



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 [論文内容及び審査の要旨]
Author(s)	北嶋, 志保
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第11762号
Issue Date	2015-03-25
Doc URL	https://hdl.handle.net/2115/58912
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Kitajima_Shiho_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 北嶋 志保

学 位 論 文 題 名

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
(Study on Drug Effects Extraction for Using Information from Patient Texts)

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、励みとなる可能性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.