

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 4日現在

機関番号：17701

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21590571

研究課題名（和文）：医薬品副作用シグナル検知システムへのテキストマイニング技術の
応用に関する研究研究課題名（英文）：A study on application of the text mining technology to the adverse
drug reaction signal detection system

研究代表者：村永 文学 (MURANAGA FUMINORI)

鹿児島大学・医学部・歯学部附属病院・講師

研究者番号：00325812

研究成果の概要（和文）：本研究では、総合病院情報システムの情報から、薬剤相互作用による白血球減少症をアソシエーション分析アルゴリズムで発見する方法の技術的な検討を行った。対象は2008年～2009年に当院に入院し前立腺癌の化学療法を受けた患者を対象とした。2009年の症例について発見したIF-THENルールのうち、2008年の症例から作成した知識データベースに含まれないものが35万レコードであった。lift値の大きい薬剤について調査したところ、多くが既知の薬剤であった。発生頻度が低くクリティカルな有害事象については、十分に整備された知識辞書が必須であった。

研究成果の概要（英文）：This study was performed to evaluate the utility of a data mining algorithm for detection of adverse drug events. We used experimental data (medication history and test results) with the assumption of "an adverse event due to an interaction between an existing drug and a new drug after introduction of the new drug on to the market". We then evaluated whether we could identify this assumed drug-drug interaction. We developed a knowledge database from association rules extracted from medication history and test results before addition of the new drug. This rule was used for filtering the knowledge database in analysis of data including the new drug. We tried to perform an association analysis using actual prescription data, test results and medical records. We developed a knowledge database by patients with prostate cancer which left the hospital in 2008, and conducted an investigation into the drug adverse event of the prostatic patients who left the hospital in 2009. As a result, it was recognized that the maintenance of the knowledge dictionary was very important.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	2,500,000	750,000	3,250,000
2010年度	500,000	150,000	650,000
2011年度	600,000	180,000	780,000
年度			
総計	3,600,000	1,080,000	4,680,000

研究分野：医歯薬学

科研費の分科・細目：境界医学・応用薬理学

キーワード：薬剤疫学, data warehouse, data mining, 医薬品相互作用, 医薬品副作用シグナル検知

1. 研究開始当初の背景

医薬品が市販される前の臨床試験（フェーズ3）は、モデル的な患者集団を対象とし、いわば「理想の世界」で実施されるが、市販後の薬は「現実の世界、医療の現場」で使われる。例えば、キノホルムによるスモンや、phenyl propanol amine による出血性脳卒中等、従来の臨床試験法では予測不可能な有害事象や相互作用が発生することは過去にも多数見られた。特に小児、妊婦、高齢者に対する影響、および既に市販されている薬との相互作用は、市販後調査 PMS(Post Marketing Surveillance:フェーズ4)によってのみ明らかになる事象といえ、最近認可された新薬の殆どが、処方された全症例のデータの提出が認可の条件となっている点からも、その重要性はうかがえる。しかし、薬剤相互作用による有害事象の発見が遅れることは多くの犠牲を払うことに繋がる。フェーズ4における薬剤の有害事象の早期検知は喫緊の課題と言える。近年、PMS の手法として、自発報告の中から優先的に検討すべき重要な adverse drug reaction (医薬品副作用。以下 ADR) の疑いを、統計的に選り出すシグナル検出法の研究が進んでいる。

一方、医療機関等は組織的に安全性の監視を行うことが不可欠である。これからの時代は、薬剤師の知識と経験のみに頼るのではなく、処方データおよび検査結果値等から、データマイニングと呼ばれる統計的なシグナル検出法を用いてシステムティックに薬剤監査支援を行うニーズも生まれてくると考えている。

2. 研究の目的

薬剤疫学データウェアハウスの医薬品相互作用シグナル検知システムの精度を向上させ、既知の副作用や薬剤併用禁忌等の監視のみならず、未知の医薬品相互作用のシグナル検知手法を確立することを目的とし、研究を行っている。

特に本研究では、総合病院情報システムに蓄積されたデータを対象に、医薬品相互作用による有害事象のシグナルを、アソシエーション分析アルゴリズムの apriori と知識辞書を組み合わせた検出アルゴリズムによって検知可能か評価するのが目的である。

3. 研究の方法

(1) 対象

対象は、2008年～2009年に当院に入院し前立腺癌の化学療法を受けた患者を対象とした。具体的には、まず薬剤疫学DWHサーバに、2008年～2009年に発生した薬歴と検

査結果及びカルテ記事を連結不可能匿名化の処理したのち抽出し登録した。

(2) データクレンジング処理

薬歴のデータクレンジング処理としては、消費メモリおよび処理速度を高速化するために薬剤名称を「薬剤001」等のように単純化したものに置き換えた。(以下、薬剤略称と呼ぶ) 同一成分で異なる規格の薬剤は同じ薬剤略称とした。アソシエーション分析アルゴリズムの apriori では量を扱えないので、最小規格のものを1とし、使用量に応じて、薬剤略称の個数を増やすことで対応した。例えば、「薬剤001」の1錠1mgを最小単位1とし、「薬剤001」の2mg規格を1錠投与している場合は、「薬剤001」を2つ登録した。整数にする必要があるため、端数は四捨五入した。

(3) 診断の推定

今回は、IF-THENのアソシエーション対象として、白血球減少症の発症を想定した。情報源としては、検査結果値と、カルテの記事のテキストマイニング処理結果で判断するようにした。

検査結果のクレンジング処理としては、単純に閾値を設けるのではなく、白血球数の過去データとの変化率が50%を超えたものを有意(白血球減少あり)とした。

カルテ記事のテキストマイニング処理については、記事の形態素分析を行い、独自の係受け判別処理を行った。形態素分析については、MeCabシステム(ver. 0.98)を利用した。係受け処理は、2008年のカルテ記事情報を形態素分析し、主キーワードとなる「白血球」「WBC」「顆粒球」から係受けた語に「減少」「低下」「↓」等の係受けキーワードが出現した場合に有意とした。

(例:「白血球」→「減少」、「顆粒球」→「減少」、「WBC」→「1500」→「↓」、 「WBC」→「1500」と→「低下が」→「見られた」等)

係受け解析システムは、汎用のものについて評価したが、予備試験を行った結果、カルテ独特の表記へ対応が困難であったため、今回は白血球減少の発見に特化した係受け処理を行うperlスクリプトを独自開発した。

検査結果もしくはカルテの記事のどちらかで有意であった場合に、白血球減少症を発症した症例とした。

(4) データベースへの格納

各個人のデータについて、時間軸を1週間ずつに区切り、同時に投与した薬剤と、その結果として白血球減少症が発症の有無をデータベースに蓄積した。

例1:患者1の2008年第1週の薬歴は「薬剤001, 薬剤001, 薬剤002, …」、結果は「WBC_N」

例 2 : 患者 1 の 2008 年第 2 週の薬歴は「薬剤 001, 薬剤 001, 薬剤 002, 薬剤 003 . . .」、結果は「WBC_L」とした。

次に、蓄積したデータを元に 2008 年のデータに対してアソシエーション分析を実施し、抽出されたルールで知識データベースを作成した。

次に、2009 年のデータに対してアソシエーション分析を実施し、ルールを抽出した。

2008 年のデータで作成した知識データベースと、2009 年にデータで発見したルールを比較し、知識データベースに存在しないルールを異常（発見した新たな知見）として抽出した。

アプリアリ (Apriori) は 1994 年に IBM アルマデン研究所の R. Agrawal によって提案された相関ルール抽出アルゴリズムである。このアルゴリズムでは、発生イベントを {0, 1} の 2 値として扱っている。現在、このアルゴリズムを改良した様々なアルゴリズムが提案されており、パッケージとして実用化もされている。

以下に、アプリアリアルゴリズムによる具体的な処理手順を示す。

手順 1) アイテム 1 つずつを候補アイテム集合 $C(1)$ (ルールとして抽出される候補) と呼び、全データベースを検索して各候補アイテム集合 $C(1)$ の出現回数をカウントし支持度を計算する。

手順 2) 各候補アイテム集合 $C(1)$ について、ユーザの定めた基準である最小支持度以上の支持度を持つものをラージアイテム集合 $L(1)$ と呼ぶ。

手順 3) ラージアイテム集合 $L(1)$ 同士を組み合わせたものを新しく候補アイテム集合 $C(2)$ として出現回数をカウントし支持度を計算する。

手順 4) 手順 2 と手順 3 の処理を k 回繰り返し、候補アイテム集合 $C(k)$ が空集合になったところでストップする。各パスにおいてのラージアイテム集合を出力とする。なお、最小支持度は、予備実験の結果より決定する。

使用したソフトウェアについて説明する。薬剤疫学 DWH 用 DBMS としては、MySQL ver 5 を利用し、薬歴、検査結果値、症例基本情報等の格納と知識データベース用として利用した。

データクレンジング&知識ベースフィルタは ActivePerl ver. 5.8 系を利用して自作した。

アソシエーション分析・頻出アイテム抽出処理については、過去の研究では R ver. 2.9.2 の `arules` パッケージを利用したが、本研究では perl スクリプトでアプリアリアルゴリズムを実装した。その理由は、R にデータを

渡し、結果を受け取るインターフェースがテキストファイル (バッチファイル) 経由であり、バッチファイル実行時の利用可能メモリの制約上、大量のデータを perl スクリプトから R に渡し分析することが困難であったためである。なお、過去の R で処理した実験系のデータと、新規に作成した perl スクリプトによるアプリアリアルゴリズム処理ライブラリで同じ結果がでることは確認できた。

(5) 実験結果

2008 年に退院した患者の前立腺癌患者数は、手術症例も含め 132 例あり、そのうち、白血球減少症を発症していた患者は 40 例 (29.6%) であった。投与薬剤種数は、注射薬 142 種、内服・外用薬 290 種であった。

検査結果によって白血球減少症と判断した例が 40 例であった。カルテ記事のテキストマイニング処理にて白血球減少症を発見できたものは 18 例であり、全て検査結果によって判断した例に包含された。

2009 年に退院した患者の前立腺癌患者数は、手術症例も含め 198 例であり、そのうち白血球減少症を発症していた患者は 51 例 (25.6%) であった。投与薬剤種数は、注射薬 148 種、内服・外用薬 337 種であった。

検査結果によって白血球減少症と判断した例が 51 例であった。カルテ記事のテキストマイニング処理にて白血球減少症を発見できたものは 22 例であり、その中で、記事上の表現にて誤検知したと思われる例が 2 例あった。(投与後の顆粒球減少に注意するよう指示したという記載を検知) 誤検知例については、処理の対象から外した。

2008 年退院と 2009 年退院の患者 ID には 75 例の重複があった。手術及び化学療法の繰り返しによる再入院によるものと思われた。

処方された薬剤について、2008 年と 2009 年で重複しているものについて調査した。内服・外用薬では、重複しているものは 223 種、2008 年だけの処方薬は 67 種、2009 年だけの処方薬は 114 種であった。注射薬剤では、一致しているものが 94 種、2008 年だけの処方薬が 48 種、2009 年だけの処方薬が 54 種であった。

IF-THEN ルール抽出において利用したアプリアリアルゴリズムの閾値パラメータは、前回の研究の結果より、support 値 = 0.05, confidence 値 = 0.5 とした。

2008 年の症例データをアプリアリで処理後、抽出した白血球減少症関連薬剤の IF-THEN ルール数は、2,798,373 ルールであった。これを知識データベースに格納した。

2009 年の症例データをアプリアリで処理後、抽出した白血球減少症関連薬剤の IF-THEN ルール数は、2,973,607 ルールであり、知識データベースに登録の無い IF-THEN ルール数は 350,468 となった。

4. 研究成果

我々は過去の研究で、実験的に「新薬が追加された後、既存のある医薬品と相互作用による有害事象が発生した」ことを想定した薬歴と検査結果値のデータ群を作成し、この想定された相互作用薬を発見可能かどうか評価を行った。その結果、相互作用の発症率を25%程度に下げてもパラメータの調整を行うことで想定する相互作用薬を発見可能であることが判明した。

本研究では、総合病院情報システムに蓄積された薬歴と検査結果、カルテの記事から、薬剤相互作用による白血球減少症の発症について相互作用薬を実験系と同様の手順で発見可能か、どのような技術的な問題があるかを検証した。

(1) テキストマイニングによる白血球減少症の判定について

今回の事例では、汎用の形態素分析システムである MeCab ver 0.98 と、独自の係受け処理スクリプトを用いて白血球減少の判断を行うようにした。

係受け処理スクリプトは2008年の症例のカルテ記事の形態素分析結果を元に、近接する品詞の係受けルールを見出し辞書化した。

(ルールの策定を行った)したがって2008年分のデータ処理については当然ながら良好な結果を得た。しかしテキストマイニング手法による白血球減少症の検知率は低い結果であり、その原因として記事の未記載や明確な表現の無いものが多かった。

2009年のカルテ記事に対する分析を同じルールで行った場合に、明らかに誤検知と思われる事例があり、この例では「将来発生する顆粒球減少に注意するように」と記載された記事であった。テキストマイニングでは、対応の難しい記事であると思われた。

これらの結果より、白血球減少症については、検査結果値による判断の方が高い精度で判断できた。もし検査結果等により判断が困難な症状についてはテキストマイニングに頼らざるを得ないが、現時点では実用的な検知率に至る方法論が確立したとは言えない。さらなる研究が必要であると考えられる。

(2) 相互作用薬の発見について

実験の結果、2009年の症例について発見したIF-THENルールのうち、2008年の症例から得た知識データベースに含まれないものが35万レコードにも及んだ。これは、2008年の症例と2009年の症例に処方された薬剤種の差が大きいことに起因する。発見されたIF-THENルールを算出されたlift値の大きい順に並び替え、ある程度白血球減少症に関連のある薬剤候補としてあげることが可能で

あったが、多くが既に白血球減少症を単独で発生させることが添付文書に明記されている薬剤と、たまたまその患者に処方されていた薬剤群であり、新規の医薬品相互作用の発見には至らなかった。

今回採用したアプリアリアルゴリズムは、閾値パラメータとして support 値および confidence 値を指定可能であり、この値を高くすることで発見するIF-THENルール数を減らすことは可能である。ただし、我々の過去の研究1) 2)において、アプリアリアルゴリズムを有害事象データを含むデータに適用する場合に、有害事象の件数が少ない場合、最小支持度50%、最小確信度80%ではアプリアリアルゴリズムではディテクトできなかった。これはアプリアリアルゴリズムがメジャーなデータルールを抽出するためのアルゴリズムであるからである。したがって、発見するIF-THENルール数を減らすために、無為に support 及び confidence を高く設定すると、件数の少ない有害事象については見落としが発生する恐れがあると考えられる。

以上より、発生頻度が低く、クリティカルな有害事象については、十分に整備された知識辞書が必須であり、その整備方法については、十分に吟味する必要があると考えた。例えば、白血球減少症を来たさなかつた症例、及び既知の副作用による白血球減少症と確定できるような症例を選別し、それを知識辞書のソースとして利用するのも1つの方法かもしれない。

(3) まとめ

アプリアリアルゴリズムによるアソシエーション分析と知識データベースフィルタリングを組み合わせた医薬品相互作用発見システムの臨床応用として、前立腺癌患者の白血球減少症に影響する医薬品相互作用による有害事象の発見の試行を行った。本研究では知識データベースとして用いた2008年の投与薬剤種数と、マイニング対象とした2009年の投与薬剤種数の差が大きく、満足できる結果は得られなかった。加えてテキストマイニングによる白血球減少の判定についても、満足できる検知率には至っていない。

今後の課題としては、知識データベースを整備する方法と、テキストマイニングによる検知率の向上について、さらなる研究を進めていきたい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- 1) 村永文学, 岩穴口孝, 宇都由美子, 熊本一朗:アソシエーション分析による医薬品相互作用検知手法の評価, 医療情報学 31 巻 Suppl. 2011; p1000-1002 査読(有)
- 2) Fuminori Muranaga, Yumiko Uto, Ichiro Kumamoto.:Evaluation of a Data Mining Technique for Detection of Adverse Drug Events using a Pharmacoepidemiological Data Warehouse. Journal of the 16th JSPE and the 5th ACPE Joint Meeting , 2010; Vol.16: p54. 査読(有)
- 3) 鳥越千秋, 向窪世知子, 村永文学, 宇都由美子, 窪田美行.:形態素分析と辞書フィルターを活用した看護記録の監査方法に関する研究. 日本医療マネジメント学会雑誌,2010;11 巻 Suppl. :p263. 査読(有)
- 4) 村永文学, 宇都由美子, 熊本一朗.:薬剤疫学データウェアハウスを用いた医薬品有害事象検知手法の評価. 医療情報学,2009;29 巻 Suppl.:p772-773. 査読(有)

[学会発表] (計 1 件)

- 1) 村永文学:アソシエーション分析による医薬品相互作用検知手法の評価: 第 31 回医療情報学連合大会, 2011 年 11 月 22 日, 鹿児島

6. 研究組織

(1) 研究代表者

村永文学 (MURANAGA FUMINORI)
鹿児島大学・医学部・歯学部附属病院・
講師
研究者番号 : 00325812

(2) 研究分担者

熊本一朗 (KUMAMOTO ICHIRO)
鹿児島大学・大学院医歯学総合研究科・
教授
研究者番号 : 40225230

宇都 由美子 (UTO YUMIKO)
鹿児島大学・大学院医歯学総合研究科・
准教授
研究者番号 : 50223582