
ポスター発表

[PB] ポスター B

2020年6月5日(金) 09:00 ~ 18:30 ポスター会場(2) (e-poster)

[PB-29] 医薬品特定に必要な形態素に関する調査

Investigation of the drug morphemes required for drug identification

*佐藤 弘康¹、蝦名 勇樹¹、真井 雄規¹、島津 智行¹、喜多 力¹、田村 広志¹、渡辺 浩明¹ (1. JA北海道厚生連 帯広厚生病院 薬剤部)

*Hiroyasu Sato¹, Yuki Ebina¹, Yuki Sanai¹, Tomoyuki Shimazu¹, Chikara Kita¹, Hiroshi Tamura¹, Hiroaki Watanabe¹ (1. Department of Pharmacy, Obihiro Kosei General Hospital)

医薬品特定に必要な形態素に関する調査

佐藤 弘康、蝦名 勇樹、真井 雄規、島津 智行、喜多 力、田村 広志、渡辺 浩明
JA 北海道厚生連 帯広厚生病院 薬剤部

Investigation of the drug morphemes required for drug identification

Hiroyasu Sato, Yuki Ebina, Yuki Sanai, Tomoyuki Shimazu, Chikara Kita,
Hiroshi Tamura, Hiroaki Watanabe

Department of Pharmacy, Obihiro Kosei General Hospital

医薬品名称は、語幹、規格量、剤型、屋号など、独自の形態素から構成されるが、実際に医療テキストとして記載あるいは入力される場合には、一部の形態素が省略される場合が多い。医療テキスト情報に包含される医薬品名に関する情報から有用な知見を得るためには、その医薬品の特定が重要である。今回、我々は独自に開発した医薬品名に特化した形態素解析アルゴリズムを使用し、承認医薬品名について、どのような形態素を指定した場合に個別医薬品コード(YJコード)が何桁まで一意に特定可能であるかを調査した。

その結果、ブランド名のみであっても41.3%は一意で医薬品を特定可能であること、規格量や剤型を追加指定することによって、特定率が向上することを明らかにした。一方で、標準的医薬品コードであるYJコードの仕様上の問題から、医薬品特定が困難である事例も観察された。

キーワード 医薬品名、形態素解析、自然言語処理、個別医薬品コード、標準コード

1. はじめに

近年、診療録をはじめとする医療テキスト情報から、自然言語処理 NLP 等により新たな知見を探索する研究が多く行われている。医療テキスト情報の中には医薬品名に関する情報が包含されることも多いが、医薬品の情報に関する利活用は十分ではない。医薬品名称は、規格量・剤型・濃度・デバイス・屋号など独特な形態素から構成されており、医療テキスト情報に含まれる医薬品に関する情報から医薬品を同定するためには、これらの形態素を特定する必要がある。一般に、医療テキスト情報に含まれる医薬品名は、正式な承認医薬品名がそのまま記載されることは少なく、剤型や規格量が省略されるケースが多いと推測される。

そこで、医薬品名に関する形態素をどの程度指定した場合に医薬品を一意に同定できるのかについて、臨床で最も使用されている医薬品標準コードである個別医薬品コード(YJコード)を用いて調査を行った。

2. 方法

1) 対象データ

2018年12月9日に診療情報提供サービスのホームページよりダウンロードした承認医薬品リストを用いた。

2) 医薬品の形態素解析

医薬品の独特な形態素に切り分けるために、我々が Visual Basic for Application にて独自に開発し、Excel のユーザー定義関数「DrugMA」として実装したアルゴリズムを使用した。本アルゴリズムにより、すべての承認医薬品名を「一般名」、「ブランド名」、「規格量」、「濃度」、「剤型」、「デバイス」、「屋号」等の各形態素に切り分けた[1],[2]。

3) 統合 YJ コードの計算

YJコードは、12桁から構成される標準医薬品コードであり、電子カルテの処方オーダー等において汎用されているコード体系の1つである。YJコードは8桁目のみがアルファベットで表記され、その他の桁は数字で表記される。

医薬品名の語幹(「一般名」または「ブランド名」)に「規格量」、「濃度」、「剤型」、「デバイス」、「屋号」の各形態素を組み合わせた場合に、どの程度 YJコードを特定可能かについて調査した。

ある形態素の組合せが示す医薬品群において YJコードの各桁が一意である場合、その数値あるいはアルファベットを用い、複数の値を取りうる場合にはワイルドカード(数値の場合は“X”、アルファベットの場合は“@”)を用いて、その医薬品群を代表する YJコード(統合 YJコード)を表現した。

4) 医薬品形態素辞書の作成

承認医薬品名の各語幹において、あらゆる形態素の組合せにおける統合 YJ コードを算出し、医薬品名形態素解析用の辞書を構築した。

5) 医薬品特定に必要な形態素に関する調査

構築した医薬品名形態素辞書を用いて、形態素の組み合わせごとに統合 YJ コードの特定率(ワイルドカードの非存在率)を、先頭 4 桁、7 桁、8 桁、9 桁、12 桁にて調査した。

3. 結果

16,708 品目の承認医薬品名を形態素に分解し、あらゆる形態素の組合せにおける統合 YJ コードを算出した結果、50,027 件の医薬品名形態素辞書を構築した。

語幹のみで他の形態素を指定しない場合、統合 YJ コードの全 12 桁の特定率は、語幹を「ブランド名」とした場合と「一般名」とした場合で、それぞれ 41.3%、16.1%であった。一方で、語幹のみで統合 YJ コードの先頭 7 桁が特定できた割合は、それぞれ 87.8%、88.4%であった。

形態素を追加指定するたびに統合 YJ コードの 12 桁特定率は上昇がみられた。5 つすべての形態素を指定した場合の統合 YJ コード 12 桁特定率は、語幹を「ブランド名」および「一般名」としたいずれの場合においても 100%であった。

形態素として、「剤型」あるいは「デバイス」を指定しない場合、統合 YJ コードの 8 桁特定率は低下する傾向が見られた。また、「規格量」あるいは「濃度」を指定しない場合には、9 桁特定率が低下する傾向が見られた。

4. 考察

医療テキスト情報に含まれる医薬品名に関する情報を有効に活用するためには、その情報から医薬品を特定する必要がある。今回、YJ コードへの紐づけを指標として、医薬品名に特化した形態素解析用辞書を構築した。

今回使用した医薬品名に特化した形態素解析アルゴリズムである DrugMA は、今後さらなる改良の余地があるものの[1]、今後、今回構築した辞書を医療テキスト情報に適用することで、医薬品名に関する情報から統合 YJ コードへの変換が可能とな

ると考えられる。

今回の調査において、語幹が「ブランド名」である場合には、他の形態素が指定されなくても約 4 割は YJ コードが一意に特定できることが示された。一方、語幹が「一般名」である場合には、語幹のみでの特定率は 16.1%と低く、一意に医薬品を特定することは困難であることが示唆された。

YJ コードは、先頭 4 桁で薬効分類を、先頭 7 桁で医薬品の成分と投与経路(内服、外用、注射)を表す仕様となっており、同一成分であっても、適応症の薬効群が異なる場合には、先頭 4 桁は一致せず、同一成分であっても、内服薬と注射薬では先頭 7 桁は異なる。今回の調査において、「ブランド名」の語幹のみで YJ コードの先頭 7 桁が特定できなかった事例が 12%程度確認されたのは、これらの理由と考えられる。

YJ コードの 8 桁目は剤型を表すアルファベットであり、「剤型」の形態素が指定されていない場合に 8 桁特定率が低下することは矛盾しない結果であった。また、「ペン」、「エリプタ」等の「デバイス」を指定した場合には、剤型が推測可能であるため、「剤型」が指定されていないでも 8 桁特定率は低下しない傾向が見られた。

5. 結語

医療テキスト情報から医薬品に関する有効性や安全性を特定するためには、YJ コードとして 7 桁から 9 桁の特定が必要であると考えられる。今回構築した辞書により今後、医療テキスト情報への応用が可能になると考える。一方で、YJ コードの仕様上の問題から、医薬品特定のために用いる標準医薬品コードとして YJ コードが適切であるかについて検証が必要である。

参考文献

- [1] 佐藤弘康、蝦名勇樹、真井雄規, 他: 医薬品名の形態素解析アルゴリズムの構築, 第 39 回医療情報学連合大会, 2019.
- [2] 蝦名勇樹、佐藤弘康、真井雄規, 他: 承認医薬品名における形態素構成の調査, 第 39 回医療情報学連合大会, 2019.