

令和 2 年 6 月 11 日現在

機関番号：17701

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K09238

研究課題名(和文)ディープラーニング応用の医薬品相互作用シグナル検知システムの開発評価に関する研究

研究課題名(英文) Study of development and evaluation of signal detection system for side effects due to drug interaction using deep learning

研究代表者

熊本 一郎 (Kumamoto, Ichiro)

鹿児島大学・医歯学総合研究科・客員研究員

研究者番号：40225230

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究では医薬品相互作用による有害事象のシグナル検知用の分析手法2つの評価研究を行った。医療従事者の自由記載記録を文章ベクトル化手法であるGenSimライブラリのDoc2Vecで分析する手法では、誤嚥性肺炎を発症した患者の看護経過記録をDB化し、コサイン類似度の閾値0.999613で判定し、AUC 0.763、感度は90.9%、特異度60.3%で判定できた。時系列の量的データである検査結果を画像化し、Google社のtensorflowを利用して機械学習させ分析する手法では、有害事象発症群として機械学習用に各500例、評価用に各10例ずつ画像化検査データ作成。疾患判別率は68-72%であった。

研究成果の学術的意義や社会的意義

我々は過去の研究において、医薬品の相互作用による有害事象のシグナルを検知するシステムを開発し、ニューラルネットワークが膨大な薬剤種であっても分析可能であり実用的であることを見出した。相互作用を起こしている可能性の高い薬剤はアプリアリアルゴリズムを応用して特定できた。本研究の成果によって、医療従事者の自由記載記録と定性・定量化された大量の検査結果から、多種多様な症状を呈する有害事象の発症の予見が可能となり、新薬の市販後調査(フェーズ4)時の有害事象の自動監視により安全な創薬を支援するであろう。また本研究の成果は、各種疾病の発症予測やインシデントの発生予防にも応用可能であると思われる。

研究成果の概要(英文)：In this study, we evaluated two analytical methods for signal detection of adverse events due to drug interactions. It is a method to analyze free description records written by medical staff with Doc2Vec of GenSim library which is a text vectorization method. A nursing progress record of a patient who developed aspiration pneumonia was made into a DB and judged by a cosine similarity threshold of 0.999613. The AUC was 0.718, the sensitivity was 90.9%, and the specificity was 60.3%.

This is a method to image inspection results, which are time-series quantitative data, and perform machine learning analysis using Google's tensorflow. As adverse event onset groups, 500 cases for machine learning and 10 cases for evaluation were prepared and analyzed. The disease discrimination rate was 68-72%.

研究分野：医療情報学

キーワード：機械学習 tensorflow doc2vec 医薬品有害事象

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

薬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」で行われるが、市販後の薬は「現実の世界、医療の現場」で使用される。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)によって、初めて明らかになる事象といえる。

従来より 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、統計的に選び出すシグナル検出法の研究が続けられている。しかし、医療従事者の気付きに頼って有害事象の報告をしているのが現状であり、このままでは有害事象の気付きが遅れ、薬害の被害者を多く発生させる恐れがある。これからの時代は、薬剤師や医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、コンピュータによる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のが主流となるだろう。

一方、近年、機械学習分野における研究の発展が著しい。人工知能にビッグデータから特徴量を学習させ、それを基にして未知データを分類させる手法である。この機械学習手法にディープラーニングはブレイクスルーをもたらした。ディープラーニングでは、自己符号化機(ニューラルネットワーク層)を多層化することで、分類させたいデータについて、どのようなパラメータで学習すれば精度が良くなるのかコンピュータ自身が最適化することを可能とした。

我々は、過去の科学研究費補助金により医薬品相互作用による有害事象のシグナルを検知する為の薬剤疫学データウェアハウスシステムを開発してきた。平成20~23年度の科学研究費補助金の支援を受け、実験的に、薬歴・検査結果等から有害事象の発生を検知し、相互作用を及ぼした可能性が高い医薬品の候補をピックアップすることに成功した。平成24~26年度の科学研究費補助金の支援を受け実施した研究では、唯一、単層ニューラルネットワークだけが、アプリオリアルゴリズムやベイジアンネットワークよりも、より実際の薬歴に近いデータ構造から想定した相互作用薬を発見できることを証明した。

2. 研究の目的

本研究の最終目標は、現場の医療従事者等に特別な入力作業等の負担を発生させず、臨床上、自然に発生するデータから、未知の医薬品相互作用による有害事象を自動的に発見し警告するシステムを開発することである。本研究期間内に、我々は以前より研究開発を続けてきた医薬品の相互作用による有害事象のシグナル検知機能の精度を向上させる為、多層ニューラルネットワーク(ディープラーニング)の応用が可能であるか、研究を行った。

3. 研究の方法

本研究では、医薬品の相互作用によって発生した有害事象のシグナルを機械的に検知するために必要となる、2つの分析手法について評価研究を行った。どちらの研究でも下記のハードウェア及びソフトウェアを利用した。

- ・利用したハードウェア：Nvidia GTX980 SLI
- ・利用したソフトウェア：
 - Python ver. 3.8.2
 - GenSim Library ver. 3.8.3
 - Nvidia 社製 CUDA ver. 10.1 update2
 - CuDDN (CUDA Deep Neural Network Library) ver. 7.6
 - Google 社製 tensorflow ver.2.2.0
 - Perl ver.5.8 (量的・質的・時系列データの画像化 自作スクリプト)

(1) 医療従事者の自由記載記録からシグナルを検知する方法に関する研究

まず、電子カルテに医療従事者が入力した自由記載文からシグナルを検知する方法として、文章ベクトル化手法である GenSim ライブラリの Doc2Vec を利用して分析する方法について評価した。2013年から2017年の間に鹿児島大学病院に入院したすべての患者から、入院中に誤嚥性肺炎を発症した患者を誤嚥性肺炎発症群とし、機械学習用データと、評価用データに分割した。本研究での誤嚥性肺炎症例群の定義は、入院期間中に誤嚥性肺炎の病名開始日を有し、かつ抗生剤の投与が開始された症例とした。加えて医師がカルテ記録及び薬歴・検査結果を詳細に調査しカルテ記事からも誤嚥性肺炎の発症を確認した。また対照症例群としては、2018年6月1日~7月20日の間に入院した患者を抽出し、年齢構成として誤嚥性肺炎症例群の95%信頼区間に相当する年齢に該当した症例をランダムに抽出し、対照症例群とした。

解析対象とした看護観察記録は、1記録あたり40文字(80Byte)の制限が設けられ、簡素な表現で患者の状態を逐次記録するように設計されている。看護師の観察記録は各症例により記録量が異なっていたため、誤嚥性肺炎発症前の4日分の記録のみを使用した。ランダム抽出された

対照症例群については、対象期間中の入院日から記録日数の調整を行った。日本語では、Doc2Vecでの解析の前に事前処理が必要である。日本語文章は単語が繋がっているため、事前に形態素解析を行い単語間にスペースを設けておく必要がある。今回は、MeCabを用いて形態素解析を実施した。なお、形態素解析用辞書として、MeCabに付属されている ipadic 辞書、Comejisyo15) (1文字単語は削除)に加え、鹿児島大学病院独自のユーザー辞書を使用した。ユーザー辞書の作成には、誤嚥性肺炎症例群・対照症例群のデータとは関係のない過去の看護観察記録を用いた。MeCabで形態素解析後に、適切に単語が分割されない用語(固有名詞等)を抽出し、ユーザー辞書として登録した。抽出された誤嚥性肺炎症例群を機械学習用データと評価用データにランダムに2分割した。評価用データには、非肺炎の対象データも混入した。

評価方法を図1に示す。機械学習用の症例データに、評価用データを1件混入し、Doc2Vecにて各学習用データとのコサイン類似度を算出し、平均値を算出した。つまり、過去の誤嚥性肺炎症例の記録との類似性の平均値を求めることになる。

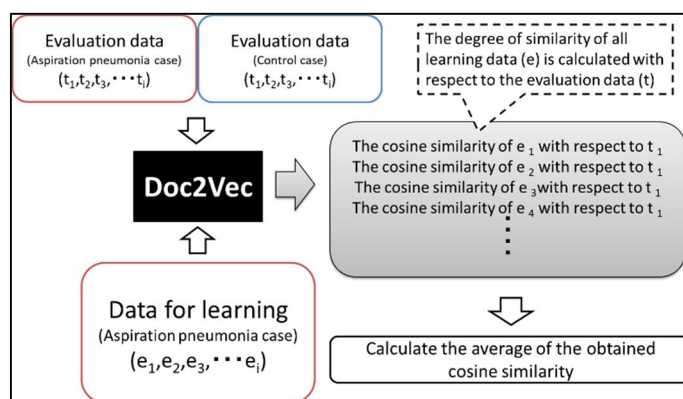


図1. Doc2Vecによる分析手順概要

(2) 時系列の量的・質的データからシグナルを検知する方法に関する研究

時系列かつ量的データである検査結果データと、定量化された症候を画像化し、多層化ニューラルネットワークである Google の tensorflow を用いて疾患を分類する方法について述べる。

画像の機械学習については学習用データ数が少なくとも数百～数万件必要であると言われていた。しかし多くの学習用データを臨床データから抽出するのは不可能であるので、本研究では臨床データをベースとし、データが正規分布するようにシミュレータで症例データを増加させた。具体的には、日本臨床検査医学会が2018に示した共用基準範囲から box-muller 法を用いて正規分布する健常者データを作成して対照群を1万件作成した。症例データについては、2013年から2017年の間に鹿児島大学病院に入院したすべての患者を対象とし、肝機能障害(22例)、腎機能障害(58例)、顆粒球減少症(72例)を抽出し、検査結果値の代表値と分散の変化について調査した。その結果、肝機能障害患者は $AST=+20 \pm 22$, $ALT=+30 \pm 34$, $LD=+80 \pm 127$ であった。腎機能障害患者は、尿素窒素 $BUN=+14 \pm 12$, $Cr=+1.2 \pm 0.9$, $UA=+4 \pm 6$ であった。顆粒球減少症は $WBC=-2502 \pm 921$ であった。これらの過去症例の特徴量を参考に、共用基準範囲に対して、上記代表値の変化が発生するように、box-muller 法を用いて正規分布する症例群を、機械学習用に各500例、評価用に各10例ずつ perl スクリプトで作成した。

量的・質的・時系列データの画像化手法を図2に示す。横軸を時間軸とし、縦軸を項目軸とし、1ピクセルを1種類の検査結果項目の1回分のデータとした。1ピクセルあたり、RGB毎に256段階の値を表現可能である。本研究では、基準範囲の中央値を黒色とし、正の値を赤色(基準範囲の上限が128)とし、負の値を青色(基準範囲の下限が128)となるように調整し、画像化した。

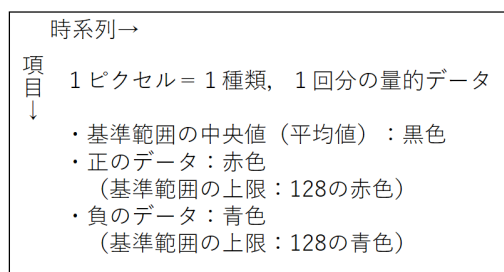


図2. 量的・質的・時系列データの画像化

作成した画像は、微細な変化として表現されており、人間の目では全く区別がつかないと思われた。作成した健常者データ及び各症例データについて、tensorflow で機械学習し、評価用データの疾患を判別可能か評価した。

4. 研究成果

(1) 医療従事者の自由記載記録からシグナルを検知する方法に関する研究

Doc2Vec には、調整パラメータが複数存在する。その中でも重要なものはベクトル size、window、min_count である (図 3)。

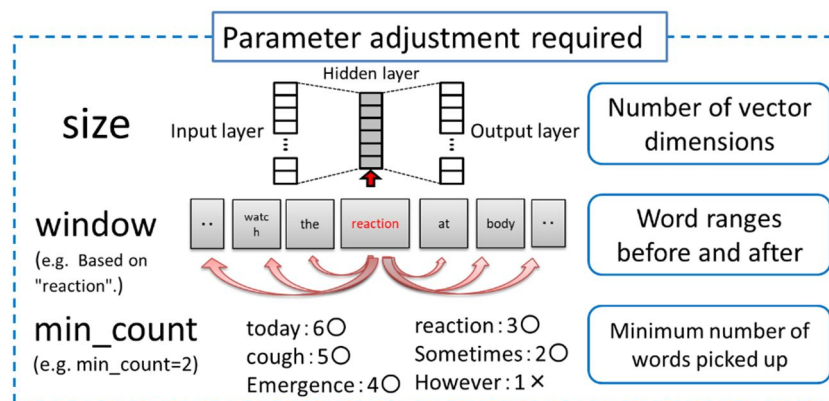


図 3. Doc2Vec の調整パラメータ

まずは最適な調整パラメータの算出実験を行った。各パラメータの最適値探索範囲を表 1 に示す。

size	10,20,30,40,50,60,70,80,90,100
window	4,5,6,7,8,9,10,11,12,13,14,15
min_count	1,2,3

表 1 各パラメーターの調整値

最適パラメータ算出用データで算出された各パラメータの AP(Average Precision)の上位 10 位を表 2 に示す。

size	window	min_count	AP
40	12	1	0.436205426
60	6	1	0.433256461
40	8	1	0.430546809
80	12	1	0.428175516
60	13	1	0.422276537
60	14	1	0.418351127
40	11	1	0.416334529
40	12	2	0.414355221
100	12	1	0.414306112
70	15	1	0.408291844

表 2 AP による最適パラメーター値の上位 10 位

これらの実験の結果、size=40, window=12, min_count=1 が最適なパラメータ値であった。最適パラメータ時で評価用データを分析した際の ROC 曲線(Receiver Operating Characteristic Curve) を図 3 に示す。

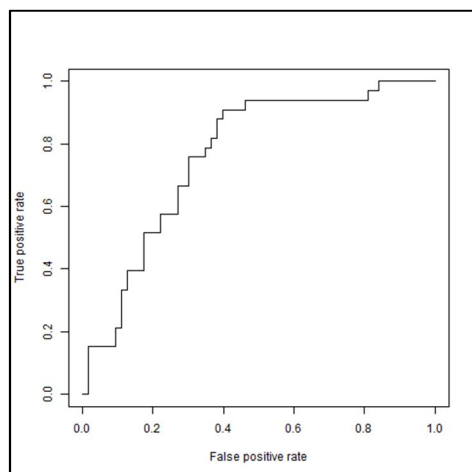


図 3 . 最適パラメータ値における ROC 曲線

AUC(Area Under Curve)は 0.763 であった。最大となる感度は 90.9%、特異度は 60.3%で、誤嚥性肺炎の発症を予測することが可能であった。

(2) 時系列の量的・質的データからシグナルを検知する方法に関する研究

研究の結果を図 4 に示す。判別率は 68-72%であった。5 回の学習回数にて判別率の上昇は頭打ちになった。10 回を超える学習では損失が上昇し、過学習となっていた。

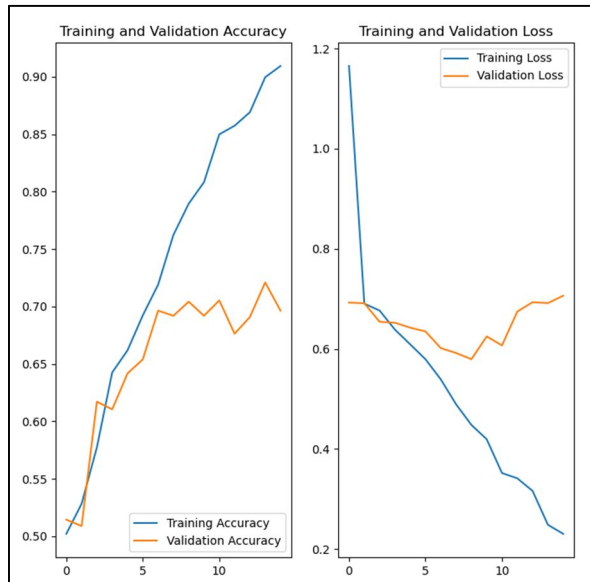


図 4. tensorflow による疾患画像判別。機械学習回数と精度（左）損失（右）の変化

これら 2 つの研究の結果より、医療従事者の自由記載記録と、量的・質的・時系列データから、過去の症例との類似性から有害事象のシグナル検知が可能になったと言える。有害事象シグナル検知後は、各症例の薬歴に対して、ニューラルネットワークまたはアプリアリアルgorithmを用いて分析し、相互作用を発生している可能性のある薬剤を特定可能であることは過去の実験より可能であると考えられる。

今回の研修では、tensorflow による画像化された検査結果データの判別率は約 70%にとどまった。今後も更なる精度の向上を目指し研究を継続したい。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 小牧祥太郎, 村永文学, 宇都由美子, 岩穴口孝, 熊本一朗	4. 巻 38 Suppl.
2. 論文標題 文章ベクトル化における調整パラメーター値とカットオフ値の検討 看護観察記録を用いた誤嚥性肺炎発見手法において	5. 発行年 2018年
3. 雑誌名 医療情報学	6. 最初と最後の頁 782-785
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 小牧祥太郎, 村永文学, 宇都由美子, 岩穴口孝, 熊本一朗	4. 巻 37
2. 論文標題 誤嚥性肺炎予防の為に、観察記録解析における文書ベクトル化技法の有用性の検討	5. 発行年 2017年
3. 雑誌名 医療情報学連合大会論文集	6. 最初と最後の頁 380-383
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 宇都 由美子, 花原 康代, 岩穴口 孝, 落合 美智子, 市村 カツ子, 村永 文学, 熊本 一朗	4. 巻 37
2. 論文標題 ケアの実施データの二次利用による持続可能な看護計画の見直しツールの開発とその評価	5. 発行年 2017年
3. 雑誌名 医療情報学連合大会論文集	6. 最初と最後の頁 695-697
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 村永文学, 岩穴口孝, 宇都由美子, 熊本一朗	4. 巻 37
2. 論文標題 高額医療費の要因分析に適したデータマイニング手法に関する研究	5. 発行年 2017年
3. 雑誌名 医療情報学連合大会論文集	6. 最初と最後の頁 484-487
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 岩穴口 孝, 宇都 由美子, 村永 文学, 熊本 一朗	4. 巻 17
2. 論文標題 継続的な看護ケアに不可欠な退院時看護要約のためのケア実施入力データの分析	5. 発行年 2017年
3. 雑誌名 日本医療情報学会看護学術大会論文集	6. 最初と最後の頁 69-70
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件(うち招待講演 0件/うち国際学会 0件)

1. 発表者名 小牧祥太郎, 村永文学, 宇都由美子, 岩穴口孝, 熊本一朗
2. 発表標題 文章ベクトル化における調整パラメーター値とカットオフ値の検討 看護観察記録を用いた誤嚥性肺炎発見手法において
3. 学会等名 第38回医療情報学連合大会
4. 発表年 2018年

1. 発表者名 小牧祥太郎
2. 発表標題 誤嚥性肺炎予防の為の、観察記録解析における文書ベクトル化技法の有用性の検討
3. 学会等名 第37回医療情報学連合大会
4. 発表年 2017年

1. 発表者名 村永文学
2. 発表標題 高額医療費の要因分析に適したデータマイニング手法に関する研究
3. 学会等名 第37回医療情報学連合大会
4. 発表年 2017年

1. 発表者名 宇都 由美子
2. 発表標題 ケアの実施データの二次利用による持続可能な看護計画の見直しツールの開発とその評価
3. 学会等名 第37回医療情報学連合大会
4. 発表年 2017年

1. 発表者名 岩穴口 孝
2. 発表標題 継続的な看護ケアに不可欠な退院時看護要約のためのケア実施入力データの分析
3. 学会等名 第17回日本医療情報学会看護学術大会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	村永 文学 (Muranaga Fuminori) (00325812)	鹿児島大学・医歯学総合研究科・客員研究員 (17701)	
研究分担者	宇都 由美子 (Uto Yumiko) (50223582)	鹿児島大学・医歯学域医学系・准教授 (17701)	
研究分担者	岩穴口 孝 (Iwaanaguchi Takashi) (80619198)	鹿児島大学・医歯学域鹿児島大学病院・助教 (17701)	

6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 協力 者	小牧 祥太郎 (Komaki Shotaro)		