

医薬品副作用情報を用いた副作用検索システムの提案

三上拓也^{†1} 駒田康孝^{†1} 野口保^{†2} 菅野敦之^{†3} 山名早人^{†4†5}

医薬品の服用に伴う副作用の早期発見と対策は、医療現場において重要な課題である。副作用の早期発見と対策のためには、医師や薬剤師が全医薬品の全副作用を把握しておく必要がある。しかし、一つの医薬品に既知の副作用は数多くあり、全ての副作用の把握は困難である。また副作用には同義に解される類似表記が多くあり、同一の副作用にも関わらず表記違いにより異なる副作用として誤認し、副作用の発見が遅れる可能性がある。さらに、医薬品との関連性が立証されていない未知の副作用も想定される。そこで本稿では副作用の表記ゆれに頑健かつ、医薬品の未知の副作用検索に対応した副作用検索システムを提案する。提案手法では、医薬品の添付文書中の副作用や、副作用が疑われる症例報告を元に医薬品の未知の副作用を推定する。実験では、実際に副作用が疑われる症例報告があった事例150件を入力し、副作用の検索結果と、副作用が疑われる症例報告にある、医薬品との関連性が疑われる副作用を比較することにより有用性を評価した。実験の結果、副作用の検出率は72.7%であり、うち42.2%を未知の副作用として検出した。また従来、表記ゆれにより同一の副作用として検出できなかった既知の副作用29.3%を、副作用の表記ゆれを解消して同一の副作用として検出でき、提案手法が有用であることを確認した。

1. はじめに

医療において、医薬品を用いた治療行為である薬物療法は、疾患の予防や治療のために、多くの医療現場で利用されている。現在、日本国内においては、医師の診断により処方される医療用医薬品は約18,000品目、医師の処方なく薬局や薬店で購入できる一般用医薬品は約12,000品目に上る。このように、医薬品は医療において必要不可欠なものとなっている。一方、医薬品の服用に伴う、意図せずに有害な症状が生じる副作用が大きな問題となっている。副作用は、めまいや食欲不振といった比較的軽い症状を呈するものから、アナフィラキシーや肺炎など重篤な症状を呈するものまで多様である。副作用を早期に発見し対策を講じることは、医療現場において重要な課題である。

しかし医師や薬剤師が、医薬品の服用に伴う副作用を早期に発見し対策を行う上で、以下の3つの課題が存在する。

1. 医薬品に対する副作用の数
2. 未知の副作用の問題
3. 副作用の表記ゆれの問題

1つ目の課題として、医薬品に対する副作用の数が挙げられる。医薬品と副作用は一对多の関係にあり、発症頻度の低い副作用や症状の軽い副作用など、各医薬品の全ての副作用を医師や薬剤師が把握することは困難である。また、医療現場の限られた問診時間の中で、医薬品の数多くある副作用を探し出すことは困難である。

2つ目の課題として、未知の副作用の存在が挙げられる。未知の副作用とは、医薬品との関連性が立証されていない副作用を指す。医薬品の副作用は、臨床試験や類似成分の医薬品の研究結果を元に立証される。そのため、臨床試験時に発見されなかったものの、市販後に医薬品と副作用の

関連性が疑われ、後に医薬品の副作用と立証される副作用も存在する。しかし、未だに医薬品の副作用として知られていない、または立証されていない未知の副作用が存在する可能性がある。そのため、未知の副作用の発見と対策が遅れる危険性がある。

3つ目の課題として、副作用の表記ゆれの存在が挙げられる。副作用の表記ゆれとは、副作用表記のばらつきを指す。医薬品の副作用等の情報は、医薬品に添付されている添付文書に記載されている。しかし、添付文書にある副作用の表記には同義に解される類似表記が少なくない。そのため、作成された添付文書によって表記ゆれが生じる。例えば、「アナフィラキシー」という用語に対し総称並びに、付随症状に関する複数の表記が存在する。また、「腸管閉塞」という用語に対して、同一の疾患でありながら全く表記が異なる用語も存在する。表記ゆれは、医師や薬剤師が、同一の副作用にも関わらず異なる副作用として誤認する一因となり、副作用を見逃す危険性がある。

医薬品と副作用の因果関係を推定する研究は以前より盛んに行われている。医薬品と副作用の因果関係推定の研究における代表例として、副作用が疑われる症例報告の統計データを用いたシグナル検出に関する研究[1][2][3][4]が挙げられる。シグナル検出は、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、シグナル検出は副作用と疑われる症例報告に依存するため、マスコミ等の社会的要因の影響を受けやすく、副作用と疑われる症例報告の件数に偏りや変動が生じる可能性がある。一方で、副作用と疑われる症例報告の偏りによる影響を少なくするために、近年普及が進む電子カルテの情報から、医薬品と副作用の因果関係を推定する研究[5]が行われている。

一方で、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いる。これらの医薬品検索サービスは、医薬品の既知の情報に基づいているため、未知の副作用の問題や副作用の表記ゆれの問題が残り、欲する情報が得られない可能性がある。また、医薬品の検索システムに関する研

^{†1} 早稲田大学大学院 基幹理工学研究科

^{†2} 明治薬科大学 薬学教育研究センター

^{†3} 明治薬科大学 実務実習部門

^{†4} 早稲田大学 理工学術院

^{†5} 国立情報学研究所

1 独立行政法人 医薬品医療機器総合機構, 医療用医薬品の添付文書情報,
http://www.info.pmda.go.jp/psearch/html/menu_tenpu_base.html

究[6]も行われている。しかし、医薬品の検索システムに関する研究も、未知の副作用の問題や副作用の表記ゆれの問題が残る。

そこで本稿では、医薬品の副作用等が記載された添付文書を用いて、頻出する副作用のパターンから医薬品の未知の副作用を推定する。また副作用が疑われる症例報告を用いて、シグナル検出により医薬品の未知の副作用を推定する。そして、副作用の表記ゆれに頑健かつ、推定した未知の副作用の検索に対応した副作用検索システムを提案する。本システムにより医師や薬剤師が、医薬品の副作用を早期に発見できるようになると期待される。

2. 関連研究

本節では、未知の副作用推定に関する研究として、医薬品と副作用の因果関係を推定する研究[1][2][3][4][5]について述べる。また、医薬品の検索システムに関する研究[6]及びサービスについて述べる。

2.1 医薬品と副作用の因果関係推定に関する研究

医薬品と副作用の因果関係推定に関する研究の代表例として、副作用が疑われる症例報告の統計データを用いたシグナル検出[1][2][3][4]が挙げられる。シグナル検出は、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。因果関係の評価値であるシグナルが検出された場合に、対象の医薬品と副作用の間には関連性があると推定される。

シグナル検出では、副作用が疑われる症例報告の報告件数を基に、因果関係を推定する対象の医薬品と対象の副作用を表1の形式で集計し、因果関係の推定を行う。

表1 対象の医薬品と副作用の症例報告件数

	対象の医薬品	その他の医薬品	合計
対象の副作用	N_{11}	N_{12}	N_{1+}
その他の副作用	N_{21}	N_{22}	N_{2+}
合計	N_{+1}	N_{+2}	N_{++}

Evans ら[1]は、対象の医薬品及びその他の医薬品に対する、対象の副作用の報告割合の比に着目したシグナル検出手法 Proportional Reporting Ratio (PRR) を提案した。PRR は、シグナル値の計算式が単純で計算量が少ない手法である。また、性別や年齢などを考慮したシグナル検出ができる。一方、副作用と疑われる症例報告が少ない場合、シグナル値が不安定となる問題がある。

Rothman ら[2]は、対象の副作用及びその他の副作用が起こる場合のオッズ比に着目したシグナル検出手法 Reporting Odds Ratio (ROR) を提案した。ROR は他のシグナル検出手法と比較し、検出感度が高いと言われている。一方、感度が高いため因果関係がない医薬品と副作用をシグナルありと誤検出する可能性がある。

Bete ら[3]は、ベイズ的アプローチに基づくシグナル検出手法 Bayesian confidence propagation neural network (BCPNN)

を提案した。BCPNN は、対象の医薬品及び対象の副作用に対する報告の確率から、対象の医薬品-副作用の組の発生する確率を推測し、実際の発生率と比較する手法である。ベイズ的アプローチに基づくため、副作用と疑われる症例報告が少ない場合でも、シグナル値は安定する。

また、BCPNN と同様にベイズ的なアプローチに基づくシグナル検出手法として、DuMouchel[4]は Gamma-Posission Shrinker (GPS) を提案した。GPS は性別や年齢などを考慮した評価を行うことができるとしている。

一方、シグナル検出手法は、副作用と疑われる症例報告がマスコミ等の社会的要因の影響を受けやすく、症例報告数に偏りや変動が生じる可能性が指摘されている。そこで、Hrpaz ら[5]は、副作用と疑われる症例報告の偏りや変動を解消するため、電子カルテを用いた副作用に対するマイニング手法を提案した。近年普及が進む電子カルテは、患者に関する情報が継続的に記載されている。そのため、社会的要因の影響を受けづらいつと考えられるためである。

2.2 医薬品検索に関する研究及びサービス

医薬品検索に関する研究として、平川ら[6]は、先発医薬品の特許が切れた後に製造販売された、同一成分、同一効能の後発医薬品の検索システムを提案した。平川らは、転置 index を用いることで、先発医薬品の医薬品名、成分、効能が類似する後発医薬品の検索を可能にした。また、n-gram による医薬品名の表記ゆれ解消が行われている。

一方、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いる。例として、医薬品添付文書検索サービス¹が挙げられる。医薬品添付文書とは、医薬品の効能や成分、副作用などが記載されている文書である。医薬品検索サービスは、医薬品の名称や成分、効能、副作用など検索したい場合には有用なサービスである。

これらの医薬品検索に関する研究及びサービスは、医薬品の既知の情報を用いて、副作用等の医薬品情報や類似医薬品の検索が行われている。そのため、医薬品の未知の副作用を検索できず、未知の副作用の発見と対策が遅れる問題が生じる。また、医薬品名の表記ゆれ解消は行われているが、副作用に対する表記ゆれの解消は試みられていない。そこで我々は、医薬品と副作用の因果関係を推定することで、副作用の表記ゆれに頑健かつ、未知の副作用の検索に対応した医薬品の副作用検索システムの構築を行う。

3. 提案手法

提案する副作用の表記ゆれに頑健かつ、未知の副作用の検索に対応した副作用検索システムは、図 1(a)に示す副作用関連性検索システム、図 1(b)に示す類似医薬品検索システム、図 1(c)に示す未知の副作用検索システムの3つから

¹ 独立行政法人 医薬品医療機器総合機構、医療用医薬品の添付文書情報、http://www.info.pmda.go.jp/psearch/html/menu_tenpu_base.html

構成される。提案するシステムにおける、副作用の表記ゆれ解消手法を図 2(a)に、未知の副作用推定手法を図 2(b)に示す。

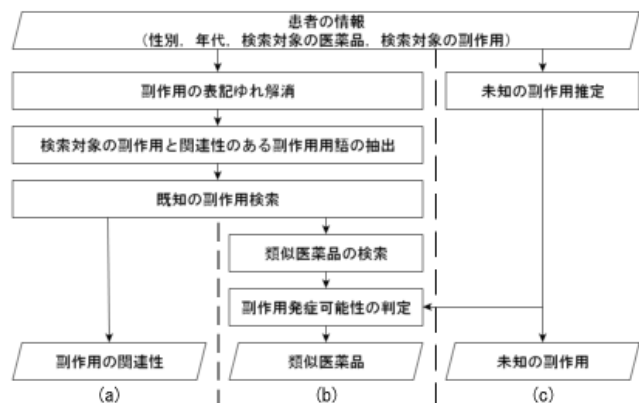


図 1 各副作用検索システムの処理手順

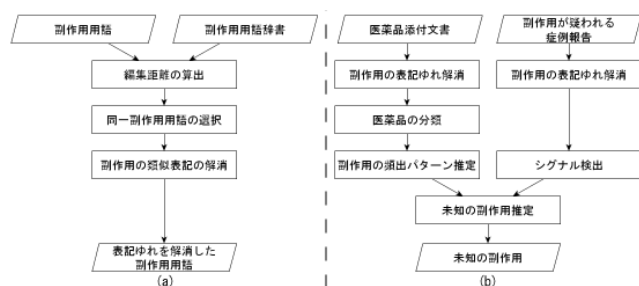


図 2 副作用の表記ゆれ解消手法(a)と未知の副作用推定手法(b)

3.1 使用データ

本手法では医薬品添付文書と副作用が疑われる症例報告を用いてシステムを構築する。

医薬品添付文書は、各医薬品の製造・販売元が作成を行う、医薬品の販売名、組成、効能、副作用等の医薬品の情報が記載された文書である。しかし、添付文書は記載項目や順序は定められているが、副作用の表記等は同義に解される類似表記が少なくないため、表記ゆれの問題がある。

副作用が疑われる症例報告は、医薬品医療機器総合機構²が公開しているデータを使用する。副作用が疑われる症例報告は、医薬品・医療機器等安全性情報報告制度に基づき医師や薬剤師などから報告された情報である。この症例報告には、医薬品によって副作用が発症したと疑われる患者の年代、身長体重、原疾患、投薬履歴、副作用と疑われる症状などが含まれる。

3.2 副作用の表記ゆれ解消

医薬品添付文書や、副作用が疑われる症例報告に記載されている副作用用語には同義に解される類似表記が少なくない。そのため副作用の表記ゆれが生じる。例えば、「アナフィラキシー」という副作用用語に対し、「アナフィラキシーショック」など、総称並びに付随症状に関する複数の表

記が存在する。また、「腸管閉塞」という用語に対し、「イレウス」という同一の疾患でありながら全く表記が異なる用語も存在する。表記ゆれの存在は、同一の副作用にも関わらず表記違いにより異なる副作用として認識してしまう一因となる。そこで、図 2(a)に示す手法で表記ゆれを解消する。なお副作用の表記ゆれ解消のため、くすりの副作用用語辞典³を基に、表 2 に示す副作用用語辞書を作成する。

表 2 作成する副作用用語辞書

副作用用語辞書	詳細
副作用基本用語辞書	基本となる副作用用語辞書及び、副作用の付随症状に関する辞書
副作用関連用語辞書	副作用基本用語辞書にある用語の類似表記用語の辞書
臨床検査関連辞書	臨床検査において生じる副作用用語及び、対応する類似表記用語の辞書
その他の副作用用語辞書	上記辞書に含まれない副作用用語

作成した副作用用語辞書に基づいて副作用の表記ゆれを解消する。副作用の表記ゆれ解消では、表記ゆれを解消したい副作用用語 $a = \langle a_1 a_2 \dots a_i \dots a_{|a|} \rangle$ と、副作用用語辞書中の任意の副作用用語 $b = \langle b_1 b_2 \dots b_j \dots b_{|b|} \rangle$ の、2 つの用語を比較することで表記ゆれを解消する。ここで a_i は副作用用語 a の i 番目の文字であり、 b_j は副作用用語 b の j 番目の文字である。また $|a|$ 、 $|b|$ はそれぞれ副作用用語 a 、 b の文字列長である。この副作用用語 a 、 b を、2 つの用語の相違具合を示す編集距離により比較する。編集距離 $d(a, b)$ は式(1)、式(2)、式(3)で定義される。編集距離 $d(a, b)$ の値が小さいほど用語 a 、 b は一致しており、 $d(a, b)$ の値が大きいくほど用語 a 、 b は相違である。

$$d(a, b) = \begin{cases} f(|a| - 1, |b| - 1) & (a_{|a|} = b_{|b|}) \\ 1 + \delta(a, b) & (a_{|a|} \neq b_{|b|}) \end{cases} \quad (1)$$

$$\delta(a, b) = \min(f(|a| - 1, |b| - 1), f(|a| - 1, |b|), f(|a|, |b| - 1)) \quad (2)$$

$$f(n, m) = \begin{cases} n & (m = 0) \\ m & (n = 0) \\ d(\langle a_1 a_2 \dots a_{n-1} a_n \rangle, \langle b_1 b_2 \dots b_{m-1} b_m \rangle) & (\text{otherwise}) \end{cases} \quad (3)$$

$(0 \leq n \leq |a|, 0 \leq m \leq |b|)$

しかし、用語の文字列長によっては編集距離 $d(a, b)$ と直感的な類似度には相違が生じてしまう。そこで式(4)により、2 つの文字列 a 、 b のうち、文字列長 $|a|$ 、 $|b|$ の値が大きい方で編集距離 $d(a, b)$ を割ることで、編集距離の値を 0 から 1 に正規化する。そして、正規化した編集距離が最小となった副作用用語辞書中の副作用用語 b を、表記ゆれを解消したい副作用用語 a と同一副作用用語 a' とみなす。

$$d'(a, b) = \begin{cases} \frac{d(a, b)}{|a|} & (|a| \geq |b|) \\ \frac{d(a, b)}{|b|} & (|a| < |b|) \end{cases} \quad (4)$$

2 独立行政法人 医薬品医療機器総合機構, <http://www.info.pmda.go.jp/>

3 くすりの適正使用協議会, くすりの副作用用語辞典, http://www.rad-ar.or.jp/03/02_fukusayo_jiten/03_fukusayoyogo.html

次に、編集距離により表記ゆれを解消した副作用用語 a' の類似表記を統一する。副作用関連語辞書と臨床検査関連辞書に含まれる副作用用語は、副作用基本用語辞書中の副作用用語と同義に解される類似表記の副作用用語である。そこで副作用用語 a' が、副作用関連語辞書または臨床検査関連辞書に含まれる副作用用語である場合、対応する副作用基本用語辞書中の副作用用語へと変換することで、類似表記を統一する。

3.3 未知の副作用推定

図 2(b)に示す未知の副作用推定手法には、副作用の頻出パターンを用いた未知の副作用推定と、Evans ら[1]のシグナル検出手法を用いる。副作用の頻出パターンを用いた未知の副作用推定では医薬品添付文書を、シグナル検出手法では副作用が疑われる症例報告を使用する。

(1) 副作用の頻出パターンを用いた未知の副作用推定

副作用の頻出パターンを用いた未知の副作用推定では、医薬品を日本標準商品分類⁴に基づき分類する。日本標準商品分類は、医薬品をその作用部位や成分を指標として分類したものである。ここで、日本標準商品分類によって分類した医薬品から、医薬品の副作用情報を抽出し、日本標準商品分類ごとの副作用情報集合を生成する。そして、生成した副作用情報集合に対し、飽和頻出パターンマイニングアルゴリズムの FP-close[7]を適用し、日本標準商品分類ごとの飽和頻出パターンを推定する。次に、推定した副作用の飽和頻出パターンに基づいて未知の副作用を推定する。推定した副作用の飽和頻出パターン集合 $C = \{C_i | 1 \leq i \leq s, s = |C|\}$ から、1 つの飽和頻出パターンの副作用集合 $C_i = \{x_j | 1 \leq j \leq t, t = |C_i|\} (i = 1, \dots, s)$ を選択する。また、未知の副作用を推定する医薬品の副作用集合を $D = \{y_k | 1 \leq k \leq u, u = |D|\}$ とする。そして、副作用集合 C_i 、 D について類似度 *similarity* を算出する。類似度 *similarity* は、式(5)と定義する

$$\text{similarity}(C_i, D) = \frac{|C_i \cap D|}{|C_i|} \quad (5)$$

類似度 *similarity* が高いほど、副作用集合 D 中に副作用集合 C_i が出現する可能性が高いと考えられる。そこで、類似度が閾値を超えた副作用集合 C_i の要素 x_j のうち、副作用集合 D の要素 y_k でない要素を未知の副作用として抽出する。そして類似度 *similarity* を、推定した副作用の飽和頻出パターン集合 C の全ての要素に対して算出する。

(2) シグナル検出手法を用いた未知の副作用推定

シグナル検出手法を用いた未知の副作用推定では、Evans ら[1]のシグナル検出手法 PRR を用いる。PRR は表 1 に基づき、式(6)のように定義される。ここで出力結果 PRR 及び各パラメータに対しての閾値として、 $PRR \geq 2, N_{11} \geq 3$ 、式(7)のカイ二乗値 $\chi^2 \geq 4$ と定義する。この閾値を満たす医

薬品と副作用の組のうち、検索対象の医薬品の既知の副作用でない副作用を、未知の副作用として抽出する。

$$PRR = \frac{N_{11}/N_{+1}}{N_{12}/N_{+2}} = \frac{N_{11} \times N_{+2}}{N_{12} \times N_{+1}} \quad (6)$$

$$\chi^2 = \frac{N_{11}(|N_{11}N_{22} - N_{12}N_{21}| - N_{++}/2)}{N_{1+}N_{2+}N_{+1}N_{+2}} \quad (7)$$

3.4 副作用検索システムの構築

3.2 項, 3.2 項の手法を用いて、以下の 3 つの副作用検索システムを構築する。

- 副作用関連性検索システム
- 未知の副作用検索システム
- 類似医薬品検索システム

なお、副作用検索システムの入力として、患者の性別、年代、検索対象の医薬品、検索対象の副作用を与える。

(1) 副作用関連性検索システム

図 1(a)に示す副作用関連性検索システムは、検索対象の医薬品の既知の副作用中に、検索対象の副作用と関連性のある副作用が含まれるかを検索するシステムである。

本システムでは、まず入力された検索対象の副作用の表記ゆれを解消する。次に副作用用語辞書を用い、表記ゆれを解消した副作用の検索を行う。検索対象の副作用に合致する副作用だけでなく、検索対象の副作用が、副作用基本用語辞書に含まれる副作用の付随症状である副作用も含め、副作用用語辞書から検索する。そして、検索で得られた副作用のうち、検索対象の医薬品における既知の副作用を検索して結果を出力する。

(2) 未知の副作用検索システム

図 1(c)に示す未知の副作用検索システムは、医薬品と副作用の因果関係は証明されていないが、医薬品と副作用の因果関係が疑われる副作用を検索するシステムである。

本システムでは、まず検索対象の医薬品について、副作用の頻出パターンを用いた未知の副作用推定を行う。次に、検索対象の医薬品について、シグナル検出手法を用いた未知の副作用推定を行う。なお、シグナル検出は以下の 3 つの場合ごとに未知の副作用を推定する。

- 性別や年代に関係なく現れる、未知の副作用
- 特定の性別のみ現れる、未知の副作用
- 特定の年代のみ現れる、未知の副作用

(3) 類似医薬品検索システム

図 1(b)に示す類似医薬品検索システムは、特定の副作用が発症しない同一効能の医薬品を検索するシステムである。

本システムでは、まず副作用関連性検索システムと同様に、検索対象の医薬品における既知の副作用中に、検索対象の副作用と関連性のある副作用が含まれるかを検索する。次に、検索対象の副作用に関連性のある副作用が、検索対象の医薬品の既知の副作用として含まれていない医薬品で、かつ検索対象の医薬品と日本標準商品分類が同一の医薬品を、類似医薬品として抽出する。さらに、抽出された類似

⁴ 総務省 統計局, 日本標準商品分類,
<http://www.stat.go.jp/index/seido/syouthin/>

医薬品の未知の副作用中に、検索対象の副作用に関連する副作用が含まれているかを検索し、検索対象の副作用が発症する可能性を類推する。検索対象の副作用に関連する副作用が、類似医薬品の未知の副作用中に含まれている場合、検索対象の副作用が発症する可能性「あり」と判定し、類似医薬品の未知の副作用中に含まれていない場合、検索対象の副作用が発症する可能性「なし」と判定する。

4. 副作用検索システム評価実験

本節では提案した副作用検索システムを用いて、実際の副作用が疑われる症例報告にある、医薬品との関連性が疑われる副作用を検索することで、本システムの有用性を評価する評価実験について述べる。

4.1 評価対象の副作用検索システム概要

3 節の提案手法に基づき構築した、評価対象である副作用検索システムを図 3 に示す。

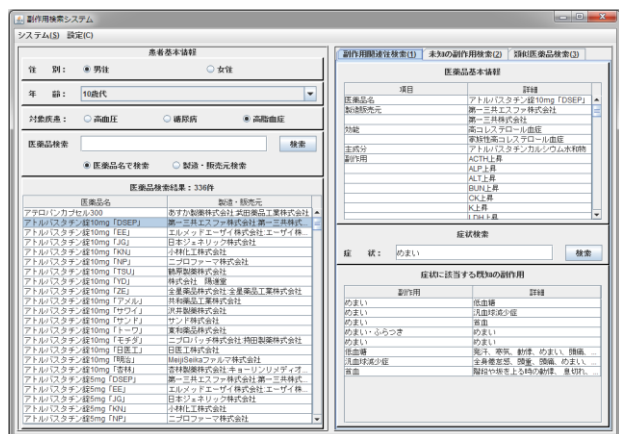


図 3 副作用検索システム

評価実験では、日本人の三大疾病である高血圧、糖尿病、高脂血症に対する効果をもつ医薬品を対象とした。評価実験で副作用検索システムの構築で用いた医薬品添付文書と副作用が疑われる症例報告は、医薬医療機器総合機構が公開するデータを用いた。また、評価実験で副作用検索システムの構築に用いた医薬品添付文書の概要を表 3 に、副作用が疑われる症例報告の概要を表 4 にそれぞれ示す。

表 3 評価実験に用いた医薬品添付文書の概要

医薬品の効能	添付文書数	医薬品数
高血圧	948 件	1,908 品目
糖尿病	440 件	821 品目
高脂血症	178 件	336 品目

表 4 評価実験に用いた副作用が疑われる症例報告の概要

症例報告期間	2003 年～2012 年 6 月
総患者数	229,799 人
医薬品の総レコード数	1,611,884 レコード
副作用の総レコード数	358,040 レコード
医薬品の総品目数	4,539 品目
副作用の総症例数	5,642 症例

4.2 評価方法

構築した副作用検索システムに対して、実際の副作用が疑われる症例報告を入力し、副作用の検出件数及び、処理時間を計測することでシステムの有用性を評価する。評価環境を表 5 に示す。

表 5 評価環境

OS	Windows 7
CPU	Intel® Core™ i7-950
メモリ	12.0GB

評価実験ではテストデータとして、2012 年 7 月から 2012 年 9 月までに医薬品と副作用の関連性が疑われると報告のあった、副作用が疑われる症例報告を用いる。そして、医薬品との関連性が疑われる副作用を、本システムで推定できるかを評価する。

テストデータから、効能が高血圧、糖尿病、高脂血症である医薬品が投薬された患者情報を、効能ごとに各 50 件抽出する。そして、抽出した患者情報のうち、患者の性別、年代、医薬品、医薬品との関連性が疑われる副作用を、副作用検索システムに入力した。

システムが出力した副作用が、入力したテストデータの医薬品との関連性が疑われる副作用と同一である場合、システムが副作用を検出できたとみなす。そして、以下のいずれの場合で副作用を検出したかを集計した。

例えば、システムに入力された副作用「アナフィラキシー」が、表記ゆれを解消せずに医薬品の既知の副作用として検出した場合、(ア)①の「入力された副作用に表記ゆれなし」として集計する。また、システムに入力された副作用「アナフィラキシーショック」の表記ゆれを解消し、「アナフィラキシー」として医薬品の既知の副作用と検出した場合には、(ア)②の「入力された副作用に表記ゆれあり」として集計した。

● 集計方法

(ア) 既知の副作用：

- ① 入力された副作用に表記ゆれなし
- ② 入力された副作用に表記ゆれあり

(イ) 未知の副作用：

- ① 副作用の頻出パターンによる推定
- ② シグナル検出による推定

(ウ) 検出なし

そして医薬品の効能別のテストデータ 50 件中、既知の副作用または未知の副作用として検出できた件数の割合を、副作用の検出率として算出した。また全テストデータ 150 件中の副作用の検出率を算出した。

また同時に、システムへのテストデータの入力からシステムが出力するまでの処理時間を測定した。

4.3 評価結果

副作用検索システムの医薬品効能別副作用検出結果を図

4 に、各副作用検索システムの医薬品効能別平均処理時間を図 5 に示す。

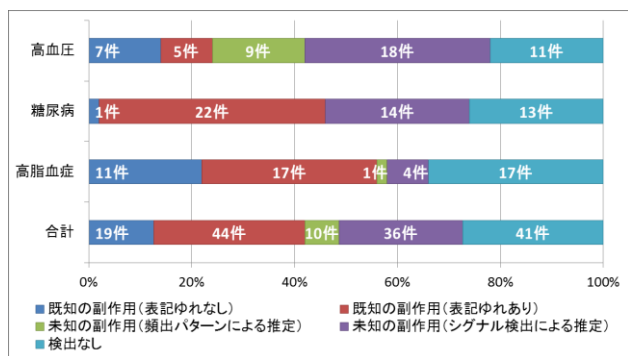


図 4 副作用検索システムの医薬品効能別副作用検出結果

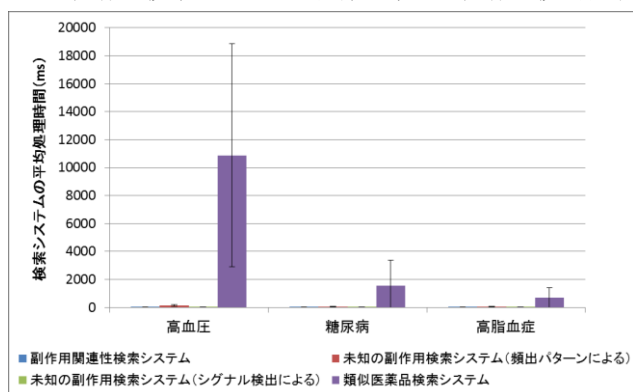


図 5 副作用検索システムの医薬品効能別平均処理時間

評価実験の結果、副作用検索システムの副作用検出率は、テストデータ全体で 72.7%であった。また医薬品の効能別では、高血圧の場合が最も高く 78.0%、糖尿病の場合が 74.0%、高脂血症の場合が 66.0%であった。

副作用の表記ゆれ解消の面では、テストデータ全体検出した副作用のうち、29.3%について副作用の表記ゆれを解消し、既知の副作用として検出した。したがって、従来は同一の副作用にも関わらず表記ゆれにより異なる副作用として誤検出されていた 29.3%の副作用を、本システムを用いることで副作用表記ゆれを解消し、同一の副作用として正しく検出できるようになったといえる。

また未知の副作用推定の面では、テストデータ全体で検出した副作用のうち、42.2%について未知の副作用として検出した。システムが未知の副作用として検出した副作用と、テストデータの医薬品との関連性が疑われる副作用が一致することから、システムが推定した未知の副作用は、医薬品の未知の副作用であると考えられる。なお、未知の副作用として検出した副作用中、シグナル検出により推定した未知の副作用が 78.3%、副作用の頻出パターンにより推定した未知の副作用が 21.7%であった。

副作用検索システムの処理時間の面では、副作用関連性検索システムや、シグナル検出による推定を用いた未知の副作用検索は、各医薬品に対しても平均処理時間は一定かつ短時間であった。一方、頻出パターンによる推定を用いた未知の副作用検索システムや、類似医薬品検索システム

は、平均処理時間ばらつきが生じた。これは、副作用頻出パターン集合 C の要素数に起因すると考えられる。副作用の頻出パターンを用いた未知の副作用推定では、副作用の頻出パターン集合 C の全要素について未知の副作用推定処理を行う。また類似医薬品検索では、未知の副作用推定を用いて副作用の発症可能性を推定する。したがって、未知の副作用検索システムや、類似医薬品検索システムの平均処理時間に影響が出たと考えられる。

これらの結果から、提案したシステムは副作用の検索において有用であると考えられる。しかし、副作用の頻出パターンによる未知の副作用推定手法は、精度向上及び処理時間の向上が必要である。

5. まとめと今後の課題

本稿では、限られた時間内での医薬品の副作用の検索、未知の副作用、副作用の表記ゆれという問題点に着目し、副作用の表記ゆれに頑健かつ、医薬品の未知の副作用の検索に対応した医薬品の副作用検索システムを提案した。具体的には、副作用の表記ゆれの統一手法や、医薬品の副作用頻出パターンを用いた未知の副作用推定手法を提案し、副作用検索システムを構築して評価実験を行った。

今後の課題としては、検索システムの処理速度の向上や副作用検出率の向上、検索対象とする医薬品の全医薬品への拡張、医薬品の飲み合わせについての考慮が挙げられる。また未知の副作用推定手法で、未知の副作用をランキングすることで、未知の副作用推定精度の向上することが挙げられる。

参考文献

- 1) Evans, S.J.W., Waller, P.C. and Davis, S.: "Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports," *Pharmacoepidemiology and Drug Safety*, Vol.10, No.6, pp.483-486 (2001).
- 2) Rothman, K.J., Lanes, S. and Sacks, S.T.: "The reporting odds ratio and its advantages over the proportional reporting ratio," *Pharmacoepidemiology and Drug Safety*, Vol.13, No.8, pp.519-523 (2004).
- 3) Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A. and De Freitas, R.M.: "A Bayesian neural network method for adverse drug reaction signal generation," *Eur J Clin Pharmacol*, Vol.54, No.4, pp.315-321 (1998).
- 4) DuMouchel, W.: "Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System," *The American Statistician*, Vol. 53, No. 3, pp.177-190 (1990).
- 5) Harpaz, R., Haerian, K., Chase, H.S. and Friedman, C.: "Mining electronic health records for adverse drug effects using regression based methods," *HIH'10*, pp.100-107 (2010).
- 6) 平川聖太郎, 小野純平, 渡辺まり子: "HTML5 LocalStorage の転置 index を用いた類似医薬品検索システム," 電子情報通信学会総合大会講演論文集, Vol.2012, No.1, p.146 (2012).
- 7) Grahne, G. and Zhu, J.: "Efficiently using prefix-trees in mining frequent itemsets," *FIMI* (2003).