医療カルテやブログ記事を対象とし，薬剤名や症状名などの医療用語の位置を機械学習により特定する研究を行ってきたが，「頭痛」のような症状名は抽出できても「頭が痛む」のように対象と評価のセットで記載された症状については抽出の対象としておらず，また非専門的な表現にも対応しきれていないなどの問題点があった．本研究で著者が対象とするブログ記事も専門家ではない患者やその家族によって書かれたものであるため，専門家によって書かれたカルテに出現する用語や言い回しとは異なる表現が使われることが多いことが問題点として考えられる．そのため，本学位論文は闘病ブログに現れる表現の多様性を損なわず，適切に抽出するシステムを構築して行った研究について述べている．

医療情報のなかでも，医薬品は治療において必要不可欠なものであるため，医薬品の効果や付随して起こる副作用についての情報を取得することは，患者にとってよりよい治療につながると考えられる．本学位論文では，目指すシステムの第一段階として，患者によって書かれたブログ記事から，薬剤の服用による変化，効果，副作用を，（薬剤，対象，効果）の三つ組で抽出するシステムの構築・提案を行った．

実際にブログに出現する薬剤の効果，副作用に関する記述の特徴を調査し，その際用いられる特定の表現を収集し，手がかりとした．その手がかり語と構文情報を考慮したパターンマッチングにより抽出を行う．専門用語や評価表現に着目した抽出を行わないため，話し言

葉で書かれたブログの表現の多様性を保持した抽出が可能であり，例えばこれまで抽出が困難であった擬音語や擬態語で書かれた評価表現も収集することができる点に本研究の独創性がある．

ブログに出現する薬剤に関する効果，副作用の記述に対し，手がかり語を用いたパターンの有効性を確認するため，ブログの要約文であるスニペットを対象に評価実験を行った．一般的に意見抽出に用いられるパターンマッチング手法の評価結果と比較したところ，適合率において 40.7 ポイント高い 42.1%という優位性のある結果を示し，本手法の有効性を確認することができた.要するに，従来の意見抽出に用いられる「薬剤名→対象」「対象→効果」（矢印は係り受け関係を示す）がブログに出現する薬剤に関する情報には不適切なことが多く，提案した「薬剤名→効果」「対象→効果」のパターンが適していることが明らかとなった．

しかし，再現率は 6.2%と低い結果となった．その原因として，大きく 2 つ考えられる．手がかり語が存在しない，助詞の省略といった理由から要素が存在する場所を特定することの誤りによって出力ミスまたは出力が得られないこと，また抽出された要素が，三つ組の要素として不適切であることである. これらの問題を解決するため, 二つの改善手法を提案した．第一に，提案したパターンが当てはまらない場合についても抽出を可能にするため，手がかり語が存在する場合，しない場合について複数のパターンを提案し，それらを組み合わせて抽出を行った．その結果，再現率が 24.8 ポイント向上し 31.0%となり，より柔軟な抽出が可能となったことが示された．また，評価表現は擬音語や擬態語など多彩な表現で記述されることが多いため，評価表現辞書に存在する表現では不十分なことが多い．しかし，薬剤の効果，副作用が現れる対象は身体の部位や感情など，ある程度限定されている．そこで第二に，対象要素の適切性を判断することにより，適合率の向上を図った．薬剤添付文書中に存在する対象要素として用いられる可能性のある単語を収集し，辞書の作成を行った．このようにして作成した辞書をフィルタとして用いることで, 適合率が 15.8 ポイント向上し 57.9%となり，対象単語に対するフィルタリングの有効性が明らかとなった．

さらに，スニペットのみならずブログ全文を対象とし，システムの有効性を確認した．システムの改善に役立てるため，薬剤の効果，副作用が書かれたテキストの特徴を解析し，複数のパターンに分類されることを示した．今後は，意味解析を用いて文の構文パターンを分類し，最適な抽出パターンを判別，適用すること，照応解析により薬剤名が含まれない文からも抽出を行うこと，モダリティを考慮した抽出が必要であると考えられる．

表目次

表 1. 薬剤名とその効果の出現位置例 17
表 2. 薬剤情報としての三つ組を含む文の例 18
表 3. 手がかり語例 18
表 4. 対象要素と効果要素間に存在する助詞 18
表 5. 三つ組要素生成例 23
表 6. 評価結果 28
表 7. 提案システムの抽出例 28
表 8. 組み合わせシステム評価結果 32
表 9. 組み合わせシステム抽出例 33
表 10. 110 辞書登録単語例 34
表 11. MPi 辞書登録単語例 34
表 12. 辞書フィルタ評価結果 35
表 13. 辞書フィルタ抽出例 36
表 14. 人手による手がかり語追加例 37
表 15. 類語辞書を用いた追加例（「おかげ」から生成） 37
表 16. 200 文アノテーション結果 40
表 17. アノテーションデータに対する評価結果 41
表 18. 手がかり語追加による評価結果 42
表 19. 全文を対象とした評価結果 44
表 20. スニペットに対する手がかり語と出現パターンの対応関係 48
表 21. パターンと出現数の関係 49
表 22. 全文データに対する手がかり語と出現パターンの対応関係 50
表 23. ランダム 100 文中に存在する三つ組数の違い 50
表 24. 全文データに対する手がかり語と出現パターンの対応関係 52
表 25. 全文を対象とした評価結果 54

図目次

図 1 . TOBYO 辞典検索結果例 15
図 2. 提案システムの概要と抽出例 20
図 3. 提案システムの抽出パターン 22
図 4. ベースライン 26
図 5. 抽出規則 A 27
図 6. 抽出規則 B 27
図 7. 薬剤服用情報が出現するパターン 47
図 8. 構文解析結果 48

第 1 章

はじめに

1.1 研究の背景と目的

患者中心の医療が進められており，患者が治療に主体的に参加し，治療方法を選択することが求められている．しかし，病気や診断方法,治療方法が複雑であるため,突然病気を患った患者がこれらを正しく理解し,判断,決定することは容易ではない.患者の意思決定や治療に対する積極性を促し,QOL(生活の質)や健康状態に良い影響をもたらすためには,医療に関する必要な情報を取得し,理解,活用する力が患者自身に求められている．患者の医療情報入手先に対する調査[1]では，入手先として医療関係者が最も多く，信頼できる情報源であると評価されている．しかし，よりよい意思決定を行うためには，集団に対する効果等が確率的に表される「エビデンス情報」と，体験者の個別的な語りである「ナラティブ情報」の両方が必要である．患者は，同病者から体験談を得ることで，不安が軽減したり，励まされることがある[2][3]が，医療者がこのナラティブ情報を十分に提供するのは困難である．また，担当医師や看護師に質問しにくいといった調査結果[4][5]からも, 患者が情報不足感を抱いていることが見て取れる.

ところで,情報化社会の進展に伴い,誰でも容易にインターネット上に情報発信が可能となっている.そのため，インターネット上に存在する患者やその家族によって発信された情報が近年注目されている[6]．患者も医療情報を入手するために，医師だけでなく，インターネットなど多様な情報源を利用していることが確認されている[7]．実体験に基づく医療情報は患者の不安を軽減させたり,励みとなる可能があることに加えて，　大規模かつ即時的な情報は,既存の調査を上回る可能性がある.しかし,人手で膨大なデータから,求める情報のみを取捨選択し,収集するには多大な労力が必要である.また,人手による検索では,悪い面ばかり無意識に注目し収集してしまうといった問題が考えられる.したがって,本研究は,インターネット上から得られる患者の実体験情報を自動的,網羅的に収集,提示することで,患者の判断材料や励みとなるシステムの構築を目的としている.

1.2 目指すシステムの構想

システムは，患者にとって理解しやすい言葉で書かれた情報を，理解しやすい形で提供することが望ましいと考えている．また，筆者の性別や年齢，疾患名などを設定することで，ユーザに近い筆者の情報を検索することができれば，過去に存在した（もしくは存在している）ユーザと類似度の高い筆者が，どのような治療により，その後どのような体調の変化をたどったかを知ることができ，ユーザが治療方法を決定する際などに大きな判断材料となる．例えば，数日前に階段から落ちて足首がズキズキ痛み，骨折を疑っているユーザがいるとする．システムに「骨折」,「足首，ズキズキ」などを入力すると，骨折に関わる記事，足首がズキズキ痛むことが書かれている記事が検索される．そして，骨折に至った原因を記述した部分をリストアップして表示し，階段から落ちたことが原因で骨折する事例があったかどうかを検索することができたり，足首がズキズキした結果どのような診断を受けた事例があるかどうかなどを調べることができる．本研究ではこのようなシステムを実現することを目指している．

さらに，即時的な情報を用いることで，いち早く病気や副作用の流行を把握する助けとなる可能性がある．数年前，子宮頸がんワクチン「サーバリックス」の副作用が大きく取り上げらた．販売開始当時は，ワクチンの有効性に焦点が当てられ推奨されていたが，その後，ワクチンの副作用が重篤であり，発生率も高いことが明らかとなり，今では積極的な接種の呼びかけが中止されている．この事件の問題の一つに，ワクチンと歩行困難，死亡といった症状との関連性を認めるまでに時間を要した点が挙げられる．本研究により実現を目指すシステムでは，例えばこのような事例に対しても，子宮頸がんのワクチンを接種した患者が発信した情報をリアルタイムに収集し，接種後の体調の変化に関する情報を抽出，統計的に解析することで,「サーバリックス」と「死亡」もしくは「歩行困難」などの発症例は少ないが重篤な症状を関連性がある出来事として捉えて共有し，被害を最小限に抑えることも可能であると考えている．

加えて，未知の原因を発見できるシステムであることも求められる．例えば，ある薬剤を服用している複数の筆者の記事を追跡し，体調の変化を抽出，数値化し，時系列で表示する．横軸を経過年数，発病年を 0 として複数の患者の体調変化を表すグラフを同時に表示することで，比較が可能となる．さらに，悪化スピードなどでグループ化し，ブログの文章に出現した情報と照らし合わせることにより,「お酒を飲む」「睡眠時間が多い」など病気の進行に関わる様々な影響因子を発見できる可能性がある．

しかし, 前述したようなシステムを用いることにより, 悪い結果にばかり注目してしまい, 結果的に精神的に悪い影響を与える危険性が考えられる．したがって，ユーザがネガティブな情報に偏った情報収集をしてしまわないように，極性判定処理を行い，ポジティブな情報もネガティブな情報も偏りなく提示する必要があると考えている．また，システムを構築する際，書き手，読み手どちらにおいても問題となるのは，素人判断による危険性である．いずれにしても，ネットの情報を用いる際には，信頼性が大きな問題として挙げられる．ネッ

ト上に存在する医療情報の信頼性を判断する研究[8]も進められているが，著者は膨大な情報を利用し，統計的な傾向を把握することにより信頼性の低さを補うことができると考えている．

1.3 本研究の位置づけ

日本で使用されている医薬品は, 医師の診断により処方される医療医用医薬品が約 18,000 品，医師の処方無く薬局や薬店で購入できる一般用医薬品が約 12,000 品にものぼる．医薬品は治療において必要不可欠なものであるが，付随して起こる副作用が問題となっており，副作用を早期発見し，その対策を講じることが医療現場においては重要な課題である．そのため，医薬品と副作用の因果関係を推定する研究[9][10]が盛んに行われ，医薬品検索サービス[11][12]も多数存在する．これらが基にしているデータは，各医薬品の製造，販売元が作成した情報や医師や薬剤師から報告された情報である．しかし，市販後に医薬品と副作用の関連性が疑われ，副作用と立証されるケースもあるため，未知の副作用には対応しきれていないのが現状であり，医薬品の効能や副作用を正確に把握するためには可能な限り多くの情報を収集する必要がある．

医薬品の副作用に関する記述を抽出する研究[13][14]では，医療カルテを対象とした抽出を行っている．しかし，これらが抽出の対象としている副作用は，医師などの医療従事者によって副作用であると認められた事象についてのみである．薬品との関連性が実証されてはいないが，患者自身が感じている関連性のある変化も未知の副作用候補として抽出することは，他の患者にとってひとつの判断材料となり有用であると考えられる．

そこで，本研究ではリアルタイムに大規模データが取得可能なネット上に存在する情報に着目した．また，実体験に基づいた情報を収集するために，患者やその家族である闘病者によって書かれたブログ記事を抽出対象としている．患者自身の疾患や医療に対する認識や，ネット上に存在する情報の信頼性は不完全だが，大規模かつ即時的な情報は，既存の調査を上回る可能性がある．また，日本では個人が匿名で情報を発信することが多い[15]．匿名の情報の信頼性は確かではないが，匿名での情報発信では体裁を気にすることなく，身近な人には知られたくないような情報も公開することができる[16]．また，多数の情報を集めることにより，同様な情報の有無などから統計的に匿名情報の信頼性を把握することが可能であると考える．以上のことから，ウェブからは書き手の率直な意見や感想，悩みと言った情報が入手可能であり，膨大な情報を利用することにより匿名の信頼性の低さを補うことができると考えている．

本研究は闘病者によって書かれたブログ記事から，医療情報を検索できるシステムを構築することを最終的な目的としている．また，患者の分布傾向や治療法についての情報だけでなく，医師が認識していないような，疾患や医薬品と関連性のある事象を発見できるシステムを目指している．このような医療情報の提供は，患者にとって治療法などを決定する際の判断材料となり, ブログから得られたナラティブ情報は, 精神的な支えとなると考えられる．

このようなシステムを構築する第一段階として，薬剤の服用による効果や変化に関する記述を，ブログから抽出する手法について研究を進めてきた．意見の主要部分は（対象，属性，評価）の三つ組で表現できる[17]とし，ユーザの意見や評判を三つ組で抽出する研究が多くされている[18][19][20]．薬剤効果は「効く」「効かない」「眠れる」「辛い」といった一単語で極性をもつ単語を用いて表されることが多い．しかし，「なくなる」といった単語による評価表現では，「髪がなくなる」と「痛みがなくなる」のように対象によって極性が大きく異なる場合が存在する．このような単語の一致では対応しきれない問題を解決するため，本研究においても（薬剤，対象，効果）の三つ組を用い，対象と効果で対となる薬剤効果の抽出を行う．

1.4 本論文の構成と概要

本論文の構成を以下に示す．

第 2 章では，関連研究について説明する．

第 3 章では，本研究が対象とするデータについて述べる．

第 4 章では，システムの構成を述べた上で，処理過程を詳しく説明する．手がかり語と構文情報を用いたパターンマッチングにより抽出を行う．評価要素に着目した抽出を行わないため，正と負どちらの薬剤効果についても抽出することができる．また，文中に出現する薬剤名と対象，評価要素との構文的関係性を考慮した抽出を行うため，複数の話題を含むブログ記事からも薬剤効果の抽出が可能である．さらに，効果や様態を記述する際，日本語では多く用いられる擬態語，擬音語の抽出も可能であるという点に本研究の新規性が存在する．

第 5 章では，システムが抽出する三つ組で表される薬剤服用情報（薬剤名，対象，評価）のうち，対象，評価要素を特定するために設定したパターンの妥当性を調査する性能評価実験について述べる．

第 6 章では, システムの性能を改善するために行った, 三つの改善手法について説明する．一つ目は，提案したパターンに当てはまらない薬剤服用情報の抽出を可能にするため，比較手法の中で最も再現率の高かった手法と組み合わせた抽出を行い，再現率を向上させる手法である．二つ目は，適合率を向上させるために，対象要素に着目し，既存の文書から作成した対象用語辞書をフィルタとして用いる手法である．三つ目は，手がかり語リストを拡張する手法である．一つ目，二つ目の手法に関しては，評価実験を行い，その結果についても述べる．

第 7 章では，第５章とは異なるデータセット二種類について行った性能評価実験について述べる．一つ目のデータセットはブログのスニペット（検索結果として得られる要約文）から５人の評価者が作成したデータを元にしたものである．二つ目はブログ記事全文から作成したデータセットである．

第 8 章では，評価実験を行うためにブログから作成した薬剤服用情報の構文的特徴の分析結果と，薬剤添付文書に書かれた情報との差異について述べる．

第 9 章は結論であり，本研究で提案した手法をまとめ，性能評価実験によって得られた結果より，提案手法の有効性について評価する．さらに，今後の展望についても述べる．

参考文献

[1] 谷口綾, “一期的乳房再建術を受けた乳がん女性の思いと看護支援”, 看護教育 45(8):pp.660-663, 2004.

[2] 松本梓, 新垣円, 鄭迎芳, 井ノ口珠喜, 酒井梢, 塩田早恵, 高野学, 高橋都, “医療情報源としてのインターネットに対する乳がん患者の認識 ―メーリングリスト参加者を対象としたインターネットサーベイから―”, 医療情報学 25(3):pp.151-159, 2005.

[3] 前田優雅, 国府浩子, 藤井徹也, “治療中の乳がん患者に及ぼす同病者からの影響と関連する要因 ―乳がん患者会会員を対象として―”, がん看護 14(6):pp.711-716, 2009.

[4] Masahiko Tsuchiya and Sandra Horn, “An Exploration of Unmet Information Needs among Breast Cancer Patients in Japan: A Qualitative Study”, European Journal of Cancer Care, 18(2):149-155, 2009.

[5] 国府浩子, “初期治療を選択する乳がん患者が経験する困難”, 日本がん看護学会誌 22(2):14-22, 2008.

[6] 荒牧英治, 荒木健治, 石井保志, 上松幸一, 北嶋志保, 熊谷晋一郎, 櫻井公恵, 瀬戸山陽子, 高橋由光, 冨井美子, 堀里子, 西舘澄人, 森田瑞樹, 矢島弘士, 和田恵美子, 則のぞみ, 島本裕子, “患者/クライ エント/当事者による記録を活かす意見交換会の活動報告”, 医療情報学, 第 34 巻 4 号, pp.197-200, 2014.

[7] 瀬戸山陽子, 中山和弘, ”乳がん患者の情報ニーズと利用情報源,および情報利用に関する困難 ―文献レビューからの考察―“, 医療と社会(0916-9202)21 巻 3 号 pp.325-336, 2011.

[8] 横井遥, 福田里砂, “インターネット上の医療情報の信頼性の検討: 開腹手術に伴う術後疼痛に関する情報について”, 医療情報学 33.1, 49-59, 2013.

[9] David M. Green and John A. Swets, “*Signal detection theory and psychophysics*“, Vol. 1. New York: Wiley, 1966.

[10] 渡邉裕之, 松下泰之, 渡辺篤, 前田敏郎, 温井一彦, 小川嘉正, 澤順悟, 前田博, “重要な安全性情報を早期に検出する仕組み: シグナル検出の最近の手法について”, 計量生物学 25.1: pp.37-60, 2004.

[11] 独立行政法人 医薬品医療機器総合機構 医薬品医療機器情報提供ホームページ. [http://www.info.pmda.go.jp/index.html (cited 2014-December-18)].

[12] ファーマ・フレンド（小林薫），お薬 110 番.
[http://www.jah.ne.jp/~kako/frame_dwm_search.html (cited 2014-December-18)].
[13] 大熊智子，三浦康秀，外池昌嗣，増市博，篠原(山田)恵美子，荒牧英治，大江和彦，“医薬品の副作用調査を目的とした統合的言語処理システム”，言語処理学会第 17 回年次大会発表論文集，pp.85-88，2011.
[14] 篠原恵美子，服部圭悟，三浦康秀，外池昌嗣，大熊智子，増市博，荒牧英治，大江和彦，“構文パターンに基づく薬剤副作用情報の自動抽出と評価”，医療情報学 31(Suppl.)，pp.521-524，2011.
[15] 吉田等明，村上武，和崎宏，五味壮平，“オンラインでの匿名性と倫理観”，コンピュータ & エデュケーション 24 ，2008.
[16] 折田明子，三木草，小川美香子，”発信しづらい情報交換における匿名性の効果 ―ダイエット食品クチコミ調査から―”，情報社会学会 Vol.2 No.2，pp.114-127，2007.
[17] 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一，“テキストマイニングによる評価表現の収集”，情報処理学会研究報告，2003-NL-154，pp. 77–84，2003.
[18] 酒井義和，荒木健治，“反対語を利用した文脈依存評価表現の感情極性判定”，電子情報通信学会論文誌 D，情報・システム J93-D(9)， pp.1778-1789，2010.
[19] 土田正明，水口弘紀，久寿居大，“評判検索のための対象，属性，評価の 3 項関係のランキング法”，第２２回人工知能学会全国大会，2P2-05，2008.
[20] 杉木健二，松原茂樹，“消費者の意見に基づく商品検索”，情報処理学会論文誌，Vol.49，No.7，pp.2598-2603，2008.

第 2 章

関連研究

2.1 副作用の関連性推測システム

薬剤の未知の副作用はシグナル検出を用いて，医師や薬剤師からの報告を基に推測される[1][2]．シグナル検出とは，今まで知られていない，または立証されていない医薬品と副作用の関連性の可能性を統計的に検出する手法である．しかしシグナル検出は，症例報告に依存するため，社会的影響を受けて報告件数に偏りや変動を生じる可能性があるという問題がある．また，報告されないような情報も全て取得するのは実質不可能である．よって，本研究では，医薬品情報をより多く取得するため，ネット上に存在する患者やその家族である闘病者によって書かれたブログ記事から情報を取得したいと考えている．

患者発信のネット上に存在する情報を利用することで，リアルタイムに，新しい，医師では把握しきれていないような情報も取得可能であると考える．また，本論文では，薬剤名を含む文から特定表現を用いて薬剤と関連性があると判断された薬剤情報の抽出を行う．今後は今まで抽出された対象や評価表現（効果）を知識源として利用し，薬剤名を含まない文からも，薬剤との関連性を推測した効果の抽出ができるシステムの構築を目指している．

2.2 医療情報の入手方法

2.2.1 患者の記録を共有する試み

同病者の情報交換手段として友の会への加入などが一般的に考えられるが，地理的な問題や自分が病気だと知られたくない，敷居が高いなどの理由からインターネットを利用した情報提供も行われている．患者のオンラインコミュニティーである「PatientsLikeMe[3]」はその先駆けであり，患者自身の病状を記録したり，患者同士での意見交換を行っている．日

本国内でも，患者 SNS の「LifePalette[4]」や「TOBYO[5]」が存在している．このようなブログサイトの情報は，匿名で気軽に発信できるため信頼性は低い．しかし，1 に述べたように，匿名だからこそ発信できる情報や，直接当事者から手に入れることは難しいような情報を多数集めることができる［6]．テキストだけでなく，乳がんや認知症などのナラティブ（語り）情報を交換するサイトとして「DIPEx[7]」があり，約 150 人の患者がインターネット上に顔を出して体験談を語っている．また，「健康情報棚プロジェクト[8][9]」では，闘病に関する書籍を疾患ごとに分類し，病院や各施設に闘病記の棚を設置することで，ネットを使えない患者に対しても闘病記を提供することを目指している．

このような患者に関連する様々な分野の現状や課題について意見を交流することを目的とした活動が，「〈患者〉による記録を活かす意見交換会[10]」であり，著者もその一員である．

2.2.2 医薬品検索サービス

医薬品の効能や副作用などを検索できる医薬品検索サービスは，企業や法人などにより多数提供されており，医薬品添付文書検索サービス[11]がそのひとつである．医薬品添付文書とは，各医薬品の製造，販売元が作成を行う，医薬品の効能や成分，副作用等が記載されている文書である．医薬品添付文書検索サービスでは，薬剤名から添付文書の検索，警告情報，禁忌情報，副作用情報の検索が可能である．このような医薬品検索に関するサービスは，既知の情報を用いているため，関係している可能性のある未知の副作用については検索することができない．

本研究では，患者によって書かれたブログを対象とすることにより，未知の副作用や効能を発見するために必要な情報も取得可能であると考える．

2.3 副作用記述箇所特定システム

薬剤と未知の副作用の関連性を推測するとき，医薬品検索システムを構築する際には，報告や添付文書から薬剤効果に該当する事象箇所の特定が必要となる．しかし，自由に記述された文章から，そのような特定作業を人手で全て行うには多大な労力が必要である．そこで，近年，自然言語処理技術を中心とした情報技術の医学的応用研究が進められている．

NTCIR-10 医療言語処理（MedNLP）パイロットタスク[12]では医師によって書かれた医療文書を対象とし，患者の個人情報や医療情報の抽出する技術の研究を進めている．最終的には，診断や治療を進める際，医療関係者に役立つシステムを目指している．このパイロットタスクの主なタスクは，匿名化タスク（年齢や性別，個人名，日時などの記述部分にタグ付けを行うタスク）と，病状と診断タスク（症状や診断に関する部分を判別，タグ付けするタスク）である．

大熊ら[13]は退院時サマリーから副作用について記述されている箇所を特定するシステム

を作成した．医薬品と副作用症状候補間の距離や構文情報，医療表現などを素性としてサポートベクターマシーン（SVM）[14]を用いた特定を行う．

大熊ら[12]の研究のように，副作用関係抽出や医療用語（病状，診断に関する用語）の特定は SVM によりされることが多いが，SVM の精度は学習データに依存するため，自然言語で書かれた文書に対する十分な抽出精度が得られていない．それに対し篠原ら[15]は SVM を用いず構文パターンを用いた薬剤副作用情報の抽出手法を提案した．

電子カルテは ICD10[1]や MeSH[2]の用語を利用して記述されることが多く，これらの専門用語辞書を利用した医療情報の特定が進められている．しかし，本研究で対象としているブログ記事の筆者は専門家ではないため使われる用語も一般的なものであり，専門用語辞典を用いた抽出では不十分であると考えられる．また，これらの研究により「頭痛」のような専門用語の症状名は抽出，特定できても「頭が痛む」のように対象と評価の対で記載された症状については対象としていないことも問題点として考えられる．

そこで本研究では，専門用語辞書を用いず，対象と評価の対で意味をなす薬剤情報の抽出を行う．また，専門家ではない患者によって書かれた文書から高精度に抽出を行うために，一般人へ向けて作成された薬に関する HP を利用したフィルタリング方法の提案を行う．

2.4 商品や場所についての評価を抽出する研究

利用者から得られる医薬品情報を，薬に対する評価情報・意見情報と捉えることもできる．

立石ら[16][17]の研究では，Web ページを対象としており，あらかじめ用意した評価表現辞書を参照して，近接演算により対象物と評価表現を含む一定範囲の文字列を，対象に対する意見として抽出する．その後，パターンマッチングを用いて構文的に意見を表しているかどうか判定する．抽出に用いる評価表現辞書は，人手で作成したものであり，ドメインに大きく依存するため，ドメインが変わるたびに辞書を作り直さなければならず，辞書作成のコストが高いという問題がある．さらに，彼らの研究は，評価表現にラベル付した「肯定・否定」によって，抽出した意見の極性判定を行う．そのため，「大きい」「小さい」といった対象や属性によって極性が左右される可能性のある単語が評価表現となる場合を考慮していない．松澤ら[18]の研究では，利用者が“肯定”・“否定”・“無属性”のラベルを付与した評価表現辞書を作成し，その評価表現と利用者が入力したキーワードが，あらかじめ用意したパターンマッチングルールを満たしている場合，キーワードの前後200文字を意見として抽出する．抽出に要する辞書は利用者が作成するため，利用者の嗜好に合わせた情報抽出が可能となるが，具体的な目的なしに，漠然とあるキーワードについての意見抽出を行いたい場合辞書の作成は利用者にとって大きな負担となる．

意見の主要部分を（対象，属性，評価）の三つ組で表現できるとし[19]，ある対象につい

[1] WHO の国際疾患分類第 10 版分類コード表
[2] Medical Subject Headings：米国国立医学図書館 NLM の文献検索用医学用語シソーラス

ての意見や評判を三つ組で抽出する研究も多くされている．

酒井ら[20]の研究では，口コミ掲示板を対象とし，辞書を用いた評価情報の抽出手法を提案している．評価表現辞書に含まれる単語を含む一文を意見文として，「(対象）の（属性）が（評価表現)」のパターンを満たし，かつ「(対象）の」が「(属性）が」に係り，「(属性）が」が「(評価表現)」に係るもののみを抽出している．土田らはブログから対象・属性・評価で表される評判情報を，確信度を付与してランキングする手法を提案した[21]．評価表現辞書に含まれる表現とセットになる属性単語を SVM により抽出し，単語間距離により対象と属性を対応付ける．杉木ら[22]は，辞書を用いず，自然言語で記述された検索クエリに対応する商品を，商品レビューテキストとクエリを意見対象・項目・値の三組へと変換することで検索するシステムを作成した．

商品や場所についての評価を抽出する研究では，評価の対象が明らかであり，使われる評価表現も限られているため，評価表現辞書を用いた抽出が可能である．また，目的や対象が明確である商品レビューなどとは異なり，本研究が対象とするブログは，薬剤評価のためだけに書かれたものではないため，日々の記録から医薬品に関する評価部分をいかに抽出するかが課題となる．また，薬剤について記述する際使われる評価表現は，商品レビューなどで用いられるものだけではなく，「パンパン」や「キリキリ」といった擬態語も多い．擬態語など口語で多く用いられる表現を網羅した評価表現辞書は存在しなく，医療用の評価表現辞書を作成するには多大な労力が必要となる．

2.5 既存研究の問題点と提案システムについて

既存研究の問題点として，大きく３点が考えられる．大熊ら[14]や篠原ら[15]の抽出対象は薬剤の負の影響である副作用のみであり，薬剤によって引き起こされる正の評価（効果）を対象としていない．また，商品の意見を抽出する研究[20][22]では，対象となるテキストが商品レビューなどであり，目的や対象が明らかなことが多く，ブログのような話題が定まっておらず複数の話題を含むテキストに適用することは難しい．さらに，ブログ記事は話し言葉で書かれていることが多く，主語や助詞の抜け落ちが多いことが挙げられる．

本研究で提案するシステムは，そのうち主に２つの問題点を解決する．第一に，本研究の提案システムでは文の構造に着目し，評価要素に着目した抽出を行わないため，正と負どちらの薬剤効果についても抽出することができる．第二に，文中に出現する薬剤名と対象，評価要素との構文的関係性を考慮した抽出を行うため，複数の話題を含むブログ記事からも薬剤効果の抽出が可能である．また，効果や様態を記述する際に，日本語では多く用いられる擬態語，擬音語の抽出も可能であるという点で新規性がある．しかし，くだけた表現で記述されたテキストから正確に情報を取得するためには，助詞の抜け落ちや照応解析の問題を解決しなければならず，これらは今後の課題である．

参考文献

[1] David M. Green and John A. Swets, “*Signal detection theory and psychophysics*.” Vol. 1. New York: Wiley, 1966.

[2] 渡邉裕之, 松下泰之, 渡辺篤, 前田敏郎, 温井一彦, 小川嘉正, 澤順悟, 前田博, “重要な安全性情報を早期に検出する仕組み: シグナル検出の最近の手法について,” 計量生物学 25.1: pp.37-60, 2004.

[3] PatientsLikeMe. [http://www.patientslikeme.com/ (cited 2014-December-18)].

[4] LifePalette. [http://lifepalette.jp (cited 2014-December-18)].

[5] TOBYO::日本の闘病記 43000. [http://www.tobyo.jp/ (cited 2014-December-18)].

[6] 折田明子, 三木草, 小川美香子, ”発信しづらい情報交換における匿名性の効果 ―ダイエット食品クチコミ調査から―,” 情報社会学会 Vol.2 No.2, pp.114-127, 2007.

[7] 認定特定非営利活動法人（NPO）健康と病いの語りディペックス・ジャパン. [http://www.dipex-j.org/ (cited 2014-December-18)].

[8] 北澤京子, 石井保志, “ 患者・家族への情報提供を模索する: 多職種協働による健康・医療情報の社会提言 (< 特集> 価値観の交差点),” 情報の科学と技術, 56(9), 406-411, 2006.

[9] 和田恵美子, “「闘病記文庫」 は患者・医療者に何をもたらすか―健康情報棚プロジェクトの多職種協働活動を通して:―健康情報棚プロジェクトの多職種協働活動を通して,” 情報管理 49.9: 499-508, 2006.

[10] ＜患者＞による記録を活かすラウンドテーブル [mednlp.jp/patients_life/ (cited 2014-December-18)].

[11] 独立行政法人 医薬品医療機器総合機構 医薬品医療機器情報提供ホームページ. [http://www.info.pmda.go.jp/index.html (cited 2014-December-18)].

[12] NTCIR-10 “Medical Natural Language Processing (MedNLP)” Pilot Task. [http://mednlp.jp/medistj-en/ (cited 2014-December-18)].

[13] 大熊智子, 三浦康秀, 外池昌嗣, 増市博, 篠原(山田)恵美子, 荒牧英治, 大江和彦, “医薬品の副作用調査を目的とした統合的言語処理システム,” 言語処理学会第 17 回年次大会発表論文集, pp.85-88, 2011.

[14] Corinna Cortes and Vladimir Vapnik, “Support Vector Networks”, Machine Learning, Vol. 20, pp.273–297, 1995.

[15] 篠原恵美子，服部圭悟，三浦康秀，外池昌嗣，大熊智子，増市博，荒牧英治，大江和彦，“構文パターンに基づく薬剤副作用情報の自動抽出と評価”，医療情報学 31(Suppl.)，pp.521-524，2011.

[16] 立石健二，石黒義英，福島俊一, “インターネットからの評判情報検索”，人工知能学会誌 19 巻 3 号，pp.75–82，2004.

[17] 立石健二，石黒義英，福島俊一, “インターネットからの評判情報検索”，情報処理学会研究報告，自然言語処理研究会報告 2001(69)，pp.75-82，2001.

[18] 松澤祐太，鈴木裕利，石井成郎，小出周之, “ユーザ嗜好に着目した評判情報の抽出手法に関する提案と評価”，情報処理学会研究報告告，情報システムと社会環境研究報告 2011-IS-118(1)，pp.1-8，2011.

[19] 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一, “テキストマイニングによる評価表現の収集”，情報処理学会研究報告，2003-NL-154，pp. 77–84，2003.

[20] 酒井義和，荒木健治, “反対語を利用した文脈依存評価表現の感情極性判定”，電子情報通信学会論文誌 D，情報・システム J93-D(9)，pp.1778-1789，2010.

[21] 土田正明，水口弘紀，久寿居大, “評判検索のための対象，属性，評価の 3 項関係のランキング法”，第２２回人工知能学会全国大会，2P2-05，2008.

[22] 杉木健二，松原茂樹, “消費者の意見に基づく商品検索”，情報処理学会論文誌，Vol.49，No.7，pp.2598-2603，2008.

第 3 章

対象とするブログとその傾向

3.1 対象とするブログ

ブログは書き手によって様々な記述方法が取られ，形式も決まっていないため，必要とする情報が書かれた部分を自動判定することは難しい．そのため，本研究では，株式会社イニシアティブが運営する，闘病者[3]のための闘病体験情報を共有するサービスである TOBYO[1] に登録された闘病ブログを対象とする．TOBYO に登録された闘病ブログには，ユーザ登録の際入力した病名，性別，発病時の年齢などがタグ付してある． TOBYO に登録されていない闘病ブログも多数存在するが，それらのブログにはこのようなタグ付は一切されておらず，ブログがどのような人物によって書かれたかなどの情報を抽出する技術も必要となるため，本研究ではあらかじめ情報が付与されている TOBYO に登録された闘病ブログを対象とした．また，TOBYO 辞典を用いることにより，闘病者によるブログが収集可能であり，このようなノイズの少ないデータは本システムのアルゴリズム考案に有効であると考える．今後さらに多くの闘病体験情報を収集し，個人に適した情報提供を行うためには，ブログ筆者や対象者の情報抽出も一つの課題である．

また，TOBYO では TOBYO 辞典を用いて，病名・薬品名・治療方法・医療機関名など固有名詞をキーワードにして闘病ブログを検索できる．実際の検索結果例を図 1 に示す．本研究では，296 種の薬剤名について TOBYO 辞典を用いて検索し，検索に用いられた薬剤名（対象薬剤名）の服用情報を，検索結果として得られるブログの要約文であるスニペットから抽出する．

[3]患者とその家族・友人

図 1．TOBYO 辞典検索結果例

3.1.1 登録されたブログ

TOBYO に登録されたブログのうち, ランダムに選出した 100 件について調査したところ, 利用されているブログサイトの内訳は以下の通りであり，登録されたブログの形式は多様であることが確認できる.

- アメーバブログ[2]　76 サイト
- FC2 ブログ[3]　7 サイト
- ライブドアブログ[4]　7 サイト
- seesaa ブログ[5]　3 サイト
- Yahoo!ブログ[6]　3 サイト
- エキサイトブログ[7]　2 サイト
- ココログ[8]　2 サイト
- その他（自作 HP）　1 サイト

また，これらのブログサイトのプロフィール情報を調査したところ，本名の可能性があるファミリーネーム・ファーストネームから成る名前が筆者名として用いられているサイトは

1 つのみであり，残り 99 サイトの筆者の名前はすべてニックネームであった．プロフィールに記述された情報の真偽は定かではないが，性別は 72 サイト（男性 8，女性 64），地域は 35 サイト，年齢は 34 サイト，職業は 30 サイトで公開されていた．この結果からも，日本語で記述されたブログ記事は匿名性が高いことがわかる．闘病ブログを TOBYO に登録する際，著者の病名，性別，発病時の年齢などが入力されている．

TOBYO に登録されていない闘病ブログも多数存在するが，それらのブログには性別などのタグ付がされておらず，ブログがどのような人物によって書かれたかなどの情報を抽出する技術も必要となるため，本論文ではあらかじめ情報が付与されている TOBYO に登録された闘病ブログを対象とした．また，TOBYO 辞典を用いることにより，闘病者によって書かれたブログが収集可能であり，このようなノイズの少ないデータは本システムのアルゴリズム考案に有効であると考える．

3.2 TOBYO データベース

本研究では，TOBYO[1]に 2011 年 11 月時点で登録されていた闘病ブログ 32,679 サイトを対象とする．TOBYO 辞典で 296 種の薬剤名について検索をかけ，得られた闘病ブログスニペットと，その検索元となった薬剤名（対象薬剤名），ブログにあらかじめタグ付けされた病名，性別，発病時の年齢，ブログタイトル，ブログ記事タイトル，URL をセットにしてデータベースを作成した．データベースには，検索により得られた 10391 データのうち，スニペット内容が重複したデータを除くと全 80745 データが格納されている．

闘病ブログは随時更新されているが，TOBYO 辞典で検索結果を表示するまでにはタイムラグがあり，すべての検索結果をその都度リアルタイムに入手することは，多大な時間が必要となる．そのため，抽出手法の提案を目的とした本研究では，作成したデータベースを利用する．

3.3 TOBYO スニペットの傾向

3.3.1 薬剤名とその評価の出現位置

TOBYO 辞典によりランダム収集された 1000 スニペットを対象とし，対象薬剤名とその服用による変化や効果，対象薬剤に対する意見などの記述が闘病ブログではどのように出現する傾向があるかを調査した．それらの内容には，三つ組の服用情報として表すことができないものも含む．

表 1. 薬剤名とその効果の出現位置例

関係性	薬剤名	スニペット
同一文中	プレドニン	残暑お見舞い申し上げます。暑すぎです…、MSの方々ウートフに苦しんでおられるかと思います。心中お察しいが変わりないかどうか見てもらう以外に、もう一つあった。プレドニンを飲み始めてから、どうもコレステロールの数値が高いのでていた栄養指導を、仕方なく受けるためだ。先生は、どうもプレドニンではなく、食事の影響だと思っているのだ。ところで、今回
前後一文	漢方	漢方を飲んで2ヶ月目。すぐに効果があったのは・・・・・オリモノが多い→減りましたね〜。これが一番最初が多いのか少ないのかってわからなかったんですよね。でも漢方を飲んですぐに量が減ったんです。以前はオリモノシートが年中いつでも乾くのは体質だとばかり思っていたんですが、漢方を飲み始めて1ヶ月たち、途中経過で薬剤師さんに『唇の乾き
その他	リュープリン	最近、寝ている間にやたらと体がほてって目が覚めて熟睡できない。肩凝りもひどくなったし、心臓が突然バ読んで理由がわかった。手術前に生理を止めるために打ったリュープリンの副作用みたい。ほとんどの症状に心当たりがあったので、

対象薬剤名とその服用による意見を発見できたものが 254 スニペットあり，そのうち同一文中に薬剤名も存在したものは 199 スニペットであった．また，対象薬剤名を含む文の前後一文中に意見が存在したものは 35 スニペット，同一文中でも前後一文中でもないがスニペット中には意見が存在しているものは 20 スニペットであった．これら意見の一部を表 1 に示す．

このことから，TOBYO スニペット中では，薬剤服用による意見を含むスニペットの約 78.3%で，同一文中に対象薬剤名を記載していることが確認された．そのため，本研究ではスニペット中薬剤名を含む文を対象とした薬剤服用情報の抽出を行う．薬剤名を含まない文に対する関連性の有無の判定や，情報抽出については今後の課題とする．

3.3.2 手がかり語

新たに TOBYO 辞典によりランダムに収集された 2,000 スニペットを対象とした．薬剤名と，（薬剤，対象，効果）の三つ組で表すことのできる服用情報が同一文中に存在する文を人手で調査し，それぞれの三つ組要素をタグ付した．その結果，服用情報を含む参考データとして 181 文を入手した．

抽出した 181 文の一部を表 2 に示す．表 1(1)〜(3)にあるように，服用情報を記載した一文中で，対象薬剤名の後に特定の服用・使用表現（表 2 の下線部）を使用した後，対象部位や性質の変化を記述している文が 138 文（76.2%）存在した．138 文中に存在した服用・使用表現を人手で収集し，取得した 32 語の名詞・動詞を手がかり語とする．なお，動詞は

活用により表層形が異なるため，原形を登録している．抽出した 32 語の手がかり語として用いる服用・使用表現を表 3 に示す．

表 2. 薬剤情報としての三つ組を含む文の例

(1) それからプレドニンを飲み始めたのですが、プレドニンの副作用で骨が弱くなり
(2) 私もおタキ様（タキソテール）の後は、肩や背中がすごく凝ります
(3) 現在の発作は、アレビアチンの注入時間前後(6時、14時、22時)に痙攣が出現している
(4) 多分ジプレキサで頭がボケ気味なんだと思います

表 3. 手がかり語例

副作用	飲む	使用	切る	処方	開始
おかげ	始める	服用	止める	増量	点滴
せい	増える	注入	やる	減量	終了
影響	打つ	投与	使う	注射	後
戻す	変える	変更	減らす	吸入	内服
入る	試す				

3.3.3 対象と効果をつなぐ助詞

2.3 により得られた服用情報を含む 181 文を対象とし， 三つ組要素である（対象）と（効果）を含む文節間が，どのような助詞を用いて関係付けられているかを著者が調査した．その結果を表 4 に示す．例えば，表 2(1)の文では，「骨が」が「弱くなり」に助詞「が」によって関連付けられている．同様に表 2(2)〜(4)においても助詞「が」が（対象）と（効果）を含む文節の関連付けに用いられている．

表 4. 対象要素と効果要素間に存在する助詞

助詞	出現回数	割合 (%)
が	104	57.5
は	23	12.7
を	20	11.0
に	7	3.9
にも	4	2.2
も	2	1.1
(その他)	3	1.7
(なし)	18	9.9

参考文献

[1] TOBYO::日本の闘病記 43000. [http://www.tobyo.jp/ (cited 2014-December-18)].
[2] アメーバブログ（アメブロ）. [http://ameblo.jp/ (cited 2014-December-18)].
[3] FC2 ブログ. [http://blog.fc2.com/ (cited 2014-December-18)].
[4] ブログならライブドアブログ(livedoor Blog).
[http://blog.livedoor.com/ (cited 2014-December-18)].
[5] Seesaa ブログ. [http://blog.seesaa.jp/ (cited 2014-December-18)].
[6] Yahoo!ブログ. [http://blogs.yahoo.co.jp/ (cited 2014-December-18)].
[7] エキサイトブログ. [http://www.exblog.jp/ (cited 2014-December-18)].
[8] ブログサービス ココログ - 無料ブログ:@nifty.
[http://www.cocolog-nifty.com/ (cited 2014-December-18)].

第 4 章

提案システムの概要

本研究では患者やその家族によって書かれたブログ記事のスニペットから,（薬剤名, 対象, 効果）の三つ組で表される薬剤の服用情報を，手がかり語と構文情報を用いて抽出する手法を提案する．ここで，対象とは薬剤名の服用もしくは使用により変化・効果が引き起こされた，あるいは引き起こされ得る部位や性質，効果とは評価表現を含む効果・変化のさまを示す表現や態度を示している．

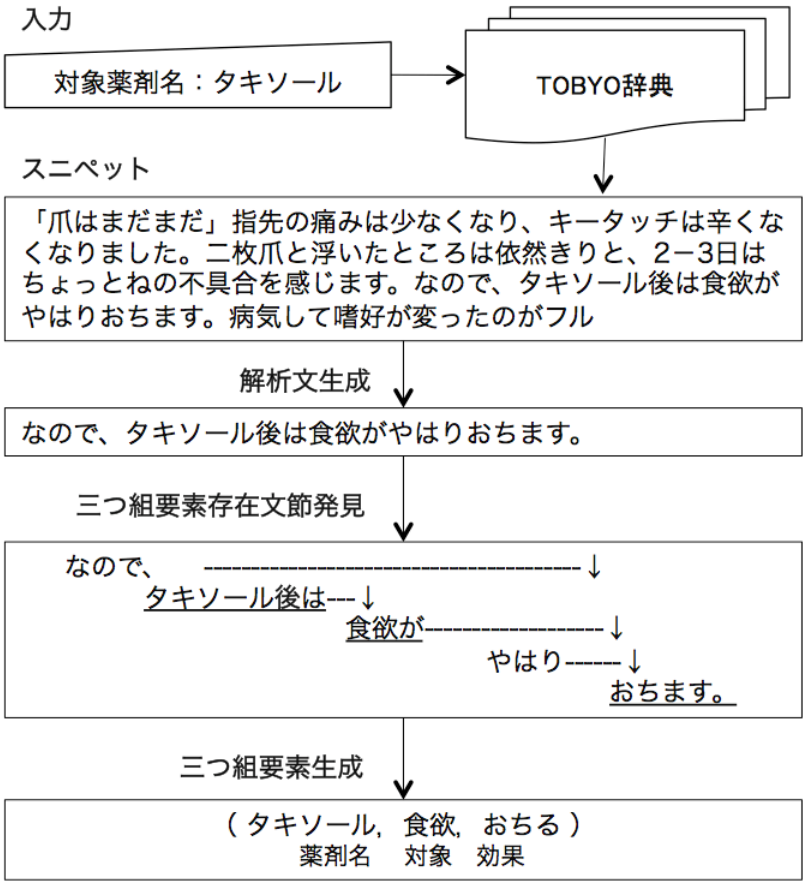

図 2. 提案システムの概要と抽出例

システムの構成要素である解析文生成部，三つ組要素存在文節発見部，三つ組要素生成部について以下で説明する．本研究の既存研究との相違点は，我々のシステムは構文情報と特定の手がかり語を用いた三つ組抽出を行うということである．システムの概要と抽出例を図 2 に示す．

4.1 解析文生成

TOBYO 辞典を用いて取得されたスニペットを，４ステップで変換する．まず，句点を区切り文字として文単位に区切り，1 文中に対象薬剤名を含む文に対し，服用情報を抽出する．単語情報の参照先である IPA 辞書[4]に存在しない薬剤名が誤って分割され，構文解析ミスにつながることを防ぐため，解析を行う際には薬剤名を 1 単語の固有名詞「MEDICINE」に置き換える．この置き換え処理は，296 種の薬剤名リストに基づき自動で行っている．なお，この処理に形態素解析結果は用いていない．

また，構文解析ミスを防ぐために，対象薬剤名を含まない括弧を削除する．この結果，表 1（3）の「(6 時、14 時、22 時)」は削除され，（2）の「(タキソテール)」は残される．対象薬剤名が含まれていなくても，括弧中に重要な情報が存在する可能性があるため，削除しても差し支えない括弧かどうかの判定をしたのち削除することは今後の課題である．

4.2 三つ組要素存在文節の発見

経験的に作成した図 3 に示す手がかり語を用いた係り受けパターンにより，三つ組の要素が存在する文節の位置を特定する． 図中の長方形は文節を，矢印は係り受け関係を示しており，いずれも CaboCha[1]による解析の結果得られたものである． 抽出ノイズを減らすため，ブログで服用表現が出現するときの特徴を用い，各文節末尾に付属する助詞の条件付けを行う． 対象要素存在文節と効果要素存在文節をつなぐ際，助詞「が」「は」「を」が用いられる割合が合わせて 81.2%と高かったため, それら三つの助詞に限定した. また, 格助詞 (例：で，から，より）は動作・状態の継続や経過に関する意味を持ち，接続助詞（例：が，ので，て，けれども）は原因・動機などの意味を持つため，手がかり語存在文節の最後尾単語の条件とした．

三つ組要素存在文節を発見する順序は「薬剤名」,「手がかり語」,「効果」,「対象」要素が含まれる文節の順である．まず，296 種類の薬剤名が含まれる文節を探し，薬剤名のすぐ後ろ，もしくは係り先の文節に **3.3.2** で定めた手がかり語が存在するかどうかを判定する．存在し，かつ手がかり語が含まれている文節の末尾助詞が格助詞，もしくは接続助詞である場

[4] MeCab[2]

合，手がかり語存在文節の係り先文節に「効果」要素が含まれていると判断する．その後，「効果」要素存在文節に最も近くで「が，は，を」を文節の末尾助詞に用いて係っている文節を，「対象」要素が含まれている文節と判断する．

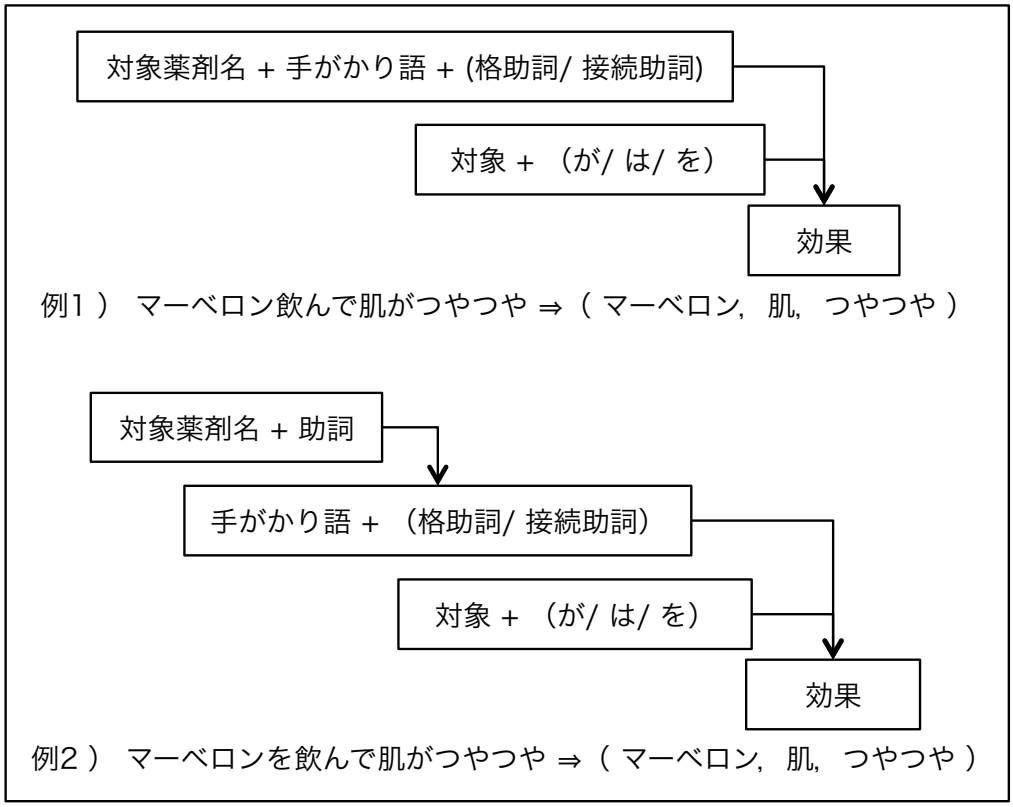

図 3．提案システムの抽出パターン

4.3 三つ組要素生成

三つ組要素が存在すると判定された文節から三つ組要素を生成する. 具体例を表 5 で示し, それぞれについて説明する.

表 5. 三つ組要素生成例

(1)	リウマトレックスを減らすと肝機能が改善	（リウマトレックス，肝機能，改善）
(2)	グリベックを飲み始めてから体毛に白髪が増えたという意見が結構	（グリベック，白髪，増える）
(3)	リオレサールの服用によって筋肉の緊張をほぐす	（リオレサール，筋肉の緊張，ほぐす）
(4)	リンデロンで右肘は曲がらなくなった	（リンデロン，右肘，曲がらない）

（1）助詞が出現するまでの自立語を連結させて要素とする.

（2）動詞は原形で抽出する.

（3）三つ組要素存在文節が連体化の助詞「の」によって修飾されている場合は，その文節を修飾節として付与する.

（4）「右」と「肘」のように連続する名詞はつなげる．文節中に打ち消しの意味を持つ「ない」が奇数個存在したときには, 必要に応じて要素の最後尾形態素の活用を変化させ,「ない」を付加し否定形にする．否定表現について前述した 181 文の調査を行ったところ,「ない」を用いた否定の評価表現が 3，その他の否定表現は用いられていなかった．そのため今回は,「〜ません」や「減量」といった手がかり語による意味の反転は考慮していないが，今後の課題とする.

参考文献

[1] CaboCha: Yet Another Japanese Dependency Structure Analyzer.
[http://chasen.org/taku/software/cabocha (cited 2014-December-18)].
[2] MeCab: Yet Another Part-of-Speech and Morphological Analyzer.
[http://mecab.sourceforge.net/ (cited 2014-December-18)].

第 5 章

評価実験

提案手法とベースライン手法，それぞれによって自動抽出された三つ組が，薬剤名を含む 1 文から読み取れる服用情報として適切であるかどうか，を著者が判断した．

5.1 データ

TOBYO 内にある検索ツールである TOBYO 辞典を用いて検索された，闘病ブログのスニペット 2,000 件中，一文中に薬剤名を含む重複無し 2,369 文から抽出を行う．これらの文に対し，あらかじめ著者が人手で服用情報の三つ組みを抽出したところ，340 組の服用情報を抽出できた．これを正解データとする

5.2 評価に用いる式

評価には式(1),(2),(3)で定義される適合率，再現率，F 値を用いる．

$$\text{適合率} = \frac{\text{正解抽出三つ組数} \times 100}{\text{抽出三つ組数}} \quad [\,\%\,] \tag{1}$$

$$\text{再現率} = \frac{\text{正解抽出三つ組数} \times 100}{\text{正解データ数}} \quad [\,\%\,] \tag{2}$$

$$\text{F 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \tag{3}$$

5.3 比較対象

図 4 に示す 3 つの抽出手法をベースラインとした．ベースライン 1 は既存手法[1]でレビューなどから意見を抽出する際用いられていたパターンである．また，既存手法[2]では意見抽出の際，評価表現に着目した抽出がされているため，ベースライン 2， 3 では一般的な評価表現辞書である EVALDIC_ver1.0.1[3]を用いたパターンマッチングにより抽出する．辞書には 5,234 単語が登録されている．なお，これらの抽出手法は三つ組要素存在文節の発見部分で比較対象として用い，解析文生成，三つ組要素生成部分では提案手法と同じ処理を行う．

また，提案手法の助詞の条件付けの有効性を確認するために，図 5，図 6 で示す抽出規則 A,B を用いた抽出も行う．抽出規則 A では，提案手法の手がかり語に後続する助詞の条件をつけない．抽出規則 B では，対象要素候補が存在する文節の最後尾の助詞として，予備調査により用いられやすいことが確認された助詞上位 4〜6 位であった「に，も，にも」も条件に加えている．

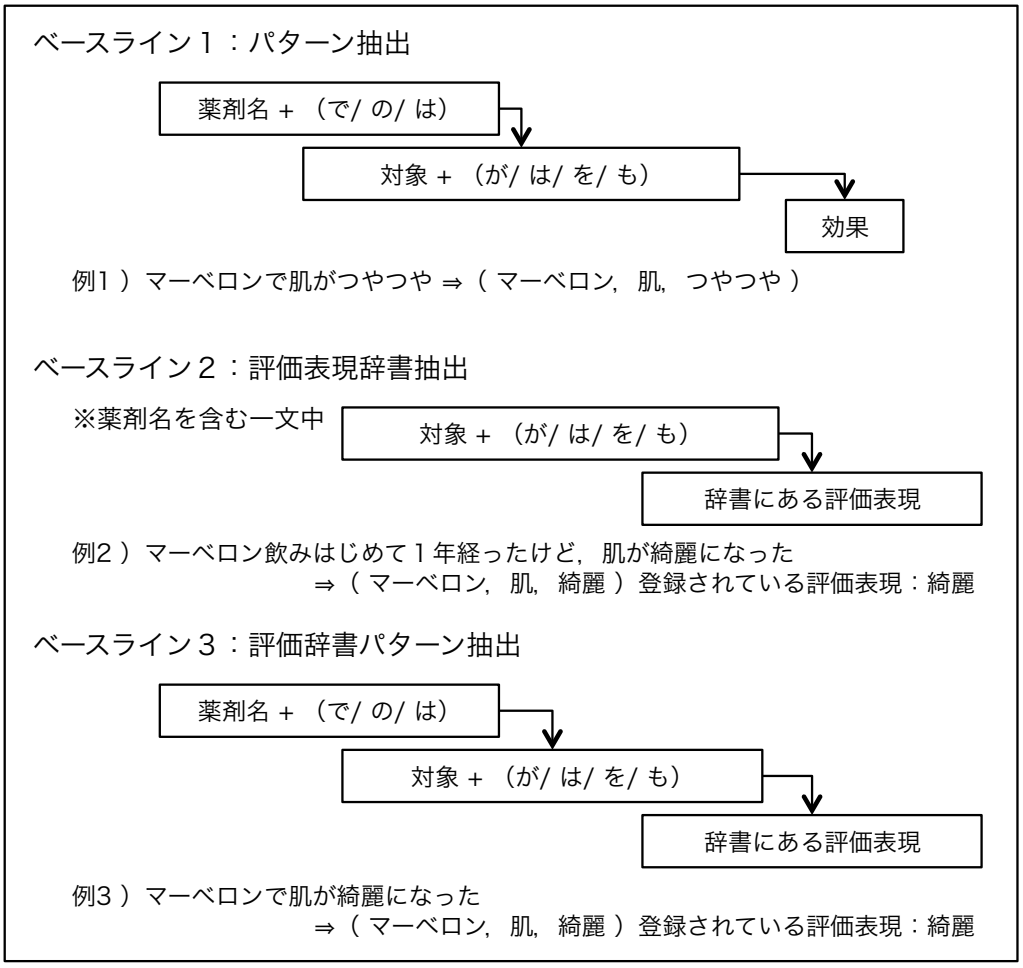

図 4. ベースライン

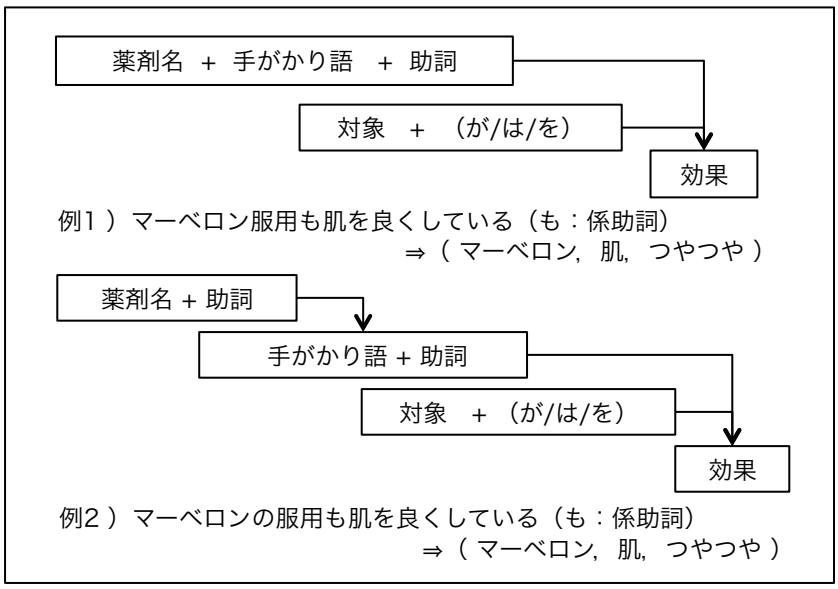

図 5. 抽出規則 A

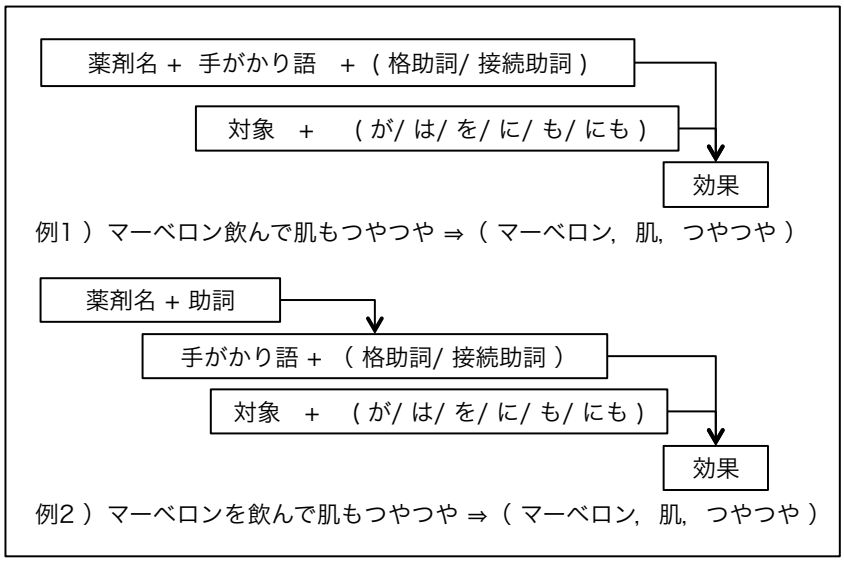

図 6. 抽出規則 B

5.4 結果と考察

実験結果を表 6 に，抽出結果例を表 7 に示す．評価に用いたデータの正解三つ組数は 340 である．提案手法の適合率が最も高かったが，再現率は低い結果となった．他システムの F 値が高く，また，最も助詞や構文パターンの条件が緩いベースライン 2 の再現率が最も高かった．実験結果から，今回提案した手がかり語を用いた抽出が，闘病ブログに対して有効で

あることが確認された．ベースライン２（評価表現辞書抽出），ベースライン３（評価辞書パターン抽出）の結果から，薬剤名と対象・効果との関係性の限定により，抽出三つ組数が大きく変化することがわかる．さらに，抽出規則 A，抽出規則 B の出力結果から，手がかり語に後続する助詞の条件付けが抽出数に大きく起因していることが確認できた．

提案手法の抽出誤りの理由として，大きく２つ考えられる．要素が存在すると判定された場所の誤りによって出力ミスまたは出力が得られないこと（表 7 の④），そして抽出された要素が，三つ組の要素として不適切であること（表 7 の⑤）である．これらについて考察を行うため，２つの分析を行った．

表 6．評価結果

	出力数	正解数	適合率 [%]	再現率 [%]	F-値
手がかり語と構文情報を用いた提案システム	47	24	51.1	7.1	0.124
ベースライン１：パターン抽出	105	11	10.5	3.2	0.049
ベースライン２：評価表現辞書抽出	866	97	11.2	28.5	0.161
ベースライン３：評価辞書パターン抽出	49	7	8.2	1.2	0.021
抽出規則 A	98	39	39.8	11.5	0.178
抽出規則 B	70	29	41.4	8.5	0.141

表 7．提案システムの抽出例

	解析文	出力	正解
①	考えたくなくても、やることはないし大量のステロイドのせいで頭が冴えちゃって…	（ステロイド，頭，冴えちゃう）	（ステロイド，頭，冴えちゃう）
②	幸いにも治療が進むにつれて自力で腎臓が動き出れて、そのまま進級ができました校長先生ありがとうでもステロイドの副作用で顔はパンパン	（ステロイド，顔，パンパン）	（ステロイド，顔，パンパン）
③	皆様(●´∀｀)ノ*【ご】*【き】*【げ】*【ん】*【よ】*【う】*遂に今年もあたしにとって憂鬱な時がやって来ました一応、プログラフはそのままでそしてセルセプトをイムランに変更してでの妊娠・出産は、いいという事でしたが・・・るんですけど・・・で、今回その変更しようとしている『イムラン』というお薬・・・その副作用で骨髄機能に悪さをする事	（イムラン，悪さ，する）	（イムラン，骨髄機能悪さをする）
④	〈喘息でも道は開ける!〉●喘息でも道は必ず開ける!〜漢方で発作がなくなって確信を持った〜（7.6/98-S.Uって発作の度に点滴するのはもうこりごりです	（漢方，×，×）	（漢方，発作，なくなった）
⑤	ヒスロンHの副作用で様々なことがあり、これを機にヒスロンHは中止になりましたです	（ヒスロン，こと，ある）	empty

5.4.1 出力誤り分析

提案手法の出力結果のうち，誤りと判定された 22 文について誤りの種類と原因の調査を行った．

三つ組が抽出されたが誤りであった文のうち３文は，スニペットとして生成される際に不適切に文が結合され，不自然な文であったことが原因であった．表 7 の③で示すように，句点を用いた文分割が不適切であり，誤りの原因となった文は 3 文存在した．この誤りを解決するためには，「☆」「〜」「……」による分割など，句点以外の要素を考慮した分割が必要

である.

また，効果を記述する際，擬態語など多彩な表現が用いられるため，それらが薬剤効果として適切な表現であるかの判定が難しい．しかし，薬剤の変化や効果が現れる対象は，ある程度限定されていると考えられる．提案手法による出力結果を調査したところ，誤りであった 22 の三つ組中，13 組は対象要素が服用情報の要素として不適切であった．表 7 の⑤がその一例である．「時間，土曜日，こと，これ」などがその対象要素例として挙げられるが，これらは薬剤の変化が起こる部位や性質として不適切，もしくは指示語であり意味的に不十分である．これらのことから，対象要素の適切性を判定することで誤った抽出を防ぎ，適合率を向上させることが可能であると考えられる.

5.4.2 正解データ出力状況分析

手がかり語を用いた提案手法により正しく抽出された 16 組の薬剤服用情報は，比較手法では抽出不可能なものであったが，提案手法の再現率は 6.2%と低い値であった．2,000 文の入力データのうち正解データ 258 組の記述形式を調査し，提案手法により出力不可能な 242 組の三つ組について，その原因を以下に述べる.

未登録の手がかり語を用いて薬剤と変化の関係が示されているものが 17 組，手がかり語が用いられていないものが 104 組存在した. 表 7 の④は手がかり語が用いられていないために提案手法で抽出不可能であった例である．スニペットの不完全な一文を対象にしたこと，ブログが口語表現で記述されているため助詞の抜け落ちが原因で, 26 組については単純な構文解析では存在する手がかり語を発見することができなかった．助詞の不一致や手がかり語不在という原因の他に，薬剤名存在文節と手がかり語存在文節との間に意味をもたない「という薬を」などの文節がはさまれているために，手がかり語を発見できない場合もあった．手がかり語を増やし，助詞の抜け落ちや解析の妨げとなる表現を考慮した抽出手法を考える必要がある.

また，提案手法のパターンマッチングに当てはまらないため抽出不可能である正解データは 242 組中 134 組存在した．134 組中 142 組は手がかり語に後続する助詞が格助詞，接続助詞以外であり，対象要素存在文節の末尾助詞が「が，は，を」以外である組は 134 組中 64 組存在した．これらの三つ組には提案したパターンの助詞の条件付けが当てはまらず，抽出が不可能であった．さらに，パターンマッチングが当てはまらない文には「慢性膵炎には漢方薬も有効的」のように「対象→薬剤名→効果」のような出現順で表記された情報や，「高温期続いてくれているのですがデュファストン効果かな〜」のように「対象→効果→薬剤名」の順で記述された情報が 258 組中約 46 組（17.8%）存在した.

指示語が薬剤名や対象部位の代わりに用いられているため，照応解析などの処理が必要な文もいくつか存在し，より高度な抽出を行うためには意味を考慮した抽出が必要である．薬剤の併用に関する記述に対し並列関係を考慮した抽出を行うことも，より多くの服用情報を正確に抽出するためには必要である.

手がかり語が存在しない，抽出パターンにマッチしないなどの問題があったために抽出することができなかった文のうちいくつかに対しては，ベースライン 1（パターン抽出），2（評価表現辞書抽出）により正解三つ組が抽出可能であった．そのため，再現率を上げるためには，提案手法と他の手法とを組み合わせたシステムが有効であると考えられる．

参考文献

[1] 酒井義和，荒木健治, “反対語を利用した文脈依存評価表現の感情極性判定,” 電子情報通信学会論文誌 D，情報・システム J93-D(9)， pp.1778-1789，2010.

[2] 土田正明，水口弘紀，久寿居大, “評判検索のための対象，属性，評価の 3 項関係のランキング法,” 第２２回人工知能学会全国大会，2P2-05，2008.

[3] 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一, “意見抽出のための評価表現の収集,” 自然言語処理，Vol.12，No.3，pp.203-222，2005.

第 6 章

精度改善手法の提案

6.1 複数パターンを用いた再現率向上手法

登録されていない手がかり語が使われていたことや手がかり語が存在しないこと，構文解析ミスなどの原因により，提案手法のパターンでは薬剤服用情報を抽出できなかった文がいくつか存在した．「前回と同じく、レンドルミンも・・便秘が酷く、苦しい。」はその一例である．しかし，**5.3** 比較対象として挙げたベースラインのうち，最も再現率が高スコアであったベースライン 2：評価表現辞書抽出（薬剤名を含む一文から評価表現辞書を用いて評価要素を特定し，その係り元を対象要素候補とする）を用いた場合には，この文から正しい薬剤服用情報が抽出可能であった．

したがって，提案手法により三つ組が得られなかった文についてはベースライン 2 を用いて抽出を行うシステムを作成し，評価を行なった．評価には **5.**で使用したデータと式を用いる．その結果を表 8 に，抽出結果例を表 9 に示す．

正解数が 89 増加し，再現率は提案手法より 26.1 ポイント上昇し F 値は最も高い値となった．また，組み合わせたことにより抽出できなくなった正解三つ組は存在しなかった．このことから，手法の組み合わせが再現率向上に有効であることが確認された．しかしベースライン 2 による出力には誤りも多く含まれているため，適合率は下がる結果となった．出力の誤りをなくすために，抽出された要素の服用情報としての適切性を判断する必要がある．

表 8. 組み合わせシステム評価結果

	出力数	正解数	適合率 [%]	再現率 [%]	F-値
手がかり語と構文情報を用いた提案システム	47	24	51.1	7.1	0.124
ベースライン２：評価表現辞書抽出	866	97	11.2	28.5	0.161
組み合わせシステム（提案システム＋ベースライン２）	890	113	12.7	33.2	0.184

表 9. 組み合わせシステム抽出例

入力	出力	正解
よかった頭痛については、プレドニンを投与しはじめて極度に視力が下がってきていてコンタクト	プレドニン, 視力, 下がる	プレドニン, 視力, 下がる
エンブレルを開始してまだ3回目なのですが、痛みが激減して喜んでいま	エンブレル, 痛み, 激減	エンブレル, 痛み, 激減
いままで甲状腺ホルモンが沢山でていたが、いくら間食・暴飲暴食しても体重が増えないはずだ!!メルカゾールのおかげで病気自体はよくなってきたものの私の体がどんど	メルカゾール, 私の体, どんど	empty
（左側目に傷がついていたせいで点眼薬を出してもらいました（ヒアレイン）次回の予約は3週間後です担当のセンセは9月30日付で異	ヒアレイン, 予約, 異	empty

6.2 辞書を用いた適合率向上手法

薬剤の服用について記述する際，評価表現は多様だが，変化や効果が現れる対象はある程度限定されていると考えられる． そのため，既存の薬剤に関する効果や副作用が書かれた文書から，対象単語として用いられる可能性がある部位や状態を示す単語を収集し，対象単語辞書を作成した． 単語の収集方法は，「 接頭詞 + 名詞 」を基本とし，接頭詞はなくてもかまわない． また，連続する名詞についてはつなげたものを一単語として登録する．また，日付や期間といった対象として不適切な単語を除外するため，数字を含む場合は登録しない（例：1 時間）． 一般向けに優しい表現で書かれたお薬 110 番[1]から作成した辞書と，薬剤添付文書から作成した専門用語を多く含む辞書，二つを用いて各手法により抽出された対象要素が適切かどうか判断する．

例えば，抽出された対象要素が「膝の青あざ」であったとする．この対象要素を MeCab[2] により形態素解析し，「膝（名詞）／の（助詞）／青（接頭詞）／あざ（名詞）」を得る．構成要素のうち接頭詞（なくても可）+ 名詞（連続する場合は結合する）である「膝」と「青あざ」が辞書に含まれているかどうかを検索する．「膝」，「青あざ」どちらかひとつでも辞書に含まれていた場合には「膝の青あざ」が服用情報の対象要素として適当であると判断し，この対象要素を含む三つ組を出力する．

6.2.1 お薬 110 番辞書 (110 dictionary)

お薬 110 番[1]とは薬専門のサーチエンジンである． 患者やその家族向けに，専門的な言葉ではなく優しい表現を用いて書き直されているのが特徴である．ウェブページから取得できた 1,143 ページの薬剤情報のうち，用法についての記載個所を除く概要，効果，副作用

などについての記載箇所から辞書を作成する．その結果，辞書には 12,170 単語が登録された．その一部を表 10 に示す．

表 10. 110 辞書登録単語例

てんかん	けいれん発作	失神発作	不機嫌	躁うつ病	躁状態	偏頭痛発作
目	副作用	肝機能検査	のど	意識障害	吐き気	症状
かゆみ	高アンモニア血症	集中力	発熱	発疹	脳	物忘れ
食欲不振	気分	嘔吐	意識障害	車	青あざ	守り
作業	手足	しびれ	毎日	頭痛	眠気	その他さまざま

6.2.1 薬剤添付文書辞書 (MPi dictionary)

医薬品医療機器総合機構が運営する医薬品医療機器情報提供ホームページ[3]から，医療用医薬品と一般用医薬品の添付文書情報を用いて辞書を作成する． 薬剤添付文書は，各医薬品の製造，販売元が作成しており，医学用語が多く用いられているのが特徴である．TOBYO 辞典を用いて取得したブログから抽出した三つ組のフィルタに用いることが目的であるため，TOBYO 辞典で検索する際クエリとして用いた 296 種の薬剤名について添付文書情報を取得し，薬効と副作用についての記述箇所から辞書を作成する．その結果，辞書には 5380 単語が登録された． その一部を表 11 に示す．

表 11. MPi 辞書登録単語例

てんかん	焦点発作	混合発作	不機嫌	躁うつ病	躁状態	偏頭痛発作
視覚異常	副作用	肝機能	腹痛	意識障害	嘔気	症状
下痢	アンモニア値	血小板減少	発熱	発疹	脳	健忘
食欲不振	悪心	嘔吐	言語障害	姿勢	硬直	厚生省
CPK	尿失禁	便秘	月経不順	頭痛	傾眠	効果毎

6.2.3 辞書の評価実験結果に対する考察

辞書フィルタを用いた抽出結果を表 12 に，フィルタ例を表 13 に示す．評価には 5.で使用したデータと式を用いる．システムが出力した対象要素を辞書フィルタにかけることにより，提案システムに対する MPi 辞書使用時で適合率が最大 21.1 ポイント増加した．また，表 13 に示した 5 種類の手法に対し，フィルタ利用前後の適合率を比較したところ，110 辞書で平均 5.7 ポイント，MPi 辞書で平均 7.5 ポイントの適合率の向上が確認でき，対象用語辞書を用いたフィルタリングが適合率の向上に効果的であることが確認された．フィルタリングなしの出力では，対象要素として日時や期間が抽出されている誤り例が存在したが，フィルタリングによりそれらの誤りを除くことができた．

例えば，フィルタを用いた出力では，対象要素である「朝の眠り」の「眠り」が辞書に登録されているため正しい対象要素として判断する．しかし，「一ヶ月半」は辞書に無い単語のため，対象単語として不適切であると判断できる．このように服用情報として不適切な対象要素を除去することで，適切な三つ組を出力することができる．

２つの辞書を比べると，MPi 辞書の登録単語数は少なく，表現も専門的である．例えば，同意義ではあるが 110 辞書には「吐き気」，MPi 辞書には「嘔気」が登録されているといった違いがある．したがって，MPi 辞書をフィルタとして利用した際には適切な対象要素も多く除かれてしまうが，110 辞書では適度に不適切な対象要素が除去されるため，組み合わせシステムに用いた際に最も高い F 値となっている．つまり，話し言葉で書かれたブログにはお薬 110 番から作成した辞書がより適切であったと考えられる．

また，辞書へ登録する際，単語の意味まで考慮していなかったため，「こと」「の」といったノイズが辞書に含まれていたため，より辞書を有効活用し適合率を向上させるためには，辞書のデータクリーニングが必要である．

表 12. 辞書フィルタ評価結果

	出力数	正解数	適合率 [%]	再現率 [%]	F-値
110 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	35	21	60.0	6.2	0.112
組み合わせシステム（提案システム＋ベースライン２）	524	92	17.6	27.1	0.213
ベースライン１：パターン抽出	67	8	11.9	2.4	0.039
ベースライン２：評価表現辞書抽出	494	72	14.6	21.2	0.173
ベースライン３：評価辞書パターン抽出	30	4	13.3	1.2	0.022
MPi 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	20	9	45.0	2.7	0.050
組み合わせシステム（提案システム＋ベースライン２）	237	44	18.6	12.9	0.153
ベースライン１：パターン抽出	28	5	17.9	1.5	0.027
ベースライン２：評価表現辞書抽出	219	35	16.0	10.3	0.125
ベースライン３：評価辞書パターン抽出	11	2	18.2	0.6	0.011

表 13. 辞書フィルタ抽出例

入力	出力	フィルタの判定		正解
		110 辞書	MPi 辞書	
マイスリーを飲むと強い倦怠感が出るので、昨日は飲まずに就寝しました	マイスリー, 倦怠感, 出る	正	正	正
たしかにタキソテールの副作用で体全体が浮腫んでいたので、腕は両方とも腫れて重くて痛いな〜とはもっと前から思ってたんだけど、これもタキソテールの副作用だと思ってた	タキソテール, 体全体, 浮腫む	正	誤	正
朝、プレドニンを飲んで時間がたつとよくなりますが	プレドニン, 時間, たつ	誤	誤	誤
うちの病院では、薬を使った採卵後と次の周期は卵巣を休ませるためピルを飲来事で何も聞けず・・・orzそういや前回の採卵後も・・・プラノバールの効き目が悪くて14日目で中止されました	プラノバール, 効き目, 悪い	正	誤	正
今、私はヒスロンの治療を終えて約5ヶ月ほど経ってますが、ヒスロン治療のこ	ヒスロン, 治療, 終える	誤	正	誤
Unknown2007-10-2122:48:19急性ではなく慢性膵炎には漢方薬も有効的と最近は良く言われてます	漢方, 慢性膵炎, 有効的	正	正	正
あの、ダルさと吐き気は一体何だったんだろう・・・?リセット5日目からクロミッド飲み始めまして・・・初☆クロミッド!!これにてすくすく良	クロミッド, 吐き気, 良	正	正	誤
もしかしてプレドニンの副作用???最近体重は変わってないというか以前より減って	プレドニン, 副作用???最近体重, 変わってない	正	誤	正
エンブレルの関与は不明ですが、多発性硬化症、視神経の炎症といった	エンブレル, 関与, 不明	正	誤	誤

6.3 手がかり語リストの拡張

手がゞかり語と構文情報を用いたパゞターンマッチングゞにより，(薬剤名，対象，効果)でゞ表される薬剤服用情報の抽出手法を提案してきた．これまで手がかり語として用いていたのは全部で31単語であったが，手がゞかり語を増やすことによりより多くのパターンに合致し，再現率が向上すると考えた．したがって，手がかり語リストを拡張するために行った二つの方法について述べる．性能評価結果は，第7章のデータを用いて行うため，7.2.2で述べる．

6.3.1 人手による追加

5.1 で述べた 2,000 スニペットから収集した 340 組の薬剤服用情報に存在しており，かつ未登録の手がかり語を人手で収集した．その結果，26 単語が追加された．追加された単語を表 14 に示す．

表 14. 人手による手がかり語追加例

お陰	のむ	なる	併用
お蔭	治療	効く	飲用
性	中止	再開	止まる
追加	復活	増やす	やめる
変わる	抜ける	する	mg
残る	食べる	落ちる	なくなる
塗る	最後		

6.3.2 類語辞典を用いた追加

Web 上の類語辞典である Weblio[4]を用いて，**3.3.2** で示した 32 単語の手がかり語の類語を収集した．提案システムは日本語で書かれたブログ記事を対象とするため，英単語の類語は収集しない．また，パターンマッチングを行う際に，手がかり語は 1 文節中から検索される．そのため，複数文節となる可能性のある助詞を含む「区切りをつける（「止める」の類語)」なども収集しない．その結果, 1,472 単語が追加された．「おかげ」について類語検索した結果得られた単語例を表 15 に示す．

表 15. 類語辞書を用いた追加例（「おかげ」から生成）

後援	ご支援	お引き立て
後ろ盾	お陰様	お陰
御陰	御蔭	お蔭

参考文献

[1] ファーマ・フレンド（小林薫), お薬 110 番.
[http://www.jah.ne.jp/~kako/frame_dwm_search.html (cited 2014-December-18)].
[2] MeCab: Yet Another Part-of-Speech and Morphological Analyzer.
[http://mecab.sourceforge.net/ (cited 2014-December-18)].
[3] 独立行政法人 医薬品医療機器総合機構 医薬品医療機器情報提供ホームページ.
[http://www.info.pmda.go.jp/index.htm (cited 2014-December-18)].
[4] 類語辞典・シソーラス・対義語 - Weblio 辞書.
[http://thesaurus.weblio.jp/(cited 2014-December-18)].

第 7 章

アノテーションデータを用いた評価実験

7.1 スニペットから生成したデータを用いた評価実験

7.1.1 正解データ

7.1.1.1 アノテーション方法

1 人の評価者では評価が偏る恐れがあるため，アノテータによる正解データを作成し，そのデータに対しても評価実験を行った．北海道大学の 20 代前半の大学生 3 名（男性 2 名，女性 1 名）・大学院生 2 名（男性 1 名，女性 1 名）がアノテーションしたデータを元に，正解データを作成する．新たに取得したスニペットを句点で区切り，文章中に薬剤名を含む 2,000 文に対し，（薬剤・対象・効果）で表される服用情報が含まれると判断される箇所について，各タグを形態素ごとに付与する．

意味的には等しい抽出範囲の幅の揺れを認めれば 2 人以上一致していた 258 の三つ組を正解データとする．

7.1.1.2 アノテーション結果に関する考察

5 人のアノテーションデータについて κ 値を求めたところ，0.41 と低い値であった．

完全一致が少ない原因を分析するため，システムの目的を最も理解している著者が，2,000 文のうち 200 文に対し服用情報のタグ付けを行い，比較した．その結果 200 文に対し，著者が抽出した三つ組数は，評価者 5 人の平均 17 に対して 32 であった．

著者と評価者とのアノテーション結果に差が出た大きな原因は，説明不足であった．我々が目指すシステムは，現れていない効果も「現れないことが現れた効果」として意味がある

表 16. 200 文アノテーション結果

一致人数	抽出三つ組数	第一著者との比較			
		一致三つ組数	適合率 [%]	再現率 [%]	F値
3人以上	12	12	100.0	37.5	0.545
2人以上	22	20	90.9	62.5	0.741
1人以上	53	30	56.6	93.8	0.706

情報であり抽出すべきだと考えている．しかし例えば，「リフレックスのおかげでかゆみがなくなった」という文に対し，（リフレックス，かゆみ，なくなった）とタグ付けした評価者は一人もいなかった．我々のシステムの目的を考えれば，（リフレックス，かゆみ，なくなった）という情報も正解データとしてタグ付けされるべきである．

さらに，「私に取って、タキソテールの脱毛率は少々高めです」について（タキソテール，脱毛率，高め）と（タキソテール，脱毛率，少々高め）と抽出する違いがあるなど，抽出範囲の揺れも完全一致率を下げる大きな要因であった．

この意味的には等しい抽出範囲の揺れを容認し，一致しているとした場合，評価者 5 人が抽出した三つ組の異なり総数は 200 文中 53 であった．

著者とアノテーションが一致しなかった例について挙げる．ブログの筆者が誤った表現を用いていることや，スニペットの文であるため不完全であること（例：「ある程度はまし」を「程度はまし」と記述）が評価者を混乱させていた．また，「今回の数値はプレドニンを飲み始めて風邪をひいた時を除けば最低数値である」について，「風邪をひいた」ことは「プレドニンを飲み始めた結果」であるのか，筆者の意図する意味かどうかを判定することは，人手でも困難でありシステムに判断させるのは難しい．「熱が上がったら病院なのでボルタレンなど適切な処置をしてくれることでしょう」という文に対して，（ボルタレン，熱，上げる）とタグ付けした評価者もいたが，上がった熱を下げる働きをもつのがボルタレンである． この文章から「ボルタレンは熱を下げる薬である」という情報を抽出するためには，表層的な情報だけではなく一般常識を用いた推論が必要となるため，システムに判定させることは難しい．将来的には推論ができるシステムを作りたいと考えているが，推論が必要な情報は本論文で目指すシステムの正解データとして不適切である．

アノテーションされた 2,000 文のうちランダム抽出した 200 文に対し第一著者がタグ付けし，評価者と第一著者との一致率を計算した結果を表 16 に示す．主旨の等しい抽出結果を一致していると見なしたとき，2 人以上の評価者が一致したタグ付けした結果の F 値が最も高かった．したがって，アノテーションされた 2,000 文のうち，2 人以上の一致が得られた 258 組の三つ組を正解データとした．この正解データは，第一著者の意図するデータとして最も適切だと考えられる．

今後は事前に評価者に対し，システムの目的についてより深く理解してもらい，アノテーションの訓練をするなど，目指すシステムの正解としてふさわしいデータを作成する必要がある．また，アノテーションにより，個人によって重要視する情報に差が見られたため，抽

出する服用情報として主要素である三つ組の代わりに，フレーズでの抽出も医療情報提供には有効であると考えている．

7.1.2 結果

7.1.2.1　提案手法，ベースラインに関する結果

アノテーションデータに対する評価実験結果を表 17 に示す．評価に用いたデータの正解三つ組数は 258 である．評価には **5.2** の式(2),(3)の他，式(4)を用いる．

5 で述べた著者が作成した正解データを用いた実験結果より表 17 の実験結果は数値的には低いが，提案手法の適合率が最も高く，ベースライン２（評価表現辞書抽出）の再現率が最も高いという結果，F 値の大小関係は一致している．したがって，提案手法のブログに対する適切性，組み合わせ手法の再現率改善効果，辞書の適合率改善効果はこの評価実験結果からも確認できる．しかし，アノテーションには目的に沿わないものも含まれていたため，再現率，適合率ともに前の評価実験と比べ全体的に低くなった．

$$\text{再現率} = \frac{\text{正抽出解三つ組数} \times 100}{\text{アノテーションにより得られた正解データ数}} \quad [\,\%\,] \qquad (4)$$

表 17．アノテーションデータに対する評価結果

	出力数	正解数	適合率 [%]	再現率 [%]	F-値
手がかり語と構文情報を用いた提案システム	38	16	42.1	6.2	0.108
組み合わせシステム（提案システム＋ベースライン２）	681	80	11.7	31	0.17
ベースライン１：パターン抽出	143	2	1.4	0.8	0.01
ベースライン２：評価表現辞書抽出	662	64	9.7	24.8	0.139
ベースライン３：評価辞書パターン抽出	35	1	2.9	0.4	0.007
110 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	29	15	51.7	5.8	0.105
組み合わせシステム（提案システム＋ベースライン２）	398	65	16.3	25.2	0.198
ベースライン１：パターン抽出	110	2	1.8	0.8	0.011
ベースライン２：評価表現辞書抽出	379	50	13.2	19.4	0.157
ベースライン３：評価辞書パターン抽出	28	1	3.6	0.4	0.007
MPi 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	19	11	57.9	4.3	0.079
組み合わせシステム（提案システム＋ベースライン２）	181	38	21	14.7	0.173
ベースライン１：パターン抽出	49	0	0	0	
ベースライン２：評価表現辞書抽出	166	27	16.3	10.5	0.127
ベースライン３：評価辞書パターン抽出	11	0	0	0	

7.1.2.2　手がかり語リスト追加による効果

3.3.2 で収集した手がかり語 32 単語を用いた場合，人手により手がかり語を 26 単語増やし，58 単語を用いた場合，類語辞典を用 いて手がかり語を増やし，1,504 単語を用いた場合について抽出性能を評価する．再現率に着目した評価を行うため，評価には 7.2.1 の式 (4)を用いる．結果を表 18 に示す．

手がかり語が増加すると，ごく僅かだが正解数が増加した．追加した数に比べて正解数の増加量が少ないため，追加した手がかり語と薬剤服用情報の各要素との関係が，提案手法のパターンとは異なる可能性があると考えられる．

表 18. 手がかり語追加による評価結果

	手がかり語数	抽出数	正解数	再現率[%]
初期リスト	32	38	16	6.2
人手による追加	58	54	17	6.6
類語辞書による追加	1504	61	18	**7.0**

7.2 ブログ記事全文に対する評価実験

スニペットを抽出対象としたために，文が不適切につなげられ情報として不十分である場合や，構文解析ミスを引き起こす原因となっている場合が約 13.6%（258 文中 35 文）あった．スニペットに起因する問題を回避し，より多くの医療情報を抽出するために，ブログ記事全文を対象とした．

スニペットよりも長く，医療情報以外の情報も多く含むと考えられるため，「´Ｉ｀」「･ω_･」「((+_+))」などの文字列はあらかじめ除去する．この操作により顔文字などの情報が消えてしまうが，それらに含まれる感情も考慮した医療情報の抽出は今後の課題とする．

7.2.1 正解データ作成方法

ブログ記事全文に対する評価実験を行うために，正解データを作成した．薬剤名を含む文のうち，第一著者が薬剤服用情報の可能性がある部分をタグ付けし，その適切性を評価者が判断する．評価者はすべて北海道大学の 20 代前半の大学生 2 名（男性 2 名）・大学院生 3 名（男性 2 名，女性 1 名）である．

その結果，ブログ記事のうち薬剤名を含む 3,000 文中，545 文から抽出された 745 組の三つ組を薬剤服用情報の正解データとする．この正解データには現れていない効果（例えば「痛みがなくなった」など）も含まれており，正と負どちらの薬剤効果についても抽出するという本論文の目的に即したデータとなっている．評価者による評価結果の一致率は κ 値 0.67 であり，実質的に一致していた．また，スニペットに対するアノテーションの際問題で

あった推論が必要な薬剤情報に対応するため，薬剤服用情報の定義を「薬剤の服用に関わる変化」とし，「ウンチが出ていないのでラキソベロンを飲まなければならなかったけど、少し時間をおいて内服する事になった」には（ラキソベロン，ウンチ，出ていない）の三つ組をタグ付けした.

7.2.2 実験結果とその考察

各抽出方法による出力結果を表 19 に示す.

ブログ全文を対象とした抽出結果は，スニペットを対象とした抽出よりもさらに再現率が 2.2 ポイント低かった. 全文ではスニペットとして表示されない範囲まで対象とされるため，より多くの情報が含まれており，その分多彩な記述表現が存在していたためだと考えられる. 「薬の副作用なのか朝が異常にだるいので、薬を追加することでそれが強くなるのが怖く、もう 2 週間デパスだけで様子を見ることにしました」のような文から「薬」が何を指すのかを，「初回のインターフェロン終了後は数ヶ月のあいだ吐き気などが継続したのですが、今回の治療後はそんなことも一切なく」では「一切ない」のは「吐気」であるなど照応解析処理も必要である.

また，適合率を向上させるためには抽出三つ組の適切性を判断する必要がある．しかし，評価表現辞書抽出により「今骨転移でナベルビン、手足の痺れ、感覚マヒを和らげるため「ピドキサール錠」をＤr から処方されてます」から抽出された（ピドキサール，感覚マヒ，和らげる），（ナベルビン，感覚マヒ，和らげる）のように，三つ組要素のみでは判断できない例も多く存在する．そのため，要素としての適切性だけではなく，薬剤名との位置関係や構文情報を用いた判断，さらには意味を考慮した抽出が必要であると考えられる.

表 19. 全文を対象とした評価結果

	出力数	正解数	適合率 [%]	再現率 [%]	F-値
手がかり語と構文情報を用いた提案システム	52	24	46.2	3.2	0.060
組み合わせシステム（提案システム＋ベースライン２）	1412	181	12.8	24.3	0.168
ベースライン１：パターン抽出	193	17	8.8	2.3	0.036
ベースライン２：評価表現辞書抽出	1386	170	12.3	22.8	0.160
ベースライン３：評価辞書パターン抽出	58	8	13.8	1.1	0.020
110 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	36	20	55.6	2.7	0.051
組み合わせシステム（提案システム＋ベースライン２）	815	141	17.3	18.9	0.181
ベースライン１：パターン抽出	130	14	10.8	1.9	0.032
ベースライン２：評価表現辞書抽出	795	133	16.7	17.9	0.173
ベースライン３：評価辞書パターン抽出	40	6	15.0	0.8	0.015
MPi 辞書フィルタ					
手がかり語と構文情報を用いた提案システム	18	10	55.6	1.3	0.026
組み合わせシステム（提案システム＋ベースライン２）	420	75	17.9	10.1	0.129
ベースライン１：パターン抽出	60	13	21.7	1.7	0.032
ベースライン２：評価表現辞書抽出	412	73	17.7	9.8	0.126
ベースライン３：評価辞書パターン抽出	17	5	29.4	0.7	0.013

第 8 章

薬剤服用情報の出現位置, 内容に関する考察

8.1 薬剤服用情報の出現特徴

安全性の観点から，正確な情報を得るアプリケーションを作成する場合には，適合率を重視した評価を行うことが適切だと考えている．一方，とにかく色々な情報がほしい，たくさん集めた情報を元に自分で判断したいと考えている患者向けのアプリケーションを作成する場合には，再現率を重視し，網羅的に多くの情報を収集できるシステムを構築することが望ましい．6.1 で述べた組み合わせ手法により提案システムの再現率は向上したが，それでもまだ31.0%と低い結果となっている. したがって, より多くの薬剤服用情報を抽出するため, 正解データとして作成した薬剤服用情報の各要素（薬剤名，対象，効果）と手がかり語の出現位置について，文法的な特徴を解析した.

その結果，1 文中に出現する薬剤名，対象，効果，そして手がかり語の出現位置について, いくつかのグループに分類することができた.

8.1.1 スニペット分析結果

8.1.1.1 手がかり語の出現状況

まず始めに，2,000 スニペットから得られた 258 組の薬剤服用情報について，調査する. 258 組中，手がかり語を用いて記述されたものは 169 組（65.5%，160 文）存在していた. この結果から, 手がかり語に着目した抽出という方針が適切であったことがわかる. しかし, 手がかり語を用いて書かれた 169 組の薬剤服用情報のうち, 37.9%にあたる 64 組において, 未登録（3.3.2 で定義した 32 単語以外の手がかり語）が用いられていた．これらの未登録の手がかり語の異なり総数は 26 であった．6.3.手がかり語リストの拡張により，これらの未登録の手がかり語のうち，10 種類をリストに追加することができている. この内訳は, 2000

スニペットから抽出した 340 組の薬剤服用情報を用いた追加による 7 種類（16 組の薬剤服用情報を記述する際用いられていた手がかり語），類語辞典を用いた追加による 5 種類（10 組の服用情報を記述する際用いられていた手がかり語）である．さらに，258 組の薬剤服用情報を記述する際存在していた手がかり語のうち 24 組（14%）については，複数の手がかり語（「飲み始める」など）が用いられていた．そのため，複数文節にわたり手がかり語が存在している場合も考え，単に手がかり語を含む文節の係り先文節を効果要素存在文節とするのではなく，手がかり語を含む文節の係り先文節に手がかり語が含まれていた場合，さらにその係り先文節を見て，手がかり語が含まれていなかった場合に効果要素存在文節とする，などの処理が必要であると考えられる．

8.1.1.2 結果と考察

手がかり語を用いて記述された薬剤服用情報の抽出数を増加させるため，手がかり語を文中に含む 169 組の薬剤服用情報について各要素の出現位置を分析した．

各要素の出現情報を簡略化すると，いくつかのパターンに分類可能であった．そのパターンを図 7 に示す．図の長方形は文節を，矢印は係り受け関係を示している．また，2 回以上用いられていた手がかり語と出現パターンとの対応関係を表 20 に示す．パターンの集計数は存在したすべての手がかり語についてカウントしている．

提案手法のパターンである（3），（4）は合計 80 組（47.3%）の薬剤服用情報の記述パターンと一致した．しかし，8.2 の結果より，提案手法では 16 組の正解三つ組しか抽出することができなかった．その原因を調査したところ，パターン（4）に当てはまる記述のときに，薬剤名存在文節と効果要素存在文節との間に 1 文節ではなく複数文節はさんでいる薬剤服用情報が 24 組存在していた．その中には，手がかり語が複数文節にわたる（例えば「飲んでるおかげ」は「飲んでる」「おかげ」二つの文節に分割される）場合が 7 組存在していた．そのため，8.1.1.1 で述べたように薬剤名存在文節から係り先をたどる際に，どこまでたどった先が効果要素存在文節となるか状況に応じて判断する必要がある．

さらに，（3），（4）のパターンが当てはまる 80 組の薬剤服用情報のうち，対象要素存在文節の末尾が提案手法で設定した「が／は／を」以外であったものが 15 組，手がかり語存在文節の末尾が格／接続助詞以外のものが 55 組存在していた．特に，手がかり語存在文節の末尾が読点であるのものが 24 組と多かったため，手がかり語存在文節の末尾の条件を「格／接続助詞，読点」とし，258 組の薬剤服用情報を含む 2000 文から抽出した．その結果，抽出数が 70 増加し 108 となり，正解数は 6 増加し 23 となったため再現率は 2.7 ポイント増加し 8.9%となった．一方，適合率は 20.8 ポイント減少したため，再現率を重視するシステムにおいて用いるパターンとしては，手がかり語存在文節の末尾の条件を「格／接続助詞，読点」とすることが有効であると考えられる．

また，提案手法では，連体化の助詞である「の」により修飾されていた際には対象要素に付加する，という操作を行っている．しかし，提案手法のパターンで記述されている 80 組の薬剤服用情報のうち，対象要素が複数の文節に分割されてしまう場合も 17 組存在した．

例えば，「ドグマチールの副作用は女性ホルモンに影響を及ぼすのでそっちのほうが…」という入力文に対し，（ドグマチール，女性ホルモンに影響，及ぼす）が正解データとしてタグ付

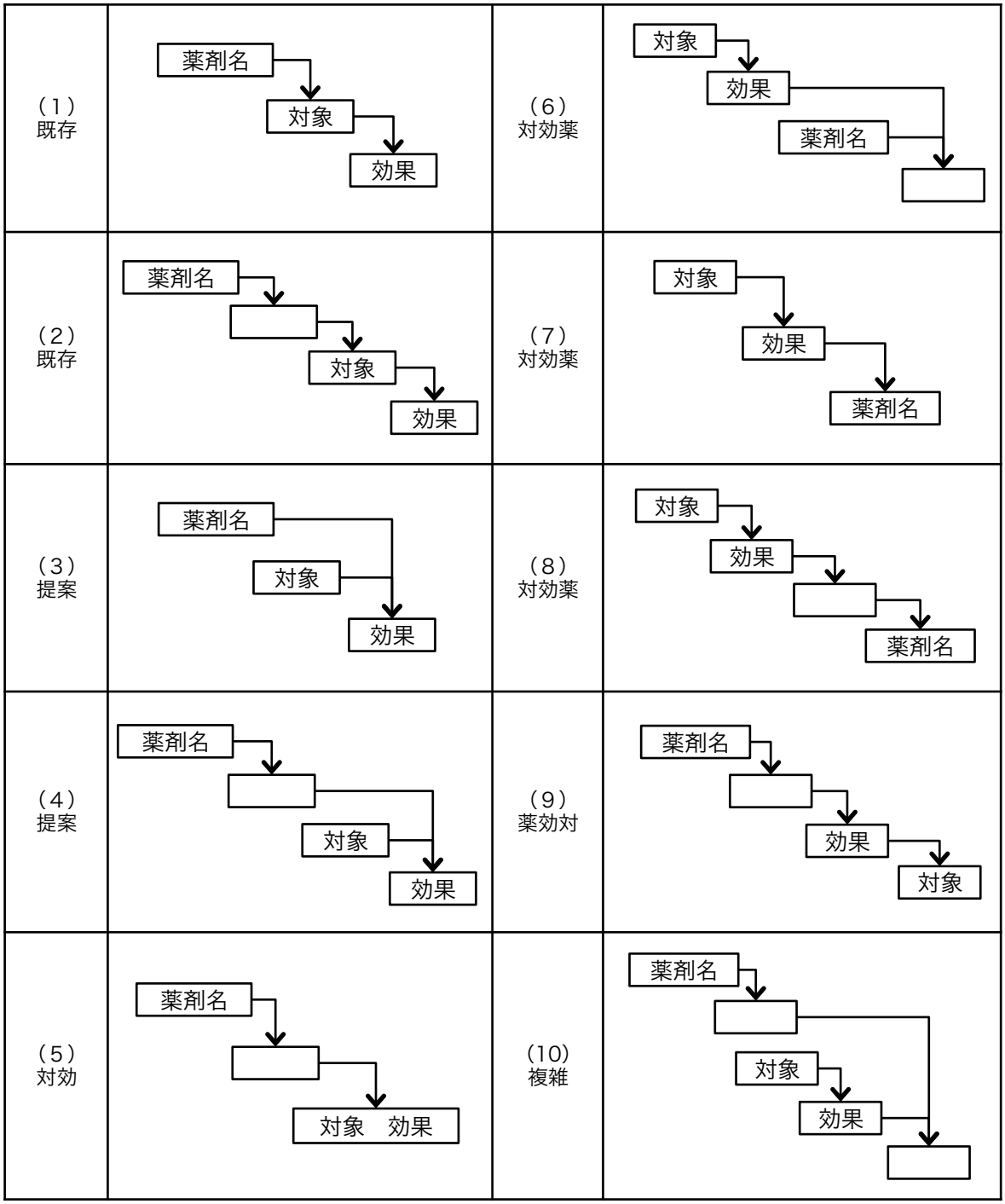

図 7. 薬剤服用情報が出現するパターン

表 20. スニペットに対する手がかり語と出現パターンの対応関係

（赤の濃淡は出現数の多寡を示している）

	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）	（10）	（11）	
	既存		提案		対効	対効薬			薬効対	複雑	その他	計
飲む	0	1	0	16	1	2	0	1	0	2	3	26
副作用	1	2	1	9	1	0	0	1	0	3	2	20
錠	1	0	0	1	2	0	0	0	0	2	2	8
服用	1	0	1	2	0	3	0	0	0	1	0	8
処方	0	0	0	2	0	1	1	0	0	2	0	6
せい	0	0	0	5	0	0	0	0	0	1	0	6
使う	0	0	0	2	0	2	0	0	0	1	1	6
使用	0	0	1	2	0	1	0	0	0	0	1	5
飲み始める	0	0	0	3	1	0	0	0	0	1	0	5
mg	2	1	0	1	0	0	0	0	0	0	1	5
効く	0	0	0	2	1	0	0	0	0	1	0	4
影響	0	0	0	3	1	0	0	0	0	0	0	4
効果	0	0	1	0	0	1	1	0	0	1	0	4
入れる	0	0	0	2	0	0	0	1	0	0	0	3
打つ	0	0	0	0	0	0	0	0	0	2	1	3
投与	0	0	0	3	0	0	0	0	0	0	0	3
続ける	0	0	0	1	0	0	0	0	0	1	0	2
お陰	0	0	0	1	0	0	0	0	0	1	0	2
効き始める	0	0	0	1	0	0	0	0	0	1	0	2
治療	0	0	1	1	0	0	0	0	0	0	0	2
併用	0	0	0	0	0	0	0	0	0	1	1	2
注射	0	0	1	1	0	0	0	0	0	0	0	2
増量	0	0	1	1	0	0	0	0	0	0	0	2
始まる	0	0	0	0	0	0	0	0	0	1	1	2
療法	0	0	0	0	0	1	0	0	0	1	0	2
計	7	4	8	72	9	14	5	3	0	28	19	169

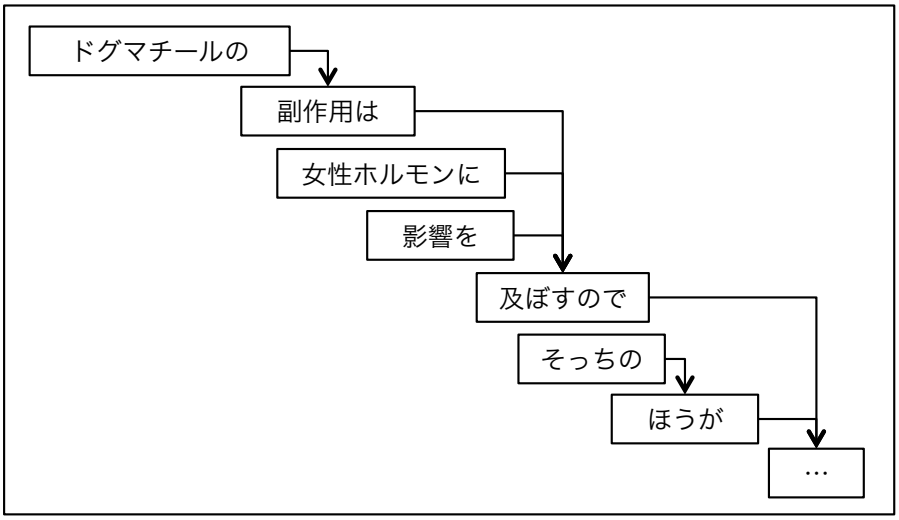

図 8. 構文解析結果

けされており，構文解析の結果，図 8 にあるような係り受け関係を得ることが出来る．図の長方形は文節を，矢印は係り受け関係を示している．しかし，正解データの対象要素である「女性ホルモンに」「影響」はそれぞれ別の文節に存在しており，二つの文節間には係り受け関係もないため，自動的に生成することは困難である．効果要素についても同様に，複数文節にわたり記述されているものが 7 組存在した．そのため，複数文節にわたる要素の抽出，結合方法を考える必要がある．もしくは，用途に合わせて，要素による抽出ではなく，薬剤服用情報を含む事象単位での抽出など適切な抽出範囲を考える必要がある．

8.1.2 ブログ記事全文分析結果

8.1.2.1 手がかり語の出現情報

3000 ブログ記事の全文からタグ付けされた正解三つ組を含む文 545 文のうち，ランダムに選出した 100 文を対象に分析を行った．この 100 文中には 134 組の薬剤服用情報が含まれている．そのうち，83 組（61.1%）で手がかり語を用いた記述がされていた．83 組中，用いられていた手がかり語の種類は 30 種類であり，7 種類（12 組）は複数の手がかり語が連結したものだった．

表 21．パターンと出現数の関係

パターン	出現回数
（4）	44
(10)	16
（6）	13
（3）	11
（1）	9
（7）	8
（2）	5
（5）	4
（9）	2
その他	22

8.1.2.2 結果

手がかり語の存在に関わらず，薬剤服用情報を含む 100 文すべてについて各要素の出現特徴を調査した．文の構造を簡略化したパターンは図 7 のものを用いる．パターンと出現数の関係を表 21 に示し，手がかり語と出現パターンの対応関係を表 22 に示す．表には 2 回以上出現した手がかり語について集計した結果をまとめたが，パターンの集計数は存在したすべての手がかり語についてカウントしている．

表 22. 全文データに対する手がかり語と出現パターンの対応関係

（赤の濃淡は出現数の多寡を示している）

	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）	（10）	（11）	
	既存		提案		対効	対効薬			薬効対	複雑	その他	計
飲む	0	0	0	10	2	0	0	0	0	3	1	16
副作用	0	3	0	2	0	1	0	0	0	1	2	9
服用	0	1	0	0	0	2	0	0	0	2	1	6
治療	0	0	0	4	0	0	0	0	0	0	1	5
併用、入れる	0	0	0	2	0	0	0	0	0	2	0	4
処方	0	0	0	0	0	2	2	0	0	0	0	4
減らす	0	0	0	2	0	1	0	0	0	0	0	3
せい	0	0	0	1	0	2	0	0	0	0	0	3
増やす	0	0	1	1	0	0	0	0	0	0	0	2
投与後	0	0	0	0	0	0	0	0	2	0	0	2
せー	0	0	0	0	0	0	0	0	0	2	0	2
注射	0	0	0	0	0	0	0	0	0	2	0	2
終了	0	0	0	1	0	0	0	0	0	1	0	2
増量	0	0	0	0	0	2	0	0	0	0	0	2
使う	0	0	0	1	0	0	0	1	0	0	0	2
飲み続ける	0	0	0	1	0	0	0	0	0	1	0	2
影響	0	0	0	1	0	0	0	0	0	1	0	2
投与	0	0	0	1	0	1	0	0	0	0	0	2
計	0	4	2	34	3	12	2	1	2	15	8	83

表 23. ランダム 100 文中に存在する三つ組数の違い

	対象データ	
三つ組数	スニペット	全文
1	91	73
2	8	22
3	1	3
4	0	1
5	0	1

8.1.2.3 スニペットとの比較

スニペットと全文，どちらにおいても「飲む」「副作用」「服用」の手がかり語が多く用いられており，要素が出現するパターンについても（4）が最も多く，薬剤名存在文節と効果要素存在文節との間に存在する係り受け関係に着目した本学位論文の方向性が適切であったことが確認された．

一文中に存在する三つ組数を無作為に選んだ100文のスニペットと全文とで比べた結果が表 23 である．TOBYO 辞典を用いて取得したスニペットの平均語長が 119.6 文字程度であ

るため，含まれる情報量も全文に比べると少ないと考えられる．パターンの出現頻度を比較すると，スニペットでは（１）（３）のようなコンパクトなパターンが多く存在しており，全文では（２）のような薬剤名存在文節と効果要素存在文節との文節間距離のあるようなパターンが多く存在していた．

8.2 薬剤服用情報の新規性

本研究で著者が対象とするブログ記事は，専門家ではない患者やその家族によって書かれたものであるため，専門家によって書かれたカルテに出現する用語や言い回しとは異なる表現が使われることが多いと考えられる．実際に本研究で対象とした TOBYO に登録されている闘病ブログに存在する表現と，薬剤添付文書[1]に存在する表現に差異があるかどうかを調査した．

8.2.1 対象データ

第 8 章の評価実験において，ブログ記事全文を対象にタグ付けした薬剤服用情報は 745 組存在していたが，異なり総数は 716 組であった．異なり総数のカウント方法について説明する．例えば「最近は、吸入ステロイドで発作の心配が無くなったので、一人暮らしを楽しみすぎている」という入力文があるとき，正解三つ組としてタグ付けされたのは（１）（ステロイド，発作の心配，無くなった）と（２）（ステロイド，発作，心配が無くなった）であった．しかし，これらの三つ組は，対象，効果の要素として抽出された範囲が異なるだけで（１）も（２）も内容的には同等である．そのため，このような場合には二つの三つ組を区別せず，ひとつの薬剤服用情報とみなしてカウントする．このようにカウントし，再集計した異なり総数 716 組の薬剤服用情報について，出現頻度順に 10 組以上出現していた薬剤名を並べたものが表 24 である．

薬剤名には，一般名（成分名）と商品名がある．参考文献[2]第一部(p.7)によると，“薬は科学的組成、使用目的、性質などいろいろな基準で分類され、一般的な名称（一般薬剤名）がつけられています．"製薬会社がこれらの薬を商品化するときには、それぞれ独自の製剤商品名（銘柄）をつけます。”とあり，区別して記述されるのが一般的である．しかし，TOBYO に登録された薬剤名は一般名と商品名との区別がされていない．TOBYO 辞典でブログを検索する際に用いた「ステロイド」や「漢方」，「免疫抑制剤」は一般名や商品名ではなく，いくつかの薬剤名の総称である．例えば，ステロイド薬は“副腎皮質ステロイド”に分類されており，“現在，日本で発売されている副腎皮質ステロイド薬は，コルチゾン酢酸エステル，ヒドロコルチゾン，デキサメタゾン，トリアムシノロン，ベタメタゾン，プレドニゾロンなどです。[2]第二部(p.36)”とある．つまり，「ステロイド」や「漢方」，「免疫抑制剤」といった語句のみでは薬剤を一意に決定することは不可能である．そのため，TOBYO に登録された薬剤名のうち，出現頻度が多く，なおかつ商品名の一部として含まれている「プレドニン」「タ

表 24. 全文データに対する手がかり語と出現パターンの対応関係

薬剤名	出現回数
ステロイド	114
プレドニン	42
漢方	34
タキソール	30
タキソテール	27
ロキソニン	27
クロミッド	21
インターフェロン	19
免疫抑制剤	19
ハーセプチン	16
タモキシフェン	16
エンドキサン	13
ノルバデックス	13
リウマトレックス	12
ワーファリン	12
ボルタレン	11
リュープリン	10

キソール」「ロキソニン」に関する薬剤服用情報を，薬剤添付文書の内容と比較した．薬剤添付文書のテキスト情報を提供している医薬品医療機器総合機構の検索ツールに「プレドニン」をクエリとして渡すと，三商品（プレドニン眼軟膏，プレドニン錠 5mg，水溶性プレドニン 10mg／水溶性プレドニン 20mg／水溶性プレドニン 50mg）がヒットする．そのうち，更新日が最新である商品（プレドニンの場合は「プレドニン錠 5mg」更新日：2014 年 12 月 10 日）の薬剤添付文書と比較する．調査は著者が行った．

比較の際，具体的すぎる，個人的すぎる内容（例えば，「漢方のお陰なのか、今日の体温は３６」という入力文があったとき，36 度が筆者にとって高いのか低いのかは，読み手からは判断がつかない）や，抽象的すぎる内容（例えば，「そうそう、ハーセプチンは私はほとんど副作用ないです」という入力文があったとき，副作用が何を指しているのか特定できない）については薬剤添付文書と比較することが難しいため，除外する．「プレドニン」「タキソール」「ロキソニン」に関する薬剤服用情報は 96 組存在していたが，このように具体的すぎる，抽象的すぎる内容の情報は 17 組であった．

8.2.2 結果と考察

比較結果の一部を表 25 に示す．添付文書中に同様の記載が含まれていた薬剤服用情報は，96 組中 61 組であった．薬剤添付文書では症状が用語（名詞）で記載されているため，①のように対象要素に薬剤添付文書に含まれているような症状を表す専門用語が含まれる場合がある．しかし，②のように「浮腫」が「むくみ」のような簡単でわかりやすい表現に変換さ

れ，記載されている場合も 38 組存在しており，ブログから収集した表現に新規性があると言える．特に③のような末梢神経障害を表す「ピリピリ」といった表現は擬態語であり，このような患者の言葉で書かれた理解しやすい表現を収集し，提示することは臨床の際に価値のあることだと考えられる．

薬剤添付文書に含まれていない情報は 12 組（例えば表 25 の④），判断不可能なものが 6 組（⑤）存在していた．このような記述には未知の効果や副作用を含む可能性があるが，筆者の持病の症状についての情報である場合あり，筆者の持病を把握し，その症状と照らし合わせ，薬剤に関する情報なのか，持病に関する記述なのかどうかを判断する必要がある．筆者の認識誤り（素人判断による誤解）も考えられるため，大規模な情報を元に，その情報が統計的に確からしいかどうか判断する必要がある．

また，「プレドニン」「タキソール」「ロキソニン」に関してタグ付けされた薬剤服用情報 96 組のうち，記述された内容を調査したところ，効果に関する記述が 44 組，副作用に関する記述が 49 組，その他判断不可能な記述が 3 組存在していた．したがって，患者が記述したブログには副作用と効果に関する三つ組で表すことのできる情報が同程度存在しているため，副作用だけではなく効果も抽出できる提案手法が適切であることが確認できた．体験情報は 67 組（69.8%）存在しており，30%程度は医師から説明された事実や薬の単なる説明（例えば「ボルタレンゲルは、眼とか粘膜に付かないようにしないとだから、結構大変です」）であるため，体験情報なのか伝聞情報なのかモダリティを考慮した抽出が今後必要である．

専門家と患者間で使われる言葉の違いを調査した研究[3]では，患者によって書かれた闘病記に含まれる症状用語を医療従事者がタグ付けし，副作用報告レポートで用いられる用語に MedDRA/J(Medical Dictionary for Regulatory Activities Terminology) [4]を用いて変換し，集計している．患者の表現を専門的な表現に変換することができれば，副作用の調査に役立つ可能性がある．しかし，本研究が目指すシステムは，患者の表現を変換せず，そのまま提示することを目指している．患者の言葉を用いることにより，優しい表現で理解しやすい情報が含まれている可能性があり，システムの利用者として考えている患者の判断，理解を助ける効果があると考えている．また，医療従事者にとっても，臨床の現場で役立つ情報を提供できると考えている．例えば，ひとくくりに「痛み」と言っても，原因や程度によって様々な表現で伝えられる．頭に対する「ズキズキ」「ガンガン」「ズキンズキン」という痛みであれば，「郡発頭痛」のときによく用いられる[5]など，一つの症状を表現するために用いられた複数の表現を関連づけておくことにより，患者の感覚的な訴えを診断に役立てることができると考えている．

表 25. 全文を対象とした評価結果

		正解データ			添付文書にあるか	添付文書にある表現
		薬剤名	対象	効果	1（あり） -1(判断不可) 0（なし）	
①	と思いつつタキソールの4クール目が始まる頃また脱毛がやってきました	タキソール	脱毛	やってきた	1	副作用　<国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「脱毛(30.5%)」、<海外臨床試験成績>「脱毛」 その他の副作用「12.皮膚　脱毛（45.3%）」
②	プレドニンのむくみは相変わらずだけど腹囲は変わりないから腹水も溜まってないしOKOK	プレドニン	むくみ	相変わらず	1	その他の副作用「8.体液・電解質　浮腫」
③	タキサン系（タキソール、タキソテールetc）、Fu−系（ゼローダetc）の薬を使用すると手先のピリピリ感がある　ビタミンB6（ピドキサール）を一緒に服用すると、症状に改善効果がみられる　保湿剤（ウレパール　牛車腎気丸　芍薬甘草湯◎乳がんの予防は	タキソール	手先のピリピリ感	ある	1	重要な基本的注意　「6.末梢神経障害が高頻度に起こるので，観察を十分に行い，症状（しびれ等）があらわれた場合には減量，休薬等の適切な処置を行うこと。」 副作用　<国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「主な副作用は末梢神経障害(37.6%)」 <海外臨床試験成績>「末梢神経障害」 重大な副作用　3. 末梢神経障害，麻痺:「しびれ等の末梢神経障害（43.8%），麻痺（0.1%），片麻痺（0.1%未満），不全麻痺（頻度不明）があらわれることがある」
④	FECに比べるとタキソールの方が吐き気は少ないとは聞いていましたが、毎週投薬になるので1回の薬の量にもよるかと思いますが、まったく出ていないことにはビックリです	タキソール	吐き気	少ない	0	
⑤	2006年01月27日平成18年1月27日金本日の体調体温　36.5°体重　46.0kg腹囲　78cm胴囲　70cm昨夜は夕食後にロキソニンを飲んで、寝る時ちょっと苦しかったがすぐ寝れて、朝もスッキリ目が覚めた	ロキソニン	目	覚めた	-1	

参考文献

[1] 独立行政法人 医薬品医療機器総合機構 医薬品医療機器情報提供ホームページ. [http://www.info.pmda.go.jp/index.htm (cited 2014-December-18)].
[2] 木村繁，医薬制度研究会(2008),『医者からもらった薬がわかる本〈2009 年版〉』，法研.
[3] 荒牧英治，増川佐知子，宮部真衣，森田瑞樹, “患者と医師が使う言葉の違い ～闘病記の医学的な応用に向けて～,” 人工知能学会特集号 28(6), 857-861, 2013.
[4] MedDRA/J. [https://www.pmrj.jp/jmo/ (cited 2014-December-18)].
[5] オノマトペラボ. [http://onomatopelabo.jp/ (cited 2014-December-18)].

第 9 章

おわりに

闘病者によって書かれたブログ記事から，医療情報を検索できるシステムを構築することを最終的な目的としている．ユーザに合わせて実体験情報を提供することにより，患者にとって治療法などを決定する際の判断材料となるだけでなく，精神的な支えとなると考えている．

そのようなシステムを構築する第一段階として，闘病ブログのスニペットに対し，（薬剤名・対象・効果）で表される薬剤服用情報を，手がかり語と構文情報を用いて抽出する手法の提案を行った．提案したシステムは，評価要素に着目した抽出を行わないため，正と負どちらの薬剤効果についても抽出することができる．また，文中に出現する薬剤名と対象，評価要素との構文的関係性を考慮した抽出を行うため，複数の話題を含むブログ記事からも薬剤効果の抽出が可能である．さらに，効果や様態を記述する際に，日本語では多く用いられる擬態語，擬音語の抽出も可能であるという点で新規性がある．

システムを改善するために，複数のパターンを組み合わせた抽出を行う手法を提案し，再現率を向上させた．また，薬剤に関する既存の文書から対象単語辞書を作成し，フィルタリングに用いる事で適合率を向上させた．さらに，ブログ全文に対しても抽出手法を適用することで，提案手法の有効性と全文に対する抽出課題を確認した．

今後の課題として，以下の 5 つが挙げられる．

- 解析文生成時に，削除して差し支えない括弧かどうかの判定，文へ分割する際の区切り箇所の判定
- 文末表現である「〜ません」や，「減量」といった手がかり語の表現による意味の反転を考慮
- 薬の併用に対応する
- 辞書をより有効利用するために，データのクリーニングを行う
- 文の構文パターンを判別し，適切なパターンを用いた抽出を行う

また，アノテーションから，服用情報を要素として抽出するには個人差が大きく，要素の抽出では情報量が不十分な場合もあることが確認された．したがって，服用情報を単語単位ではなく，フレーズ単位で抽出することも効果的であると考えられる．

さらに，抽出した情報に対しそれが伝聞情報なのか体験情報なのか，事実なのか仮定なのかを判定したいと考えている．

TOBYO に登録されていないブログサイトからも今後さらに多くの医療情報を入手するためには，筆者の性別や年齢など個人情報の抽出も課題である．

今後さらに多くの情報を抽出するためには，助詞抜け落ちの補完や，意味解析技術や照応解析技術を用いた指示語の解析など，より高度な言語処理技術が必要であると考えられる．

最終的な目的として，ナラティブ情報を簡単に患者同士でシェアできるようなシステムを作成し，医療情報検索の助けをしたいと考えている．この医療情報とは，すでに把握されている情報だけでなく，新しい情報も含む．患者にとって判断材料や支えとなる情報を提供でき，また，未知の医療情報発見にも役立つシステムを目指している．

オンラインでの健康情報交換が発展しているため，もっと多くの人たちが情報を共有することができれば，医師が発見できないような病気の傾向などを突き止めるチャンスも多いにあると考えている．

謝辞

本研究の遂行のみならず，あらゆる面において，終始ご懇篤なご指導とご高配を賜わりました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，荒木健治教授に深く感謝の意を表します．

ご多忙のなか，快く副査をお引き受けくださいました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座情報メディア環境学研究室，山本強教授に深く感謝の意を表します．

ご多忙のなか，快く副査をお引き受けくださいました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座メディアダイナミクス研究室，長谷山美紀教授に深く感謝の意を表します．

本研究を進めるにあたり，有益なご指導，ご討論をいただいた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，伊藤敏彦准教授に深く感謝の意を表します．

本研究を進めるにあたり，数多くのご指導，ご討論をいただき，公私にわたりご協力いただいた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，ジェプカ・ラファウ助教に深く感謝の意を表します．

著者の研究期間中，公私にわたりご協力いただいた本研究室の学生諸氏に厚くお礼申し上げます．

研究業績

査読付論文

(101) Shiho Kitajima, Rafal Rzepka and Kenji Araki, " Blog Snippets Based Drug Effects Extraction System Using Lexical and Grammatical Restrictions", International Journal of Multimedia Data Engineering and Management (IJMDEM), 2014.

(102) 北嶋志保, ジェプカ・ラファウ, 荒木健治, "闘病ブログに出現する (薬剤名, 対象, 効果) で表される薬剤服用情報の抽出", 知能と情報(日本知能情報ファジィ学会誌), Vol.27,No.1, 2015 (採択決定済).

国際会議

(103) Shiho Kitajima, Rafal Rzepka and Kenji Araki, "Extraction of Drug Information Using Clue Words from Japanese Blogs", Proceedings of the 14th Conference of the Pacific Association For Computational Linguistics 2013, Tokyo, Japan (Paper Sep4-8).

(104) Shiho Kitajima, Rafal Rzepka and Kenji Araki, “Performance Improvement of Drug Effects Extraction System from Japanese Blogs”, in the Proceedings of 2013 IEEE Seventh International Conference on Semantic Computing, pp. 383-386, Irvine, USA, 2013.

研究会・国内大会など

(105) 北嶋志保, 荒木健治, ジェプカ ラファウ, “闘病ブログを対象とした手がかり語を用いた薬剤服用情報の抽出手法”, 第28回ファジィシステムシンポジウム講演論文集 pp.251-256, 2012.

(106) 北嶋 志保, ジェプカ ラファウ, 荒木 健治, “闘病ブログを対象とした薬剤服用情報抽出手法の医療対象用語辞書を用いた改善”, 人工知能学会第2種研究会 ことば工学研究会資料, SIG-LSE-B301-4, pp.19-23, 2013.

(107) 北嶋志保,ジェプカ ラファウ,荒木 健治, “手がかり語の追加による薬剤情報抽出システムにおける有効性”（"Effectiveness of Increasing Clue Words on System for Drug Effects Extraction"）, 電気・情報関係学会北海道支部連合大会, 講演番号111, 2014.

研究業績（共著）

査読付論文

(108) 荒牧英治, 荒木健治, 石井保志, 上松幸一, 北嶋志保, 熊谷晋一郎, 櫻井公恵, 瀬戸山陽子, 高橋由光, 冨井美子, 堀里子, 西舘澄人, 森田瑞樹, 矢島弘士, 和田恵美子, 則のぞみ, 島本裕子, “患者/クライ エント/当事者による記録を活かす意見交換会の活動報告”, 医療情報学, 第34巻4号, pp.197-200, 2014.

国際会議

(109) Magnus Ahltorp, Maria Skeppstedt, Shiho Kitajima, Rafal Rzepka and Kenji Araki, “Medical vocabulary mining using distributional semantics on Japanese patient blogs”, Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine (SMBM 2014), pp. 69-73, Aveiro, Portugal, October, 2014.

研究会・国内大会など

(110) Magnus Ahltorp, Hideyuki Tanushi, Shiho Kitajima, Maria Skeppstedt, Rafal Rzepka and Kenji Araki, “HokuMed in NTCIR-11 MedNLP-2: Automatic Extraction of Medical Complaints from Japanese Health Records Using Machine Learning and Rule-based Methods”, Proceedings of the 11th NTCIR Conference, pp. 158-162, Tokyo, Japan, December 2014.

付録 A

薬剤服用情報の出現パターン

各要素の出現情報を簡略化すると，いくつかのパターンに分類可能であった．それぞれのパターンについての説明を以下にまとめる．

（1）薬剤名存在文節と対象要素存在文節が係り受け関係にあり，対象要素存在文節と効果要素存在文節が係り受け関係にある．そのため，従来のオピニオンマイニングのパターンが適用可能であると考えられる．

（例）「昨日、ブログを書いてから眠剤を飲む前にセロクエル（安定剤）を 2 錠服用し、それから 15 分ぐらい経ったあたりで中を歩くのが少々苦痛になってきた」

正解（セロクエル，歩く，苦痛になる）

（2）出現順は薬剤，対象，効果であり，「い」のパターンと類似している．相違点は，薬剤名存在文節と対象要素存在文節との間にいくつかの文節をはさんでおり，直接係り受け関係がある分けではない．薬剤名存在文節と対象要素存在文節との間に存在する文節には，手がかり語が含まれることがある．

（例）「いつも、プラノバールの主な副作用である吐き気を感じない」

正解（プラノバール，吐き気，感じない）

（3）提案手法のパターン．薬剤名存在文節と効果要素存在文節，対象要素存在文節と効果要素存在文節との間にそれぞれ直接の係り受け関係がある．

（例）「さっき、ボルタレン使用しましたが、すでにお腹がぐるぐるしています」

正解（ボルタレン，お腹，ぐるぐる）

（4）提案手法のパターン．薬剤名存在文節と効果要素存在文節との間に，いくつかの文節をはさんで係り受け関係がある．間に存在する文節のなかには，手がかり語が含まれることがある．また，対象要素存在文節と効果要素存在文節との間にも直接の係り受け関係がある．

（例）「最近は漢方のおかげか、長かった生理周期もだいぶ短くなって、よくな」（スニペットのため文末が不自然）

正解（漢方，生理周期，短くなる）

（5）薬剤名存在文節と対象，効果要素存在文節との間に，いくつかの文節をはさんで係り受け関係がある．間に存在する文節のなかには，手がかり語が含まれることがある．助詞の抜

け落ちなどにより，対象要素と効果要素が同文節に存在しており，品詞名だけでは分割が困難な場合が多い．

（例）「ステロイド治療のあとに気道感染起こしました」

正解（ステロイド，気道感染，起こした）

（６）要素の出現順が対象，効果，薬剤である．対象要素存在文節と効果要素存在文節との間に係り受け関係がある．対象要素存在文節と効果要素存在文節との間に「ために」などの文節が存在することもある．薬剤名存在文節の係り先文節が，効果要素存在文節の係り先文節と同じである．

（例）「今回も、吐き気を抑える為にステロイドを使用しています」

正解（ステロイド，吐き気，抑える）

（７）要素の出現順が対象，効果，薬剤である．対象要素存在文節と効果要素存在文節との間に係り受け関係があり，効果要素存在文節と薬剤名存在文節との間に係り受け関係がある．対象要素存在文節と効果要素存在文節との間に「ために」などの文節が存在することもある．

（例）「8 時に、子供の肺の発育をよくするためのステロイド注射をうつ」

正解（ステロイド，肺の発育，よくする）

（８）要素の出現順が対象，効果，薬剤である．対象要素存在文節と効果要素存在文節との間に係り受け関係があり，効果要素存在文節と薬剤名存在文節との間にいくつかの文節をはさむ．対象要素存在文節と効果要素存在文節との間に「ために」などの文節が存在することもある．

（例）「そのため下痢になる副作用があるセルセプトを一旦中止に」

正解（セルセプト，下痢，なる）

（９）要素の出現順が薬剤，効果，対象である．薬剤名存在文節と効果要素存在文節との間にいくつか文節をはさんで係り受け関係があり，効果要素存在文節と対象要素存在文節との間に係り受け関係がある．

（例）「オキサリプラチン投与後しばらくして起きる末梢神経障害」

正解：（オキサリプラチン，末梢神経障害，起きる）

（１０）各要素が離れて存在している複雑なパターン．対象要素存在文節と効果要素存在文節の間に係り受け関係がある．効果要素存在文節の係り先をたどっていくと，薬剤名存在文節の係り先をたどった先と一致する．

（例）「ちょっとプレドニンが減るので、ほてりや動悸が少なくなるといいなあって思っ」

正解：（プレドニン，ほてりや動悸，少なくなる）

（１１）そのパターンにも当てはまらない，その他のパターン

（例）「それなら熱上がったらロキソニン飲む〜って帰ったとたん 37・6→38.6 と上がった（驚）急いでロキソニン飲んだらあっという間に下がった〜よかった〜♪」

正解：（ロキソニン，熱，下がった）

付録 B

ブログに記載された薬剤情報と薬剤添付文書との比較結果

8.2 で薬剤服用情報の新規性を調査するために行った結果を以下にまとめる.

	薬剤名:出現回数	正解データ			効果:1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用:-1	yes: 1/ no: 0/ 判断不可: -1				
	プレドニン:41									
1	プレドニンを増やした状態で呼吸器夜間症状が改善している気配もあり、このまま様子をみることに	プレドニン	呼吸器夜間症状	改善している	1			1	1	効能／効果「(7) アレルギー性疾患:気管支喘息, 喘息性気管支炎(小児喘息性気管支炎を含む)」
2	プレドニンの数ある副作用の中に、発汗過多がある	プレドニン	発汗過多	ある	-1			?	1	その他の副作用「11.皮膚　発汗異常」
3	プレドニンが10錠になり、経過も良好	プレドニン	経過	良好	1	1		1		
4	心臓にも良くないからってすぐプレドニン止める様に先生はしてくれたんですけど、今も半年に1回は内分泌に行ってます	プレドニン	心臓	良くない	-1				1	重大な副作用「心筋梗塞」
5	」とこぐまくん同様プレドニン(ステロイド)を内服してみましたが、咳は止まらず、、	プレドニン	咳	止まらず	-1			1	1	効能・効果「(13) 肺疾患　(14) 結核性疾患(抗結核剤と併用する)」
6	プレドニンのむくみは相変わらずだけど腹囲は変わりないから腹水も溜まってないしOKOK	プレドニン	むくみ	相変わらず	-1			1	1	その他の副作用「8.体液・電解質　浮腫」
7		プレドニン	腹囲	変わりない	1			1	0	
8		プレドニン	腹水	溜まってない	1			1	0	
9	水曜日からプレドニンを飲み始めて4日、もう血糖値に変化がでた	プレドニン	血糖値	変化が出た	1	1		1	0	
10	プレドニンを飲み続ける限り血漿を溶かす薬も切れませんが、奇跡的に肺の機能をとりもどし、走り回るチビたちと散歩ができるくらいにはなりました	プレドニン	血漿を溶かす薬	切れない	-1				0	
11		プレドニン	肺の機能	取り戻した	1			1	1	効能・効果「(13) 肺疾患」
12	と気を引き締めました(プレドニン減らして、顔も引き締めたいと思いつつ	プレドニン	顔	引き締める	-1			1	1	その他の副作用「8.体液・電解質　浮腫」
13	私としてはプレドニンは副作用があって好きではないんで拒否したんですが	プレドニン	副作用	ある	-1	1		1		
14	|記事一覧|プレドニン減薬中»コメント[コメント記入欄を表示]1■変な汗もしやお昼過ぎの変な汗は低血糖になっているかもプレドニン減量したから、、というのは	プレドニン	低血糖	なっている	-1			?	1	効能・効果「(17)特発性低血糖症」
15	どっちにしても脱毛にはプレドニンも効くらしい	プレドニン	脱毛	効くらしい	1				1	効能・効果「(24)円形脱毛症(悪性型に限る)」その他の副作用「11. 皮膚　多毛、脱毛」

	薬剤名：出現回数	正解データ			効果：1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用：-1	yes: 1/ no: 0/ 判断不可: -1				
	プレドニン:41									
16	プレドニンを服用した当初は、食欲が旺盛で、妻からも「よく食べるなー」と言われて、「病院食の反動」と言い訳しながら、食欲に任せて食べた	プレドニン	食欲	旺盛	-1			1	1	その他の副作用「3. 消化器　食欲不振、食欲亢進」
17	誘発剤のフェマーラの服用も昨日で終わり、フェマーラ＆プレドニンパワーで元気なタマゴちゃんに育って欲しいです(*	プレドニン	元気なタマゴちゃん	育つ	1				1	効能・効果「(22)副腎皮質機能障害による排卵障害」
18	なおプレドニンの1日量8mg以下なら骨粗鬆症の原因にはならないとの事です	プレドニン	骨粗鬆症の原因	ならない	1				1	重大な副作用「5.骨粗鬆症」 高齢者への投与「骨粗鬆症」
19	プレドニン飲み始めて風邪引いたこともないし、寝込んだこともない、	プレドニン	風邪	引いたこともない	1			1	0	
20		プレドニン	寝込んだこと	ない	1			1	0	
21	二週間前からプレドニンもゼロになり、今の所体調に変化もなく生活できています	プレドニン	体調	変化もなく	1	1		1		
22	今回、プレドニンの効きが良すぎてちょっと心配していましたがやはり切れた時の脱力も相当のようです	プレドニン	効き	良すぎ	1	1		1		
23		プレドニン	脱力	相当のよう	-1				1	薬効薬理「(v)中枢神経に対する興奮作用等」
24	プレドニンにより直接食欲増加やむくみなどがでた気配はなし	プレドニン	食欲増加やむくみなど	でた気配はなし	1			1	1	その他の副作用「3. 消化器　食欲不振、食欲亢進」 その他の副作用「8.体液・電解質　浮腫」
25	プレドニンは満腹感がなくおなかが空いてつらいうえに、顔はどんどん丸くなるし体も太る	プレドニン	満腹感	なく	-1			1	0	
26		プレドニン	おなか	空いてつらい	-1			1	0	
27		プレドニン	顔	どんどん丸くなる	-1			1	1	その他の副作用「8.体液・電解質　浮腫」
28		プレドニン	体	太る	-1			1	1	その他の副作用「12. その他　体重増加」
29	-】腫れが少しづつ大きくなるので、治療に入るまでプレドニンでおさえる	プレドニン	腫れ	おさえる	1			1	0	
30	K医師の興味は、プレドニンによる倦怠感改善、および、プレドニン減量による倦怠感増悪の有無、のようだった	プレドニン	倦怠感	改善	1				1	薬効薬理「(v)中枢神経に対する興奮作用等」
31		プレドニン	倦怠感増悪	有無	0				1	薬効薬理「(v)中枢神経に対する興奮作用等」 重大な副作用（頻度不明）「精神変調、うつ状態」
32	それにしてもプレドニンの副作用で気分が高揚してるのかやけに働き者になってる私です	プレドニン	気分	高揚している	1			1	1	薬効薬理「(v)中枢神経に対する興奮作用等」 その他の副作用「5. 精神神経系　多幸症」
33	>でも、杉村さんがプレドニンを勧めて下さったのですが、プレドニンでアレルギーがでたかた見えますか?わたしはプレドニゾロンを1年くらい使っていますが、何事も起こっていません	プレドニン	アレルギー	でた	-1				1	その他の副作用「1. 過敏症注 発疹」
34	副作用の心配は勿論消えないでしょうが、プレドニンの効果を感じてしまうかもしれません	プレドニン	効果	感じてしまう	1	1				
35	ゆうべは離脱症状の関節痛もまったくなくよく眠れたのでトイレには何度か起きたけどプレドニンを今日から20mgにする事にしました	プレドニン	離脱症状の関節痛	まったくなく	1			1	-1	
36	ベッド待ちをしている患者さんが多くプレドニンを服用してから最低2週間は重大な副作用が出ないか様子を見ないといけないため服用が遅れると退院も遅れること	プレドニン	重大な副作用	出ないか	0	1				
37	そして、プレドニン減らして、おニクへらすんだもんねっ	プレドニン	おニク	へらす	-1				1	その他の副作用「12. その他　頻度不明　体重増加」
38	腫れを引くプレドニンの点滴は徹底的にする」とのムンテラでした	プレドニン	腫れ	引く	1				-1	
39	一時増やしたプレドニンは元通り10mg日に、それでも、夜眠れるなど、呼吸器状態は改善中	プレドニン	呼吸器状態	改善中	1			1	1	効能／効果「(7) アレルギー性疾患：気管支喘息，喘息性気管支炎（小児喘息性気管支炎を含む）」
40	プレドニン15mg服用中のだったこともあり、副作用が少なめで済んだ可能性あり	プレドニン	副作用	少なめで済んだ	1	1		1		
41	それを病院で言ったら、「プレドニン」を処方してくれて、痛みの辛い時は飲んでます	プレドニン	痛み	辛い	-1				1	効能・効果「(3)リウマチ性多発筋痛」 「(12)重症消耗性疾患の全身状態の改善」

	薬剤名:出現回数	正解データ			効果:1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用:-1	yes: 1/ no: 0/ 判断不可: -1				
	タキソール:29									
1	と思いつつタキソールの4クール目が始まる頃また脱毛がやってきました	タキソール	脱毛	やってきた	-1			1	1	副作用 <国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「脱毛(30.5%)」、<海外臨床試験成績>「脱毛」 その他の副作用「12.皮膚 脱毛(45.3%)」
2	以前、タキソールで再発治療をしていて、効果があったのですが、耐久性のある癌を抑えきれず、タキソールは中止してしまったからです	タキソール	効果	あった	1	1		1	1	効能又は効果 「再発又は遠隔転移を有する頭頸部癌，食道癌，血管肉腫，進行又は再発の子宮頸癌，再発又は難治性の胚細胞腫瘍(精巣腫瘍，卵巣腫瘍，性腺外腫瘍)」
3		タキソール	癌	抑えきれず	-1			1	0	
4	FECに比べるとタキソールの方が吐き気は少ないとは聞いていましたが、毎週投薬になるので1回の薬の量にもよるかと思いますがまったく出ていないことにはビックリです	タキソール	吐き気	少ない	1				0	
5	結果は肝臓転移、21日にタキソール1回目点滴投与、ぼちぼち副作用が出てきてるようです	タキソール	副作用	出てきてる	-1	1		1		
6	思うようにならなくて悔しい思いをされてらっしゃると思います私だって、予定どうりにいかないときっと同じ思いでしょう>先生からは・「骨髄機能の回復が悪いので・CEFは4回で打ち切り・あとは4回タキソールにしてみましょうか」と>言われました>タキソールの場合は・副作用として脱毛・手足の抹消にしびれがみられることが多い・とのことでした>されている方・そうなのですか?私はタキソールは未経験ですが、どなたかのホームページでタキソールよりタキソテールの方が副作用は強いとありました違っていたらごめんなさい	タキソール	脱毛	みられる	-1				1	副作用 <国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「脱毛(30.5%)」、<海外臨床試験成績>「脱毛」 その他の副作用「12.皮膚 脱毛(45.3%)」
7	思うようにならなくて悔しい思いをされてらっしゃると思います私だって、予定どうりにいかないときっと同じ思いでしょう>先生からは・「骨髄機能の回復が悪いので・CEFは4回で打ち切り・あとは4回タキソールにしてみましょうか」と>言われました>タキソールの場合は・副作用として脱毛・手足の抹消にしびれがみられることが多い・とのことでした>されている方・そうなのですか?私はタキソールは未経験ですが、どなたかのホームページでタキソールよりタキソテールの方が副作用は強いとありました違っていたらごめんなさい	タキソール	手足の抹消	しびれ	-1				1	(※1)重要な基本的注意 「6.末梢神経障害が高頻度に起こるので，観察を十分に行い，症状(しびれ等)があらわれた場合には減量，休薬等の適切な処置を行うこと。」 副作用 <国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「主な副作用は末梢神経障害(37.6%)」 <海外臨床試験成績>「末梢神経障害」 重大な副作用 3. 末梢神経障害，麻痺:「しびれ等の末梢神経障害(43.8%)，麻痺(0.1%)，片麻痺(0.1%未満)，不全麻痺(頻度不明)があらわれることがある」
8	追記:なんだか妙にハイです、タキソールの副作用か、眠れぬ夜を過ごしています	タキソール	眠れぬ夜	過ごしています	-1			1	1	その他の副作用「14. 精神神経系 5%未満 不眠」
9	[14841]Re:[14838]アドバイスお願いします投稿者:すーざん投稿日:20060717(Mon15:46タキソールの夜、私も不眠になりました	タキソール	不眠	なりました	-1			1	1	その他の副作用「14. 精神神経系 5%未満 不眠」
10	白血球も先週に比べて落ち込んでなかったので、タキソール3分の1の点滴をした	タキソール	白血球	落ち込んでなかった	?			1		副作用 <国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「臨床検査値異常は白血球数減少(48.5%)」
11	抗がん剤のタキソールで随分副作用が出て痩せた	タキソール	副作用	出て	-1	1		1		
12		タキソール	副作用	痩せた	-1			1	1	その他の副作用 20. 全身症状 5%未満「体重減少」
13	4月の始めから、手足の先が痺れる副作用が出てきて、これは明らかにタキソールの副作用だったので、しばらくタキソールの投与をやめて様子を見ていたのです	タキソール	手足の先	痺れる	-1			1	1	(※1)
14	他に、抗がん剤は毎週投与の方が3週に1回よりも良いが、タキソールのみ、3週に1回でも同じ効果が得られるとのこと	タキソール	同じ効果	得られる	1	1				
15	タキソールの副作用で、顔のTゾーンにものすごい湿疹ができて以来3ヶ月お化粧していなくて、久々にお化粧するのですが	タキソール	顔のTゾーン	湿疹ができて	-1			1	1	その他の副作用「1. 過敏症 5%～20%未満 発疹」 13. 皮膚 5%未満 皮膚炎」

	薬剤名：出現回数	正解データ			効果：1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用：-1	yes: 1/ no: 0/ 判断不可: -1				
	タキソール:29									
16	September14・2007このブログを購読する気分が上がらない[患者心理]なんだか、とっても気が重い主治医に話すと「タキソールのせいかもしれませんね」と言われたたしかに、タキソールの副作用で抑うつ状態になることがあるらしいこんな日に限ってずっと昔に起こった悲しい思い出の夢を見た昨夜、大学の同級生から突然電話がかかってきた友達という仲でもなく、卒業以来付き合いがあったわけではなく突然の電話「会いたい」と言うなぜ	タキソール	気	重い	-1			1	1	その他の副作用「14. 精神神経系　5%未満　不安，うつ病，精神症状，気分変動」
17		タキソール	抑うつ状態	なることがある	-1				1	その他の副作用「14. 精神神経系　5%未満　不安，うつ病，精神症状，気分変動」
18	タキサン系（タキソール、タキソテールetc）、Fu−系（ゼローダetc）の薬を使用すると手先のピリピリ感がある　ビタミンB6（ピドキサール）を一緒に服用すると、症状に改善効果がみられる　保湿剤（ウレパール　牛車腎気丸　芍薬甘草湯◎乳がんの予防は	タキソール	手先のピリピリ感	ある	-1			1	1	（※1）
19	タキソールは比較的副作用が強く無い方だと主治医から聞いています	タキソール	副作用	強く無い方	1	1				
20	タキソールは殆ど副作用が無く、指先の痺れくらいでした	タキソール	副作用	無く	1	1				
21		タキソール	指先	痺れ	-1			1	1	（※1）
22	みかさんはタキソール終了後1ヶ月半にもなると考えられますけど・それでも浮腫みと痺れはあるんですね	タキソール	浮腫み	ある	-1			1		その他の副作用「20. 全身症状　5%未満　浮腫」
23		タキソール	痺れ	ある	-1			1	1	（※1）
24	[乳がんのこと]毎週1回、合計54回にわたるハーセプチンの治療が終了した乳がんであることがわかった2007年3月15日から願をかけて治療が一区切りつくまでつまりアジュバントのハーセプチンが終了までアルコールを断っていた昨年の夏は、AC、タキソールという結構副作用のある抗がん剤を使っていたため、あまり飲む気自体にならなかった今年は暑い日が続いたので、飲みたくなるかな	タキソール	結構副作用	ある	-1	1				
25	投与後の数日間はちょっと辛そうですが、『タキソール』の方が副作用も少なくなり、家で元気にしております	タキソール	副作用	少なくなり	1	1		1		
26	抗がん剤もCEからタキソールに移り、心配していた副作用もCEの辛さに比べるとかなりやさしく感じますね	タキソール	副作用	やさしく感じます	1	1		1		
27	>3週間に一回がうちの病院の標準治療と言われ、一回目のタキソールを受けましたが、あまりの足の痛さに、治療を投げ出したいくらいでした	タキソール	足	痛さ	-1			1	1	重要な基本的注意「4.関節痛及び筋肉痛が高頻度に起こる」 副作用　<国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「関節痛(23.5%)，筋肉痛(18.7%)」 <海外臨床試験成績>「主な副作用は筋肉痛・関節痛」 その他の副作用　20. 全身症状　5%未満「疼痛」，　21. 筋骨格　20%以上又は頻度不明「関節痛（32.3%），筋肉痛（28.8%）」
28	」と不安な気持ちに陥る自分がいるいま、タキソールの治療を受けているので、その副作用でも関節痛などは出るらしいが、副作用の恐怖よりはるかに再発の恐怖が大きいため、思考は自然とそちらに向かう普段は、いたって前向きで、次は何をしよう	タキソール	関節痛	出るらしい	-1				1	重要な基本的注意「4.関節痛及び筋肉痛が高頻度に起こる」 副作用　<国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「関節痛(23.5%)，筋肉痛(18.7%)」 <海外臨床試験成績>「主な副作用は筋肉痛・関節痛」 その他の副作用　20. 全身症状　5%未満「疼痛」，　21. 筋骨格　20%以上又は頻度不明「関節痛（32.3%），筋肉痛（28.8%）」
29	ハーセプチンとタキソールの治療中と言う事で痺れなどの副作用はでてないでしょうか	タキソール	痺れなどの副作用	でてないでしょうか	-1				1	（※1）

	薬剤名:出現回数	正解データ			効果:1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用:-1	yes: 1/ no: 0/ 判断不可: -1				
	ロキソニン:26									
1	10日から3日間1000mgのパルス、その後プレドニン55mg、サンディミュン200mg開始、相変わらずロキソニンを飲んでいれば何の苦痛もないのでのんきに酸素ボンベをひっぱってウロウロ遊んでいました	ロキソニン	何の苦痛	ない	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
2	いつもは鎮痛剤(ロキソニン)1錠で効くのに効かなくて、もう1錠追加で飲んだら、痛みが治まったよ	ロキソニン	痛み	治まった	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
3	痛みや発熱はロキソニンで抑えた	ロキソニン	痛み	抑えた	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
4		ロキソニン	発熱	抑えた	1			1	1	効能又は効果　薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」
5	あわせてロキソニンも飲み続けるので、胃腸への負担は増し吐き気は相変わらず	ロキソニン	胃腸への負担	増す	-1			1	1	副作用等発現状況の概要「消化器症状(胃部不快感、腹痛、悪心・嘔吐、食欲不振等2.25%)」 その他の副作用「5. 消化器　0.1～1%未満　腹痛、胃部不快感、食欲不振、悪心・嘔吐、下痢」「7. 消化器　0.05%未満　消化不良」
6		ロキソニン	吐き気	相変わらず	-1			1	1	副作用等発現状況の概要「消化器症状(胃部不快感、腹痛、悪心・嘔吐、食欲不振等2.25%)」 その他の副作用「5. 消化器　0.1～1%未満　腹痛、胃部不快感、食欲不振、悪心・嘔吐、下痢」「7. 消化器　0.05%未満　消化不良」
7	その後は和らいできたのだが、AM10時頃再び頭痛が強くなり、再度ロキソニンを服用	ロキソニン	頭痛	強くなる	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
8	2008年04月14日このブログを購読する痛みで、時間外外来受診[通院]実は、前日から腰や背中に痛みを感じていた前夜も、朝もロキソニンを服用しかし、痛みは治まらないそのうち、 腰の痛みは続き 足の感覚がピリピリしだし 胸の辺りもピリピリしだし 腰骨の辺りで締めつけられ感を感じ 足の筋肉の緊張も上がってきたで、これは大変とばかり、大学病院に電話をし、救急外来を受診した診察はいつもの先生だった診察をして、これだけでは分かりにくいので、ということで胸椎のMRI撮影をした結果は、少し影があると言えばあるが、新たに再発したとか、病変が広がったとは言えない模様このまま様子を見て脱力とか感覚異常とかがあったら来てくださいと言われるとにかく再発でないようで安心した腰痛は、足の力が出なくて、少し反ってしか立てないので、そのせいだろうか	ロキソニン	痛み	収まらない	-1			1	-1	
9		ロキソニン	腰の痛み	続き	-1			1	-1	
10		ロキソニン	足の感覚	ピリピリ	-1			1	1	(※2)重要な基本的注意「4.関節痛及び筋肉痛が高頻度に起こる」 副作用　<国内副作用の概要>A法(再審査終了時及び効能追加時の集計)「関節痛(23.5%)，筋肉痛(18.7%)」 <海外臨床試験成績>「主な副作用は筋肉痛・関節痛」 その他の副作用　20. 全身症状　5%未満「疼痛」，　21. 筋骨格　20%以上又は頻度不明「関節痛(32.3%)，筋肉痛(28.8%)」
11		ロキソニン	胸の辺り	ピリピリ	-1			1	1	その他の副作用「21. その他　頻度不明 胸痛」
12		ロキソニン	腰の辺り	締めつけられ感を感じ	-1			1	1	(※2)
13		ロキソニン	足の筋肉の緊張	上がってきた	-1			1	1	(※2)
14	胃痛がきついと、ロキソニンを止め、座薬だけにします	ロキソニン	胃痛	きつい	-1			?	1	副作用等発現状況の概要「消化器症状(胃部不快感、腹痛、悪心・嘔吐、食欲不振等2.25%)」 その他の副作用「5. 消化器　0.1～1%未満　腹痛、胃部不快感、食欲不振、悪心・嘔吐、下痢」「7. 消化器　0.05%未満　消化不良」
15	2006年01月27日平成18年1月27日金本日の体調体温　36.5°体重　46.0kg腹囲　78cm胴囲　70cm昨夜は夕食後にロキソニンを飲んで、寝る時ちょっと苦しかったがすぐ寝れて、朝もスッキリ目が覚めた	ロキソニン	目	覚めた	1			1	-1	

	薬剤名:出現回数	正解データ			効果:1	抽象的	具体的	体験情報	含まれているか	添付文書にある表現
		薬剤名	対象	効果	副作用:-1	yes: 1/ no: 0/ 判断不可: -1				
	ロキソニン:26									
16	少し頭痛も発生してきたので、ロキソニン(頭痛薬)を飲む	ロキソニン	頭痛	発生してきた	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
17	痛みを抑えるのは、ロキソニンか座薬のボルタレンで何とか対応してきましたが、麻薬系の痛み止めを使うことになりました	ロキソニン	痛み	抑える	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
18	今日は頭痛があるのでロキソニンを服用	ロキソニン	頭痛	ある	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
19	私は、手首、指、膝、足首などがかなり痛く、整形で処方されたロキソニンを服用してました	ロキソニン	手首、指、膝、足首など	痛く	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
20	でもこの人、血液製剤とか入るの嫌なんだよねとか言われるが、もともと再発予防として考えていたので、遅かれ早かれ使うつもりだったと答える主治医代行の先生は熱の原因が多発性硬化症による脊髄損傷で体温調整ができにくくなっている可能性があること頭MRI画像にでた病変について、画像の後に症状が出る可能性があることをあげ、多発性硬化症の治療をした方が良いとの意見免疫吸着でも免疫グロブリンでも、と言われるが発熱が収まらない中、吸着は怖いので、グロブリンにしようかと思う今日も午後、37.9℃の熱なかなか下がらないので、ロキソニンを飲んで下げるmixiチェックシェア最終更新日2008年01月20日22時26分05秒コメント(2|コメントを書く[入院(07.6～12胸髄炎、下肢麻痺1医大)]カテゴリの最新記事6椎体の炎症のMRI画像2007年12月29日コメント(2明日から転院2007年08月02日コメント(6採血OK明日、2回目の吸着へ2007年07月31日コメント(6この記事に関連する検索ワード■コメントお名前タイトル顔選択普通うれしい悲涙アイタタえ～っムムッないしょメッセージメッセージ本文は全角で800文字までです	ロキソニン	熱	下げる	1			1	1	効能又は効果　薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」
21	ロキソニンという痛み止めの薬を毎食後飲んで痛みを止めています	ロキソニン	痛み	止めています	1			1	1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
22	4.PostedbyとびうおMay19・200803:34＞こころさま打ち始めた時は、注射後ドドッと疲れが出て、帰りの運転が出来ずに、車の中で寝込んでいたものですが、事前にロキソニンを飲んでおけば、その日は気合で仕事できるようになりました	ロキソニン	仕事	できるようになりました	1			1	-1	
23	ボルタレンも全く効果がなくて、ロキソニンだけが私の痛みを和らげてくれる唯一の救いだった	ロキソニン	痛み	和らげてくれる	1				1	効能又は効果「鎮痛」 薬効薬理「すぐれた鎮痛・抗炎症・解熱作用を有するが、特に鎮痛作用が強力である。」(1)「鎮痛作用」
24	「ロキソニンはまた飲んでみて余計にヒドイ薬疹になったらいけないので、できるだけ別の薬に代えてもらったほうが	ロキソニン	ヒドイ薬疹	なる	-1				1	副作用等発現状況の概要「発疹・蕁麻疹等(0.21%)」 その他の副作用「1. 過敏症　0.1～1%未満　発疹」

付録 C

医療カルテに記載された病歴認識手法

1. はじめに

NTCIR-10 MedNLP-2 パイロットタスク[1]に参加し，行った内容について述べる．著者はテキストの内の病歴（症状や病名）を検出し，それらのモダリティ（陰性，疑惑，そしてそれらが家族の情報かどうか）を判断するタスクに，HokuMed チームとして参加した．タグ付けする情報として日付と時刻の表現を対象としていない．

2. 手法

システムを開発するためのコーパスは，医師によって書かれた模擬患者の病歴約 3300 データであり，タスク主催者によって公開された．データ中に出現する症状，病名のうち，陰性クラスに分類されているものは 1040 カ所，疑惑クラスは 110 カ所，70 カ所は家族クラス（患者の家族に関する情報）であり，3 カ所は陰性クラスと家族クラスどちらにも属している．また，テストコーパスを参考に開発されたシステムが評価できるよう，テストコーパスも配布された．このコーパスには約 2140 カ所の症状，病名が含まれており，700 カ所は陰性クラス，60 カ所は疑惑クラス，40 カ所は家族クラスとしてタグ付けされており，家族クラスと陰性クラスどちらにもタグ付けされたのは 1 カ所のみであった．

提供されたコーパスでは，負例が含まれておらず，学習データとして不十分であるため，モダリティの検出に機械学習を用いることは不適切であると判断した．それゆえ，機械学習の代わりに，我々は手がかり語辞書と構文解析結果に基づいたルールを用いる．しかし，アノテーションされた実例数は機械学習に十分な量であった．従って，最終的な病歴抽出システムは二つのパートから成る．まず始めに，症状・病名が書かれている箇所を特定するために機械学習を用いる．その後，特定された箇所のモダリティを手がかり語辞書とルールを用いて判定する．

2.1. 存在箇所の判定

Lafferty ら[2]によって導入された条件付き確率場 (CRF) を病歴用語を認識するための機械学習モデルとする．CRF は入力データに対しラベルを付与するのに適した学習アルゴリズムである．例えば，トークンが病歴を示すシークエンス（連続したかたまり）である用語の一部であるかどうかをラベル付けすることができる．CRF を実装した直鎖 CRF である CRF++[3]が，医療テキストの用語を識別するために用いられている[4,5,6,7].

CRF モデルを訓練する為に必要な単語列へとコーパスを変換するための機能は存在しないが，日本語はスペースを用いて単語分割されていない[8, p.17]ため，トークン単位に分割する必要がある．文字レベルの分割ではあまりに小さすぎるため，CRF の学習に用いるトークン単位にコーパスを分割するために，日本語構文解析器である CaboCha[9]を用いる．CaboCha を用いた分割結果が適切であるかを調査したところ，一般的にタグ付けされた病歴用語のトークンと境界が一致しており，わずか 5 例において用語の途中で境界が存在していた.

CRF の学習に用いるタグは，IOB タグ[10, pp.763-764]を用いた．つまり，CRF は B（識別したいトークンの始まりのチャンク），I（識別したい病歴を示すトークンの内部），O（識別したいトークンの外側，無関係な部分）の３つのタイプに分類する.

CRF モデルをトレーニングする際，過学習[11, p.10]されないように正規化が用いられる．変数 C の値により決定される L1 ノルム[11, p.146]は CRF++の正規化を行うための手法のひとつある．適切な C の値[11, p.145]により決定された L1 ノルムを用いることにより，不必要な特徴量を生成したモデルの特徴量に含まない．つまりこれは，たくさんの特徴量を与えて学習させたとしても，適切な変数 C を与えることにより，特に影響を与える貢献度の強い特徴量のみが学習モデルとして生成されることを意味する．したがって， L1 ノルムの正規化を用いて CRF モデルを学習させるアプローチを採用し，たくさんの特徴量から学習モデルを生成した．C の値は初期値 4 の山登り法を用いて決定した．各 C の値を用いて学習し，10 分割交差検定によりその結果を評価した．学習したモデルによって得られた IOB で示される出力結果は，一行に一文字ずつ配置され，CoNLL 2000 共有タスク[12]の評価用スクリプトと照らし合わせることで評価する.

図１に示すように，以下の特徴量を CRF モデルを学習する為に与える.

・トークン（現在，先に出現した２つ，後に出現する４つのトークンを用いる）
・トークンの品詞（現在，先に出現した４つ，後に出現する４つのトークンの品詞を用いる）
・病歴用語の辞書にトークンが含まれているかどうか（現在，先に出現した４つ，後に出現する４つについて判断した結果を用いる）用いた辞書については後に説明する.
・トークンが発生したヘッダのカテゴリ．例えば，「主訴」や「現病歴」など．（現在のトークンの情報のみ用いる）
・トークンの接尾辞．病歴辞書中で頻繁に使われるような接尾辞を含んでいるかどうか．（現在のトークンと後に出現する４つのトークンの情報を用いる）

・トークンの文字の種類．ひらがな，カタカナ，漢字，ローマ字，数字など．（現在のトークンと，先に出現した４つ，後に出現する４つのトークンの情報を用いる）
・最も一般的なモダリティを示す手がかり語の何かが，チャンク（文節）内，もしくは係り先チャンクに含まれているかどうか．文節の判定には構文解析器を用いる．

辞書を用いたマッチングに利用したのは，以下に示す病歴に関する辞書である．

・病名診断リスト[13]
・病気（C）と精神障害（F03）のノードに分類される MeSH に登録された用語[14]
・調査，社会状況，外科的および内科処置として分類されるものを除いた MedDRA 用語[15]
・英語の SNOMED CT[16]の障害カテゴリに属する英語の用語
・英語の SNOMED CT[16]の障害カテゴリに属する用語のうち，JMDict 辞書[17]を用いて日本語に翻訳することができた用語
・日本語で書かれた患者のブログコーパスから自動的に分布セマンティクスを用いて抽出された病歴を示す用語を，手動でフィルタリングしたもの[18]

接尾辞のリストは，MedDRA と MeSH から重複なしリストを作成し，そのリスト内に接尾辞として少なくとも２回発生した１文字もしくは２文字の接尾辞を抽出することにより構築した．CRF の特徴量として与える時には，１文字で一致した接尾辞よりも，２文字で一致した接尾辞を優先する．

Word	POS	Dictionary	Header	Suffix	Orthography	Modality	Result	
嚥下	名詞, サ変接続	B-complaint	noheader	嚥下	kanji	O	B-complaint	
障害	名詞, 一般	I-complaint	noheader	障害	kanji	O	I-complaint	
を	助詞, 格助詞	O	noheader	O	hiragana	O	O	
主	接頭詞, 名詞接続	O	noheader	O	kanji	O	O	
訴	名詞, 一般	O	noheader	訴	kanji	O	O	← current
に	助詞, 格助詞	O	noheader	O	hiragana	O	O	
来院	名詞, サ変接続	O	noheader	O	kanji	O	O	
し	動詞, 自立	O	noheader	O	hiragana	O	O	
た	助動詞,*	O	noheader	O	hiragana	O	O	
.	記号, 句点	O	noheader	O	symbol	O	O	

図 1. CRF の学習に用いる特徴量
（現在のトークンに対しては赤字の特徴量を用いる）

2.2 モダリティ検出

モダリティ検出の一般的な考え方は，陰性，疑惑，家族クラスの手がかり語辞書に頼るものである．つまり，病歴は，陰性の手がかり語に影響を受けていれば陰性クラスと判定され，疑惑の手がかり語に影響を受けていれば疑惑クラスと判定され，家族の手がかり語に影響を受けていれば家族クラスと判定される．「陰性」「陰性」などの名詞，ならびにこれらの名詞+「する」「だ」は陰性を表現することができる．しかしながら，陰性表現として最も一般的な

ものは，動詞や形容詞の語尾を変化させて「ない」「ぬ」[8, p.54]が付いた形にすることである．また，テストデータ中に存在しない陰性辞書に含まれる手がかり語を削除することで，頻繁に利用される手がかり語のみを網羅することが可能であると考えている．確実性のレベルを記述するための文法的な形も存在する[8, p.118]が，頻繁に用いられる疑惑を示す非文法的な手がかり語が多数存在する．

それゆえ，陰性を示す文法的特徴や他の陰性を示す手がかり語を見つける為に，開発用コーパスを調査した．疑惑クラスを判定するための手がかり語も多数存在しているだろうが，今回は英語の疑惑に関する手がかり語リストを翻訳する方法を用いる．英語の手がかり語リストは英語で書かれたテキストに対する今までの研究で構築されたものであり，英語のモダリティ検出をスウェーデン語に適応させるためにスウェーデン語に翻訳されている[19]．これらの手がかり語リストのうち, 英語のものを JMDict[17]を用いて自動的に日本語へと翻訳した．その後，第一母国語が日本語である１人により手動で評価し，日本語において疑惑を示す手がかり語として適切なものだけを残した．

最後に，構築したモダリティ検索システムをタグ付けされた病歴を分類するために利用し，その出力結果のエラー分析を受けて，手がかり語リストをわずかに変更した．

手がかり語が，言及されている病歴に影響を与えるかどうかの判断は，一連のルールにより行う．陰性クラスと疑惑クラスの判断のために実装されたルールは，構文解析器である CaboCha[9]の出力結果に依存している．一般的に，同じチャンク（文節）もしくは係り先のチャンク（文節）に手がかり語が存在するとき（すなわち，係り受け関係のグラフで見たときに近い位置にあるとき），病歴に影響を与える．しかし，フレーズ分割によりそれらの手がかり語が病歴に関する記述と分割されたときは，影響を与えないとした．動詞に後続する接続後である「が」「から」「による」「ため」「に対して」，普通名詞以外の後ろに付いているコンマ（読点），連用形の動詞をフレーズ分割には用いた．

３つの陰性手がかり語リストをチェックし，リストに含まれる手がかり語が奇数個存在していた場合に，陰性と判断する．３つの陰性手がかり語リストは以下の通りである．

・陰性の名詞（(-)，陰性，否定，予防，消失，清明，良好，正常，不可能，整，清）
・陰性の述語（ない，無い，ぬ，なし）
・改善に関する単語（改善）

疑惑に関する判断には，上述した７１の手がかり語を含む疑惑手がかり語語リストを用いた．家族クラスのモダリティを判断するためには，行の先頭（ヘッダー）が家族手がかり語リスト（母親，父親，姉，兄，家族，息子，母，娘，父,弟,妹,両親,家族歴,妻,夫）に含まれているかどうか，もしくは「家族歴」のヘッダーの下に存在しているかどうかによって判断した．もし，いくつかのモダリティが病歴に付与された場合は，同時に複数のモダリティを付与する（例えば「陰性，疑惑」）．

また，病歴と同チャンクに陽性モダリティの手がかり語である「あり」が存在していた場合は，付与されている陰性や疑惑モダリティを却下する．「止まらず」や「(+)」が同チャンクに存在していた場合についても同様に，陰性や疑惑モダリティを却下する．最後に，「意識

清明」は開発用コーパス中で常に陰性として出現していたため，内容に関わらず，常に陰性モダリティを付与した．

2.3 テストコーパスへのタグ付与

開発用コーパスを用いて学習された CRF モデルを 10 分割交差検定で評価したときに，最も病歴が高精度に検出できたときの C 値を用いる．そして，このモデルを病歴を認識し，タグ付けするために用いた．その後，最終的なバージョンである手がかり語辞書や病歴を発見するためのルールを用いて，モダリティを判定した．

最適な C 値はコーパスによって変動するため，二つ目のタグ付けには最も”安全”な値である 1 を C 値として用い, 三つ目のタグ付けには少ない特徴量だけ（前後一単語のみ用いたもの）で学習させ，10 分割交差検定を行った際最も高精度であった C 値を用いる．

テストコーパスに自動的にタグ付けした出力結果は，手動でタグ付けした結果と比較する為に，主催者に提出した．

3. 結果

開発用コーパスでタグ付けを行ったなかで，最大の F 値は 83.19 であり，そのとき用いた C 値は 0.69141453 であった．この F 値に対する適合率と再現率は表 1 で示される．また，表 1 には開発用コーパスを用いて作成したルールと，最終バージョンの手がかり語辞書を用いて判定したモダリティの検出結果も載せている．テストコーパスに対する評価結果は，表 2 に示す．最終バージョンの手がかり語辞書，ルールと，開発用データから学習された 3 つのモデルを用いて検出した結果について載せている．

表 1.　開発用コーパスに対する結果

Submission		NER (complaints)	Positive	Family	Negation	Suspicion	Fam.+Neg.	Susp.+Neg.
1	Precision	87.30%	82.14%	96.88%	85.84%	69.70%	14.29%	0.00%
(c=0.69141453)	Recall	79.45%	75.77%	86.11%	74.90%	63.30%	33.33%	0.00%
	F-score	83.19	78.83	91.18	80.00	66.35	20.00	0.00
2	Precision	87.03%	81.87%	96.88%	85.84%	69.47%	14.29%	0.00%
(c=1)	Recall	78.42%	75.05%	86.11%	73.75%	60.55%	33.33%	0.00%
	F-score	82.50	78.31	91.18	79.34	64.71	20.00	0.00
3	Precision	87.63%	82.29%	96.92%	86.12%	70.83%	14.29%	0.00%
(c=0.629,	Recall	78.87%	74.76%	87.50%	74.90%	62.39%	33.33%	0.00%
fewer features)	F-score	83.02	78.34	91.97	80.12	66.34	20.00	0.00

表 2. テストコーパスに対する結果

Submission		NER (complaints)	NER (time & complaints)	Positive	Family	Negation	Suspicion
1	Precision	87.21%	87.21%	87.21%	77.78%	76.18%	48.84%
(c=0.69141453)	Recall	76.92%	65.59%	76.92%	80.77%	68.99%	51.22%
	F-score	81.74	74.87	81.74	79.25	72.41	50.00
2	Precision	87.47%	87.47%	87.47%	80.77%	76.44%	48.84%
(c=1)	Recall	76.12%	64.91%	76.12%	80.77%	68.54%	51.22%
	F-score	81.40	74.52	81.40	80.77	72.27	50.00
3	Precision	88.62%	88.62%	88.62%	77.78%	76.56%	46.67%
(c=0.629,	Recall	75.09%	64.03%	75.09%	80.77%	68.99%	51.22%
fewer features)	F-score	81.30	74.35	81.30	79.25	72.58	48.84

4. 考察とまとめ

開発用コーパスに対する病歴の認識結果でもっとも良い結果であったものは，適合率が 87.30%，再現率が 79.45%であった．これには時間表現に対するタグ付け結果は含まれていない．テストセットに対する病歴の認識結果でもっとも良い結果であったものは，適合率が 87.21%，再現率が 76.92%であった．これについても，時間表現に対するタグ付け結果を含んでいない．我々は時間表現のタグ付けは行っていないため，NER の再現率はいずれについても 65.59%と低くなっている．開発用コーパス，テストセットどちらについてもサブミッション 1 の値が高くなっている．リコールがやや劣るが，適合率は英語コーパスに対する既存研究の結果と一致している（例えば, Wang と Patrick[20]の結果では, 適合率が 84%, 再現率が 82%であり，Jiang 達の結果では適合率が 87%，再現率が 84%であった[6]. ）．

著者らのモデルは開発用コーパスに対して最適化してあるため，テストセットに対して実行したときリコールはわずかに低くなった．これは，開発用コーパスに依存しない，英語の手がかり語を翻訳することで得られた疑惑手がかり語リストを除いて，開発用コーパスに依存している他のモダリティを判断する為の手がかり語リストや範囲のルールをより拡張させる必要があることを示している．

開発用コーパスに対して，手がかり語とルールに基づき手動でモダリティを割り当てたところ，いつも成功した訳ではなかった．例えば，「意識清明」は開発用コーパス中では常に陰性クラスに割り当てられていたため，いつも陰性とタグ付けしていた．しかしながら，テストセットで出現した「意識清明」は全て陽性であり，このルール設定は適切ではなかったと考えられる．

開発用コーパスに対するモダリティ割り当て結果は，陽性モダリティについて適合率 82%,

再現率 76%，陰性モダリティについて適合率 85%，再現率 74%，疑惑モダリティについて適合率 70%，再現率 63%，家族モダリティの検出が最も高精度であり，適合率 97%，再現率 86%という結果となった．開発用コーパスに適合させたために生じた，手がかり語の分類に関わる矛盾と係り受け解析ミスは，より高精度な結果を出すことを困難にしている．

テストセットの結果は，一般的に予想されるように，開発用コーパスに対する結果よりも低くなった．陽性モダリティについて適合率 87%，再現率 77%，陰性モダリティについて適合率 76%，再現率 69%，疑惑モダリティについて適合率 49%，再現率 51%，家族モダリティについて適合率 78%，再現率 81%という検出結果となった．既存の英語のコーパスに対する機械学習ベースのモダリティ検出研究では，より良い結果となっている．例えば，Clark ら[21]の研究では，存在カテゴリ（陽性に相当）への検出は 94%の適合率，98%の再現率であり，不在カテゴリ（陰性に相当）への検出は適合率 95%，再現率 92%であり，そして可能性カテゴリ（疑惑に相当）への検出は 77%の適合率，53%の再現率であった．しかしながら，この機械学習モデルはこの共有タスクで利用できた開発用コーパスよりも大きい開発用セットを用いて訓練している．それゆえ，機械学習モデルは固有表現抽出には適しているものの，モダリティ検出については学習に利用できるデータの量から考えて，手がかり語とルールベースを用いた手法が適切であったと考えている．

参考文献

[1] NTCIR-10 “Medical Natural Language Processing (MedNLP)” Pilot Task. http://mednlp.jp/medistj-en/ , Accessed 2014-12-16.

[2] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proc. 18th International Conf. on Machine Learning, pages 282–289. Morgan Kaufmann, San Francisco, CA, 2001.

[3] T. Kudo. CRF++: Yet Another CRF toolkit. http://crfpp.sourceforge.net/, 2012. Accessed 2012-06-15.

[4] Y. Wang. Annotating and recognising named entities in clinical notes. In Proceedings of the ACL-IJCNLP Student Research Workshop, pages 18–26, Singapore, 2009.

[5] J. Patrick and M. Li. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. J Am Med Inform Assoc, 17(5):524–527, Sep-Oct 2010.

[6] M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc, 2011.

[7] M. Skeppstedt, M. Kvist, G. H. Nilsson, and H. Dalianis. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. J Biomed Inform, 49:148–58, Jun 2014.

[8] M. Kamermans. An Introduction to Japanese Syntax, Grammar & Language. Sjgr Publishing, 2010.

[9] CaboCha. Cabocha: Yet another Japanese dependency structure analyzer. https://code.google.com/p/cabocha/, 2012.

[10] D. Jurafsky and J. H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, second edition, Feb. 2008.

[11] C. M. Bishop. Pattern recognition and machine learning. Springer, New York, NY, 2006.

[12] Conll. CoNLL-2000. http://www.cnts.ua.ac.be/conll2000/chunking/, Accessed 2011-10-09.

[13] MEDIS-DC. MEDIS Byomei Master v3.11. http://www2.medis.or.jp/stdcd/byomei/index.html, 2014.

[14] U.S. National Library of Medicine. MeSH (Medical Subject Headings). http://www.ncbi.nlm.nih.gov/mesh, 2014.

[15] IFPMA. Meddra introductory guide version 14.0. http://www.who.int/medical_devices/innovation/MedDRAintroguide_version14_0_March2011.pdf, 2011.

[16] IHTSDO. International Health Terminology Standards Development Organisation: SNOMED Clinical Terms User Guide, July 2008 International Release. http://www.ihtsdo.org, 2008. Accessed 2011-01-24.

[17] JMdict. The jmdict project. http://www.edrdg.org/jmdict/j_jmdict.html, 2013.

[18] M. Ahltorp, M. Skeppstedt, S. Kitajima, R. Rzepka, and K. Araki. Medical vocabulary mining using distributional semantics on Japanese patient blogs. In Proceedings of SMBM 2014 - The 6th International Symposium on Semantic Mining in Biomedicine, 2014.

[19] S. Velupillai, M. Skeppstedt, M. Kvist, D. Mowery, B. E. Chapman, H. Dalianis, and W. W. Chapman. Cue-based assertion classification for Swedish clinical text – developing a lexicon for pycontextswe. Artif Intel l Med, Jan 2014.

[20] Y. Wang and J. Patrick. Cascading classifiers for named entity recognition in clinical notes. In Proceedings of the Workshop on Biomedical Information Extraction, pages 42–49, 2009.

[21] C. Clark, J. Aberdeen, M. Coarr, D. Tresner-Kirsch, B. Wellner, A. Yeh, and L. Hirschman. Mitre system for clinical assertion status classification. J Am Med Inform Assoc, 18(5):563–7, 2011.