

# レセプトデータを用いたデータマイニングによる シグナル検出に関する検討報告書

平成 23 年6月

独立行政法人 医薬品医療機器総合機構

安全第一部 調査分析課

略語・略号一覧

略語	正式名称
JMDC	株式会社日本医療データセンター
maxSPRT	Maximized sequential probability ratio test
MGPS	Multi-item Gamma Poisson Shrinker
NEC	日本電気株式会社
PMDA	Pharmaceuticals and Medical Devices Agency 独立行政法人 医薬品医療機器総合機構
ROR	Reporting Odds Ratio
PRR	Proportional Reporting Ratio
SPRT	Sequential probability ratio test

## 用語一覧

用語	内容
シグナル	<p>ある有害事象とある医薬品に因果関係があるかもしれないとされた情報のことで、それまでに因果関係が不明か明記されていないもの。</p> <p>注) 別添の報告書における定義では、有害事象が発生した患者とその発生時期を指す。</p>
ナショナルレセプトデータベース	<p>高齢者の医療の確保に関する法律（昭和 57 年法律第 80 号）第 16 条の規定に基づき、厚生労働省保険局が構築しているレセプト情報・特定健診情報等データベース。医療費適正化計画の作成等に資する調査・分析を行う等の高齢者医療確保法に基づく利用の他、医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進への利用についても、厚生労働省においてデータ利用の公益性等に関し検討が進められている。</p>
副作用	<p>病気の予防、診断もしくは治療、または生理機能を変える目的で投与された（投与量にかかわらず）医薬品に対する反応のうち、有害で意図しないもの。医薬品に対する反応とは、有害事象のうち当該医薬品との因果関係が否定できないものを言う。</p>
有害事象	<p>医薬品が投与された患者または被験者に生じたあらゆる好ましくない医療上のできごと。必ずしも当該医薬品の投与との因果関係が明らかなもののみを示すものではない。</p>
レセプト	<p>診療報酬明細書。主に「内科レセプト」、「歯科レセプト」、「調剤レセプト」の 3 種類に分類され、「内科レセプト」は更に「入院」と「入院外」に分類される。レセプトは、審査支払機関へ診療報酬の請求を行うために各医療機関より患者単位、一か月単位で発行されている。診療行為ごとに診療報酬点数が決められており、医療機関はこの点数を合算して保険者に医療費を請求する。</p>

## 目次

---

目次 .....	iii
1. 背景 .....	1
2. 目的 .....	2
3. データマイニング検討 .....	3
3-1. シグナル検出手法 .....	3
3-2. レセプトデータ .....	3
3-3. データマイニング手法検討 .....	4
4. 別添報告書に対する考察 .....	6
5. 総括 .....	9

## 1. 背景

---

PMDA は、第二期中期目標において、医薬品の安全対策の高度化を目指すため、「IT 技術の活用により、新規安全性情報の検出や解析を行うための手法を検討し、効率的・効果的な安全性情報評価体制を構築する」こと及び「講じた安全対策措置について、企業及び医療機関等における実施状況及び実効性が確認できる体制を構築する」ことを定めた。その実施計画として、「診療情報データベースのアクセス基盤を整備し、副作用の発現頻度調査や薬剤疫学的解析を実施できる体制を構築する」こと及び「安全対策措置の効果について、企業が実施する評価と並行して、必要に応じて独自に調査・確認・検証できる体制を、平成 23 年度から構築する」ことを定めた。

これらの第二期中期計画を実行するため、平成 21 年度より新規プロジェクトである MIHARI-Medical Information for Risk Assessment Initiative を安全第一部調査分析課にて立ち上げ、各種電子診療情報データの収集方法、データ特性分析、副作用等の安全性情報の抽出方法、解析手法の検討を開始した。また、検討内容に対して専門的助言を得ることを目的に、医療情報学、薬剤疫学、生物統計学等、関連する学術分野の専門家から構成される「電子診療情報等の安全対策への活用に関する検討会」を平成 21 年度に PMDA 内に設置した。

各種電子診療情報とは、主にレセプトデータ、DPC データ、病院情報システムデータを指すが、このうちレセプトデータについては、厚生労働省においてナショナルデータベースの構築が進められており、そのデータの二次利用に向けた議論が平成 22 年 10 月より開始された。PMDA では、医薬品の安全対策にナショナルデータベースの利用が近い将来可能となるであろうことを視野に入れつつ、当面は市販されているレセプトデータを用いてデータの活用方法を検討していく計画である。

レセプトデータを医薬品の安全性評価に活用するため、平成 21 年度は、市販のレセプトデータを用いて基本的な集計を実施することにより、データの特性を把握することに注力した。平成 22 年度は、同じ市販のレセプトデータを用いて、活用目的が異なる試行調査を複数並行して実施し、各々の目的に適した分析手法の検討を行い、これらの分析手法の実効性をレセプトデータの特性や限界と合わせて評価した。

本報告書は、平成 22 年度の試行調査の一つとして実施した、データマイニングによるシグナル検出の検討結果をまとめたものである。

## 2. 目的

---

現在、PMDAにおける医薬品の市販後安全性評価では、自発報告を中心にシグナルの検出を行っている。このシグナルの検出には、報告症例をレビューすることにより人がシグナルと認識する場合とデータマイニング手法により計算の結果得られる場合の両方が含まれる。いずれの方法であっても、自発報告は、臨床現場で実際に生じている副作用が全て報告されるとは限らないため、検出されるシグナルにも偏りがある可能性が考えられる。自発報告に対し、レセプトデータは診療報酬請求対象の一般の診療情報が全て含まれるため、自発報告からのシグナル検出を補完するものとして、レセプトデータをシグナル検出に活用できるのではないかと考えた。自発報告で検出されないシグナルを発見し既存のシグナル検出の補完の役割を果たすには、シグナル検出の対象は、特定の有害事象や医薬品に限定せず、様々な有害事象及び医薬品とする必要がある。そこで、本検討においては、レセプトデータより網羅的にシグナルを検出する手法を開発・構築することを目的として探索的検討を行う。得られた検討結果について、安全対策業務に適用できるような実効性のある手法が構築されたか、もしくはその見込みがあるのかを評価する。

## 3. データマイニング検討

---

### 3-1. シグナル検出手法

---

レセプトデータよりシグナルを検出する手法については、自発報告に適用される単純なデータマイニング手法（ROR 法、PRR 法や MGPS 法等）とは異なる手法が必要と考えられる。自発報告は、副作用が疑われる情報のみで構成されるデータであり、マイニングの対象とする有害事象／副作用は確定した情報として扱うことができ、医薬品は既に被疑薬と標識化されている。これに対し、レセプトデータは一般診療の情報から成るデータであり、明示された有害事象／副作用情報がないため、病名や医薬品の処方タイミング、診療行為（臨床検査等）の有無等、レセプトデータ上の様々なイベントの組み合わせから有害事象／副作用の発生を推測しなければならず、簡単に有害事象／副作用や被疑薬と判断できるものはごく僅かである。このようなデータから同じ手法を用いて目的のシグナルを検出することは困難と考えられる。また、海外においてレセプトデータを用いたシグナル検出手法（SPRT 法、maxSPRT 法等）の検討が進められているが、これらの手法は主に特定の有害事象と医薬品に注目してシグナルを検出するものであり、今回目的とする網羅的な有害事象と医薬品のシグナル検出には適用できない。従って、これまでに有害事象／副作用の検出目的で検討されている手法とは異なる金融やコンピュータセキュリティの分野等において使用されている大規模なデータから目的とする事象を検出するデータマイニング手法を用いて、レセプトデータからのシグナル検出を検討する。

### 3-2. レセプトデータ

---

本検討には、JMDC のレセプトデータを使用した。このデータは、複数の健康保険組合のレセプトデータを統合したもので、2005 年 1 月から 2008 年 12 月に発行された医科レセプト、調剤レセプトのデータが含まれる。JMDC において予め名寄せ作業が施されているため、組合員が当該健康保険組合から離脱しない限り、複数の医療機関にまたがる保険医療の情報を患者単位に追跡することが可能である。また、組合員の扶養家族の情報も含まれるため、乳幼児や小児のデータも含まれている。ただし、定年退職に伴い健康保険組合から離脱することにより、60 歳以上のデータは減少し、特に 65 歳以上では著しく減少している。その他、JMDC レセプトデータの特徴は平成 21 年度の「レセプトデータを用いたアナフィラキシーに関する試行調査報告書

([http://www.info.pmda.go.jp/kyoten\\_iyaku/file/e\\_rece-report1006.pdf](http://www.info.pmda.go.jp/kyoten_iyaku/file/e_rece-report1006.pdf)) を参照のこと。

本データセットは、2008年12月時点でデータ固定されており、対象者の年齢は、全て誕生年の1月1日を誕生日として算出されている。また、本データセットのデータ項目を図1に示す。

### 3-3. データマイニング手法検討

---

データマイニング手法の検討は専門技術を有する業者へ委託することとし、入札の結果、NECに決定した。

検討内容については、別添の「診療報酬明細書のデータを用いたデータマイニングに関する技術的検討業務 2010年度検討報告書」（NEC作成）を参照のこと。



レセプトデータを用いたデータマイニングによるシグナル検出に関する検討報告書

①患者

項目	1_0	1_1	1_2
患者ID	性別	年齢 (2008年1月 時点)	

②診療(医科入院外)

項目	2_0	2_1	2_2	2_3	2_4	2_5	2_6	2_7	2_8
患者ID	入院外 医科レセID	診療年月 (受診年月)	医療施設ID	経営体	病床数	DPC対象 病院フラグ	医療費(点 数)	診療実日数	

③診療(医科入院)

項目	3_0	3_1	3_2	3_3	3_4	3_5	3_6	3_7	3_8	3_9
患者ID	入院 医科レセID	診療年月 (受診年月)	医療施設ID	経営体	病床数	DPC対象 病院フラグ	医療費(点 数)	入院日	診療実日数	

④医薬品(調剤)

項目	4_00	4_01	4_02	4_03	4_04	4_05	4_06	4_07	4_08	4_09	4_10
患者ID	調剤レセID	明細ID	調剤レセ年 月	医療施設ID	薬局ID	調剤日	処方月日	医薬品名	成分名	一般名	

	4_11	4_12	4_13	4_14	4_15	4_16	4_17	4_18	4_19	4_20	4_21	4_22	4_23
薬価基準収 載医薬品 コード	87分類コ ード (再再分類 まで)	ATCコード 【細または 小】	ATC名称 【細または 小】	剤型大分類	剤型中分類	規格単位	投与量(1日 当り)	日数(回数)	後発品フラ グ	頓用フラグ	医師ID	JDM診療科	

⑤医薬品(入院)

項目	5_00	5_01	5_02	5_03	5_04	5_05	5_06	5_07	5_08	5_09	5_10
患者ID	入院 医科レセID	明細ID	診療区分	レセ年月	医薬品名	成分名	一般名	薬価基準収 載医薬品 コード	87分類コ ード (再再分類 まで)	ATCコード 【細または 小】	

	5_11	5_12	5_13	5_14	5_15	5_16	5_17	5_18	5_19
ATC名称 【細または 小】	剤型大分類	剤型中分類	規格単位	投与量(1日 当り)	日数(回数)	後発品フラ グ	頓用フラグ	JDM診療科	

⑥医薬品(入院外)

項目	6_00	6_01	6_02	6_03	6_04	6_05	6_06	6_07	6_08	6_09	6_10
患者ID	入院外 医科レセID	明細ID	診療区分	レセ年月	医薬品名	成分名	一般名	薬価基準収 載医薬品 コード	87分類コ ード (再再分類 まで)	ATCコード 【細または 小】	

	6_11	6_12	6_13	6_14	6_15	6_16	6_17	6_18	6_19
ATC名称 【細または 小】	剤型大分類	剤型中分類	規格単位	投与量(1日 当り)	日数(回数)	後発品フラ グ	頓用フラグ	JDM診療科	

⑦診療行為

項目	7_00	7_01	7_02	7_03	7_04	7_05	7_06	7_07
患者ID	医科レセID	明細ID	診療区分	レセ年月	標準化診療 行為名	標準化診療 行為コード	回数	

⑧傷病

項目	8_00	8_01	8_02	8_03	8_04	8_05	8_06	8_07	8_08	8_09	8_10	8_11
患者ID	医科レセID	明細ID	標準傷病 コード	標準傷病名	ICD-10コ ード	ICD-10細分 類名(頭 コード付き)	ICD-10小分 類名(頭 コード付き)	疑いフラグ	主傷病フラ グ	診療開始日	転帰	

図 1. JMDC レセプトデータセットのデータ項目

## 4. 別添報告書に対する考察

データマイニングの専門的観点からの考察は別添の報告書に示されているため、ここでは、主に PMDA が検討用に提供した有害事象に関する情報の観点から検討結果に対して考察する。

本検討では、大量のレセプトデータから効率的にシグナルを検出するため、有害事象が発生したと考えられる患者 ID と発生時期をシグナルとして検出する方法を中心に検討が進められた。この検討において、PMDA は主に有害事象に関する情報の提供と、データマイニングによるシグナル検出結果の評価を担当した。シグナル検出器の構築には教師付き学習法が採用されたことから、学習用に正例（有害事象発現時のデータ）及び負例（有害事象非発現時のデータ）の用意が必要であった。そこで、PMDA にて資料や臨床の専門家の意見等を参考に、レセプト上に現れると考えられる有害事象名、治療薬名、臨床検査名等をまとめた辞書と原因薬の中止、治療薬や臨床検査の実施等時間的なタイミングを定めたスクリーニングルールを作成した。また、検出されたシグナルについて傷病名、医薬品名、診療行為名等の情報を目視により確認し、有害事象が疑われるか否かの判定を行った。

データマイニングによるシグナル検出の検討は、課題 1、課題 2 の二段階で検討され、課題 1 では比較的重篤な 6 つの有害事象（劇症肝炎・肝不全、間質性肺炎、間質性腎炎・急性腎不全、痙攣、意識障害、急性脳症）に限定し、辞書に依存したシグナル検出の検討を行った。結果は、別添の報告書 図 3-10~13 に示されたとおり、実験 1（同一患者集団において、有害事象発現時を正例、そうでない時期を負例と設定）では、感度 0.8 で陽性的中率 0.9、また、上位 1,800 件中 1,600 件が正例と非常に高い検出精度を示した。一方、実験 2（異なる患者集団で、有害事象発現時を正例、そうでない時期を負例と設定）では、感度 0.6 で陽性的中率 0.7、また、上位 200 件で 94 件が正例と実験 1 より精度は低下したものの、工夫次第では精度向上が見込まれる結果であった。この検討においては、比較的重篤と考えられる有害事象を選択したため、これらに対応する治療薬や診療行為の開始、原因薬の中止等、特徴的な変化がレセプトデータ上に認められることが多かったのではないかと考えられた。実験 1 の同一患者集団を正例と負例に用いた場合では、この特徴的な変化を判別するように検出器が設定されたことにより、理想的な結果が得られたものと思われた。実験 2 では、正例と負例で患者集団が異なるため、正例に特徴的なデータが多かったものの、負例の背景情報が多様化し、それがノイズとなり、検出精度がやや低下したものと考えられた。

課題 2 では、辞書に依存せずシグナル検出器を一般化することで辞書に記載されていない有害事象も検出する方法の検討を行った。ここでは、別添報告書に記載されたとおり、学習用データに 3 パターンを使用した。パターン 1 では、課題 1 の 6 つの有害事象

に加え、脳梗塞、心不全、低血糖、血小板減少等、有害事象の種類を拡大し、重篤とは限らない血小板減少等の有害事象も含めた。また、スクリーニングルールを6つの有害事象に限定的なものから一般化し、さらに課題1に比べより適切に正例を抽出するためルールを厳格化した。これにより、有害事象の数は増えたものの、パターン1の正例数は57件と課題1で使用した正例数1,959件より大きく減少した。パターン2では、本検討以外の試行調査等にて特定した5つの有害事象（光線過敏症、骨粗鬆症、パーキンソニズム、無顆粒球症、肝障害）を用い、発現時点を正例、それ以外の時期を負例とした。正例の数は162件であった。光線過敏症以外は、治療薬や診療行為等傷病名以外の情報も組み合わせて発現時点を特定し正例とした。光線過敏症は傷病名で特定した上で、その前後の処方、診療行為名、傷病名の情報を目視にて確認し正例とした。これらパターン2の有害事象についても、重篤でない例が含まれていることが考えられた。パターン3では、医薬品が原因と考えられる傷病名「薬剤性〇〇」が付与された時点を正例とし、それ以外の時期を負例とした。正例の数は4,249件であった。傷病名に敢えて「薬剤性」が付与されているということは、その傷病が医薬品によるものである可能性が高いと考えた。病名の内訳では、薬疹が最も多く1,263件、2番目が薬剤性肝障害1,146件、3番目が薬剤性パーキンソン症候群691件でこれらが正例の6割を占め、その他の病名は数件ずつで病名の種類としては合計52種類であった。

課題1と課題2のシグナル検出精度の比較では、パターン1の学習データを用いて検出精度の比較を行ったが、課題1の方が課題2よりも精度が高いという結果であった（別添 図4-3）。また、課題1では図3-10~13と比較すると検出精度は低下した。これらの結果は、有害事象の種類の多様化によりシグナル検出するために判別すべきデータが多様化したこと、それに比べて学習用の正例が減ったことが、検出精度の低下の一因と考えられた。さらに、課題2では、辞書に依存せず不特定の有害事象の検出を試みたため、検出精度は課題1よりも大幅に低下しており、課題2で設定した目標を到達することはかなり困難であることが示唆された。

学習に用いたデータの有害事象パターン別に精度を比較した結果では、シグナル検出において上位にランクされたもので比較すると、最も精度が低かったのはパターン1、中間がパターン1+パターン2、最も精度が高かったのはパターン3であった（別添 図4-6）。パターン1とパターン2で有害事象の特徴が大きく異なるとは考えられないため、正例に含まれるデータの特徴は共通性があったのではないかと思われた。従って、検出精度の差は、正例がパターン1+パターン2の方が多かったことが主な原因と考えた。パターン3は正例が多く、3つの有害事象（薬疹、薬剤性肝障害、薬剤性パーキンソン症候群）が正例中約6割を占めたことから、比較的共通性を持ったデータが多かったことが考えられた。しかしながら、課題1の結果と比較するとパターン3であっても検出精度はかなり低かった。

上記以外にも複数の観点から検出精度向上の検討がなされたが、いずれも飛躍的な精度向上には至らなかった。別添報告書「4-11 まとめと考察」にもあるとおり、本検討においては、学習データとする正例・負例を用意すること、及び検出したシグナルの評価が困難であったことが非常に大きな課題であった。レセプトデータの情報のみを用いて有害事象発現時点を特定することは医師等臨床の知識を有する専門家であっても難しく、この方法により正例、負例を作成することには限界があると思われた。更なる精度向上には、マイニング技術の高度化も必要と思われるが、より適切な正例・負例の用意が最も重要な改善ポイントと考えられた。そのためには、レセプトデータ以外の情報源（病院情報システムや自発報告等）と連結した正例・負例の作成等が要求されるであろう。また、同じく別添報告書にも記載されているように、多様な特徴を持つ有害事象を検出するには、有害事象を似た特徴ごとにグルーピングして学習させる等、シグナル検出工程に前処理を加えることにより検出精度向上が見込まれると考えられた。

本検討では、課題2のシグナル検出検討結果を利用して、検出上位の症例における有害事象と医薬品、検査の組み合わせを抽出する特徴抽出の検討も併せて実施された。この検討の中で、原因薬、治療薬、診療行為、傷病名の組み合わせ 61,604 件を目視確認したところ、明らかに組み合わせとして適切でないものが大半であったためそれらを除外し、さらに精査した結果、9 件のみが有害事象発現時の組み合わせとしてあり得るものであるものであった。別添報告書の「5-7 まとめと考察」にあるとおり、シグナル検出の結果を利用した今回の特徴抽出方法では非常に効率が低く、実用化にはかなり遠いことが示された。しかしながら、抽出条件を工夫することで改善の余地があることが示されたこと（別添 表 5-2）から、角度を変えて引き続き検討する価値はあるのではないかと考えられた。

## 5. 総括

---

レセプトデータを用いたデータマイニングによるシグナル検出手法の検討は、一定範囲での検討結果が得られ、多様な有害事象のシグナル検出精度の向上には、より適切な正例と負例の提供が必要という課題が示された。また、特徴抽出については、副次的に実施した検討であったことから限定的な結果が示されるに留まったが、更なる検討の余地はあることが示された。

今回の検討結果では、レセプトデータのシグナル検出が自発報告のシグナル検出を補完するという目的を達することが可能であるのかを評価するには不十分であると思われた。今回中心的に検討を進めたデータマイニングによるシグナル検出に加え、特徴抽出の技術を利用することにより総合的に目的とするシグナルの検出が実現する可能性が考えられる。従って、シグナル検出について更なる検討は必要であるが、特徴抽出についてもシグナル検出とは独立した方法により一定の検討が必要と考えられた。ただし、検討継続に際しては、正例、負例の提供の課題への対応方法、検出対象とする有害事象のグループ化や絞り込み、特徴抽出に用いる学習データ等について予め検討する必要がある。

以上

別添

診療報酬明細書のデータを用いた  
データマイニングに関する技術的検討業務  
2010年度検討報告書

平成23年3月

日本電気株式会社

## 用語一覧

クレンジング	データ集合に対して、誤記や入力規則に違反しているデータを修正、あるいは削除すること。これにより、機械的なデータ分析の効率が向上する。
コーディング	データを事前に決められた値の集合（コード）のどれかに変換すること。事前に決められた値の集合のみを処理対象とすることにより、機械的なデータ分析の効率が向上する。
シグナル検出器	時系列データを入力とし、特定の時刻における対象事象発生疑わしさを出力するシステム。特に本報告においては、患者のレセプト系列情報から生成した時系列データを入力とし、当該期間の中心の時刻における副作用発生疑わしさをスコアとして出力するプログラム。
正例・負例	<p>正例：副作用発現が強く疑われる患者のレセプト系列、あるいはそれから生成した時系列データ</p> <p>負例：副作用発現ではないと強く推量される患者のレセプト系列、あるいはそれから生成した時系列データ</p> <p>なお、本報告においては、実際に副作用が発生していたか／いなかったかを感知する手段がないため、一定のルールに基づいて正例と負例の集合を定義することとしている。</p>
機械学習	学習用のデータ集合を解析し、傾向やルールを抽出することで、特定の目的を達成するシステムを構築／調整すること。本報告においては、学習用の正例と負例（および正例か負例かが不明なレセプト系列）を解析することで、シグナル検出器に入力すべき時系列データの選択と、シグナル検出器の可変パラメータの調整を行う。
クロスバリデーション（交差検定）	データを学習用と評価用に分け、学習用データを解析して得られた結果の有用性を、評価用データを用いて評価すること。本報告では、学習用データを用いてシグナル検出器

	を構築・調整し、評価用データを当該シグナル検出器に入力した結果を用いて、シグナル検出器の精度などを評価している。
副作用シグナルスコア	シグナル検出器が出力する「特定の時刻における対象事象発生の疑わしさ」を表す指標。単に、スコアやシグナルスコアとも表記する。
過学習	モデルが学習データに対して過適合（オーバーフィット）し、評価データに対する汎化性能が悪化する現象。モデルの複雑性が高すぎる、データ数に対して変数の数が多すぎるなどの原因によって発生する。
スクリーニング	副作用発現時のレセプト系列を絞り込むために、臨床的観点から設定した基準に沿ってレセプトデータを抽出すること。本報告では、副作用辞書に記載されている項目が、特定の発生パターンをもって現れているか否かを当該基準として抽出を行っている。
特性または特性候補	レセプト系列から生成される特徴量（レセプトの特徴を表現する量）のこと（特性候補は、その候補）で、シグナル検出器への入力となる。本報告では、医薬品や診療行為等の時間的なパターンを特性として利用した。
特徴抽出	シグナルスコアが大きいレセプト系列に固有の臨床イベントを抽出すること。本報告では、当該レセプト系列において、シグナルスコアを大きくすることに寄与している薬剤、診療行為、傷病名、あるいはそれらの組み合わせを抽出している。
ナショナルレセプトデータベース (NRDB)	高齢者の医療の確保に関する法律（昭和57年法律第80号）第16条の規定に基づき、厚生労働省保険局が構築しているレセプト情報・特定健診情報等データベース。医療費適正化計画の作成等に資する調査・分析を行う等の高齢者医療確保法に基づく利用の他、医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進への利用についても、厚生労働省においてデータ利用の公益性等



	に関し検討が進められている。
副作用仮説	未知の副作用の「原因薬、治療薬、診療行為、傷病名」の組み合わせ、あるいは、その部分的組み合わせに関する仮説。本報告では、未知であるという状況を擬似的に作り出すために、特徴抽出によって出力された臨床イベントの組み合わせの中で、シグナル検出器の学習に用いた副作用辞書の内容に該当しないものを、副作用仮説とみなしている。
副作用辞書	劇症肝炎・肝不全や間質性肺炎、脳梗塞や心不全といった副作用のタイプ別に、原因となる薬効群、治療薬、診療行為、副作用傷病名のそれぞれの項目をまとめたもの。本報告では、スクリーニングに利用することで、正例と負例の作成を行っている。
レセプト	診療報酬明細書。主に「内科レセプト」、「歯科レセプト」、「調剤レセプト」の3種類に分類され、「内科レセプト」は更に「入院」と「入院外」に分類される。レセプトは、審査支払機関へ診療報酬の請求を行うために各医療機関より患者単位、一ヶ月単位で発行されている。診療行為ごとに診療報酬点数が決められており、医療機関はこの点数を合算して保険者に医療費を請求する。
PMDA	Pharmaceuticals and Medical Devices Agency 独立行政法人 医薬品医療機器総合機構
レセプト系列	特定患者の特定期間のレセプト情報からなる時系列データ。本報告では、特定期間としてシグナル検出の対象年月を中心とした12ヶ月や同6ヶ月、3ヶ月などを採用している。

## 目次

目次 .....	iv
1. 本検討業務の概要 .....	1
2. はじめに .....	3
2-1 目的 .....	3
2-2 レセプトデータ .....	3
2-3 検討課題 .....	3
3. シグナル検出課題 1 : .....	6
副作用辞書に依存した副作用検出方法の検討 .....	6
3-1 副作用辞書の作成 .....	9
3-2 データ構造化／可視化 .....	10
3-3 マイニング用学習データの抽出 .....	11
3-4 レセプト系列特性の候補設計 .....	12
3-5 特性の同定と検出器の構築 .....	14
3-6 シグナル検出結果の評価 .....	15
3-7 医療費情報の活用検討 .....	20
3-8 まとめと考察 .....	20
4. シグナル検出課題 2 : .....	21
副作用辞書に非記載の副作用検出方法の検討 .....	21
4-1 マイニング用学習データの追加と再定義 .....	23
4-2 レセプト系列特性の一般化 .....	24
4-3 レセプト系列特性の追加 .....	25
4-4 シグナル検出課題 2 の評価方法 .....	26
4-5 シグナル検出結果の評価 1 （課題 1 と課題 2 の比較） .....	27
4-6 シグナル検出結果の評価 2 （特性の追加） .....	29
4-7 シグナル検出結果の評価 3 （負例数増加） .....	30

4-8 シグナル検出結果の評価4（副作用パタンの比較）	32
4-9 シグナル検出結果の評価5（特性データの発生パターン長）	34
4-10 シグナル検出結果の評価6（誤検出に関する考察）	35
4-11 まとめと考察	38
5. 特徴抽出による副作用仮説抽出の検討	40
5-1 特徴抽出と副作用仮説抽出	40
5-2 特徴抽出手順	40
5-3 特徴抽出結果の評価1（シグナル検出結果／検出器の挙動把握）	42
5-4 特徴抽出結果の評価2（マスクされた副作用仮説の再現）	43
5-5 特徴抽出結果の評価3（そのほかの副作用仮説の抽出）	44
5-6 副作用仮説発見効率の向上可能性	45
5-7 まとめと考察	47
6. NRDB への適用に向けた課題	48
7. まとめ	50
付録	51
A. シグナル検出アルゴリズム	51
B. 特徴抽出アルゴリズム	52

## 1. 本検討業務の概要

<p><b>目的</b></p>	<p>本業務の目的は、数十万人規模のレセプトデータを用いて、副作用発現等の臨床イベントのシグナルを網羅的に検出するのに適したマイニング手法を検討することである。特に、副作用発現の有無はレセプトデータには明示的に記載されていないため、処方、診療行為、傷病名、医療費などの代替指標からこれを推定・抽出する方法の検討を本業務の目的とする。</p>
<p><b>検討課題</b></p>	<p><b>シグナル検出課題 1：辞書に依存した副作用検出方法の検討</b></p> <p>ターゲットとする副作用に関する原因薬、治療薬、診療行為、傷病のリスト（副作用辞書）が与えられた場合に、ターゲット副作用のシグナルを精度よく検出する方法を検討する。</p> <p><b>シグナル検出課題 2：辞書不記載副作用の検出方法の検討</b></p> <p>辞書依存の副作用検出結果から得られる知見をもとに、ターゲットとして学習した辞書記載の副作用だけでなく、類似の特性を示す辞書不記載の副作用を検出する方法を検討。</p> <p><b>特徴抽出</b></p> <p>シグナル検出の結果から「疑いが強い」とされたレセプト系列で、どのような臨床イベントが起きているかを機械的に分析する方法を検討。</p>
<p><b>レセプトデータ</b></p>	<p>本検討業務においては、日本医療データセンター（JMDC）社が有する健康保険組合のレセプトデータベースからのデータ（約40万人分、2005年1月～2008年12月）を使用。</p>
<p><b>副作用辞書</b></p>	<p>劇症肝炎・肝不全や、NSAIDsによる光線過敏症、そのほか薬剤性甲状腺機能障害といった既に知られている副作用とその原因となり得る薬効群の組み合わせ合計約170タイプをターゲットとして、PMDAにて臨床的知見を基に作成。</p>
<p><b>結果概要</b></p>	<p>レセプト系列の医薬品や診療行為パターンを特性とし、「副作用発現の疑いの強さ」を出力するシグナル検出器を機械学習技術によって構成し、副作用辞書に記載された副作用と記載されていない副作用の両者についてシグナル検出の可能性を評価した。実験結果からは、学習した副作用イベントだけでなく、類似の特性をもつ未学習の副作用イベントに関しても、副作用が発現しているレセプト系列を特定できる可能性がある事が示唆された。一方で、正例数の不足、正例と類似の特性を示す負例の</p>

	<p>存在等に起因し、直接実務で利用するに十分な検出精度は得られなかった。</p> <p>また、シグナルスコアが大きいレセプト系列固有の臨床イベントを、機械的に抽出する「特徴抽出」分析に関しても検討を行った。実験からは、シグナルスコアが大きいレセプト系列の特徴を抽出するだけでは、副作用仮説（薬剤、診療行為、傷病名等の組み合わせ）発見の効率は低かったが、仮説発見を目的とした分析手法で効率向上できる可能性は十分に期待できることが確認された。</p>
<p><b>考察</b></p>	<p>シグナル検出精度向上に向けては、特性データの高度化（組合せ、外部情報、非線形等）やシグナル検出ロジックの高度化（ルールとの融合、学習データ獲得の効率化等）等が今後の重要な課題として挙げられる。</p> <p>また、特徴抽出については、本検討業務は副作用仮説の抽出という視点に関しては、重要でありながらも、副次的な取り組みによって可能性を示したにとどまっている。「副作用仮説の抽出」という視点からの本格的な取り組み、および今回の結果との融合方法の検討は重要な今後の課題である。</p>

## 2. はじめに

---

### 2-1 目的

---

本業務の目的は、数十万人規模のレセプトデータを用いて、副作用発現等の臨床イベントのシグナルを網羅的に検出するのに適したマイニング手法を検討することである。特に、副作用発現の有無はレセプトデータには明示的に記載されていないため、処方、診療行為、傷病名、医療費などの代替指標から副作用発現を推定・抽出する方法の検討を本業務の目的とする。

「副作用発現等の臨床イベントのシグナルを検出する」という目標に向けては、

- シグナル検出：副作用が発現しているレセプト系列の特定
- 特徴抽出：副作用仮説（原因薬、治療薬、診療行為、傷病の組合せ）の抽出

という2つの視点があり、一般的にはデータマイニングのアプローチがそれぞれ異なる。本業務では、シグナル検出を主たる目標とし、マイニング手法によるシグナル検出の可能性検討を行った。なお、特徴抽出に関しても、前者の副次的な取り組みとして検討を行ったため、その結果を報告する。

### 2-2 レセプトデータ

---

検討に用いたレセプトデータは、JMDC社のレセプトデータベースからのデータである。患者数は約40万人であり、レセプト数にしてそれぞれ、入院外医科：約713万件、入院医科：約11万件、調剤：約306万件相当のデータからなる。データ期間は2005年1月から2008年12月までの48ヶ月間で、JMDC社によりクレンジング、コーディングが施された上で、リレーショナルデータベースとして構造化されているものである。データ構造および主な項目は付録参照。

### 2-3 検討課題

---

本業務では、大目標としてシグナル検出を検討し、副次的な取り組みとして特徴抽出を行った（2-1参照）。以下では、各検討の概要を説明する。

### 2-3-1 シグナル検出

シグナル検出は、以下の二つの課題を設定し検討を進めた。

#### シグナル検出課題1：副作用辞書に依存した副作用検出方法の検討

ターゲットとする副作用に関する原因薬、治療薬、診療行為、傷病のリスト（以下、「副作用辞書」。副作用辞書については3-1節参照。）が与えられた場合に、ターゲット副作用のシグナルを検出する方法を検討。

#### シグナル検出課題2：副作用辞書に非記載の副作用検出方法の検討

辞書依存の副作用検出結果から得られる知見をもとに、ターゲットとして学習した辞書記載の副作用だけでなく、類似の特性を示す辞書不記載の副作用を検出する方法を検討。

本業務の目的である網羅的な検出方法の検討は課題2に相当するが、その検討を進めるために必要な「副作用シグナルを有するレセプト系列の傾向」等に関する知見を得るために、課題1を実施することは必須となる。逆に課題1において、十分な精度でシグナル検出ができないのであれば、課題2は検討するまでもなく網羅的なシグナル検出が非現実的であると結論付けることができる。よって本検討では、まず課題1に取り組み、一定の見通しが立った上で課題2に取り組むこととした<sup>1</sup>。

### 2-3-2 特徴抽出

上記シグナル検出の出力は、入力されたレセプト系列に対する「副作用発現の疑わしさ」であり、これを用いることにより、副作用が発現していると疑われる「患者とそのタイミング（年月）」を大量のレセプトデータから抽出することができる。一方、そこでどのような臨床イベントが起きていたかについては、原理的には元のレセプト系列を可視化・目視することで確認できるはずであるが、疑わしいとして抽出されるレセプト系列が多数の場合は、非常に大きい手間を必要とする。また、疑わしいレセプト系列固有の臨床イベントを確認したい場合には、疑わしいとして抽出されないレセプト系列に

---

<sup>1</sup> また、課題1において、辞書記載の「原因薬、治療薬、診療行為、傷病」の同時発生によりシグナルを検出する場合には、ノイズとなる誤報（偽陽性）が多数発生する事が想定されるが、辞書情報にマイニング技術を合わせる事で偽陽性のシグナル数を低減する事が可能かを副次的に観察する事が可能である。

おける臨床イベントとの比較も必要になり、元データの可視化・目視という方法は現実的ではなくなる。

そこで本検討では、スコアが大きいレセプト系列固有の臨床イベントを、機械的に抽出する「特徴抽出」分析に関しても検討を行った。特徴抽出には様々な方法が考えられるが、今回は、特にシグナル検出に関する検討の文脈にそった方法を採用した。すなわち、シグナル検出器が大きいスコアを発するレセプト系列において、検出器が出力するシグナルのスコアを増大させることに寄与していた薬剤、診療行為、傷病名の共起組み合わせを抽出し、それらの頻度を、シグナルスコアの大きさの異なるレセプト系列群で比較するという動作原理に基づく、特徴抽出方式を検討の対象とした。

このような特徴抽出は、上記のようなシグナル検出結果の確認や、それを通したシグナル検出器の挙動把握をするための分析手法であるが、本検討ではこれをさらに副作用仮説の発見、すなわち「副作用の原因薬、治療薬、診療行為、傷病名の発見」に利用できないかの検討も行った。これは、スコアが増加するレセプト系列における固有の臨床イベントとして、そのような副作用仮説の組み合わせが抽出されるのではないかと期待に基づいた検討である。

一方、副作用仮説の発見という目的に対しては、副作用自発報告データからの「注目すべき医薬品と副作用の組み合わせ」の抽出等を目的に行われている、ROR (Reporting Odds Ratio)法やGPS (Gamma Poisson Shrinkage)法等を、今回検討したシグナル検出という文脈を離れて、レセプト系列に存在する全ての傷病と処方情報の組み合わせに対して直接的に適用するということも考えられる。今回はこれらの分析手法との比較検討を行ってはいないが、これらの分析は「注目する副作用と医薬品の組み合わせ」について、他の副作用と医薬品の組み合わせと比較して、偏って頻度が大きいものを抽出する手法であるため、これらの分析手法をレセプト系列に直接的に適用しても「原疾患の治療薬、診療行為、傷病名の組み合わせ」が主に抽出されるに過ぎないことが容易に推察される。その意味からは、本検討における「シグナルスコアを増加させる固有の組み合わせの抽出」という原理は、抽出すべき対象を副作用関連のものに絞ることを期待した方式であるともいえる。

特徴抽出のアルゴリズム、および検討の結果は第5章で述べる。



### 3. シグナル検出課題1： 副作用辞書に依存した副作用検出方法の検討

---

課題1においては、副作用辞書が与えられた場合に、ターゲット副作用のシグナルを精度よく検出する方法を検討する。

副作用辞書が与えられた場合、シグナル検出の方法として、辞書にマッチするレセプトデータを抽出する方法が考えられるが、単純な辞書とのマッチングでは精度よく副作用シグナルを検出することはできないと考えられる（また、予備評価を実施し、単純マッチによるシグナル検出ではノイズが多くなる事を確認した）。その理由としては、原因薬、治療薬、診療行為の順序や同期性の考慮が必要であることや、レセプト記載の傷病名を参考とするにしても、適切に臨床症状が反映されていない場合があることなどが挙げられる。また、転院等の辞書非記載の情報を参考にすることによって、さらに検出精度が向上すると考えられるので、上記のような単純マッチを今回の検討の結果とすることは不適當である。

そこで課題1においては、副作用発現が強く疑われる少数のレセプト系列例（学習用データ）を臨床的観点から人手で抽出し、その特性をマイニングにより発見・利用することによって、精度の高いシグナル検出方法を構成できるかを基本問題構造として設定する。なお、学習データのうち、副作用発現が強く疑われるレセプト系列を正例、副作用発現の疑いがない（低い）レセプト系列を負例とよぶ。

なお、

- ・臨床的観点による学習用データ抽出
- ・当該レセプト系列例を特徴付ける特性の同定
- ・同定した特性を組み合わせる精度よく検出を行う方式の構成

が、極端なマンパワーの使用や、特殊な高性能コンピュータなどに頼ることなく、数ヶ月程度の期間で完了できるかどうかを検討の主たるポイントとする。シグナル検出課題1については、これらの検討ポイントについてクリアできる可能性が確認できた段階でシグナル検出課題2へ進むこととする。

課題1の具体的な検討手順は以下の通りである。また、その概要を図3-1に概観する。

## 1. 副作用辞書の作成

PMDAにて、ターゲット（既に知られている副作用と原因薬）を設定、専門委員の意見等専門知識に基づき作成する。

## 2. レセプトデータ系列の構造化／可視化

特定患者のレセプト系列を直感的に概観可能にする。

下記3や6の目視や7の考察、4や5の設計を効率よく行うために必要となる。

## 3. マイニング用学習データの抽出

スクリーニングや目視により、副作用発現が強く疑われるレセプト系列例（正例）、副作用発現ではないと強く推量されるレセプト系列例（負例）を臨床的観点から人手で抽出

## 4. シグナル検出用レセプト系列特性の候補設計

基本特性列挙、スクリーニングルールの一般化を行う。

## 5. 学習データを特徴づける特性の同定とシグナル検出器の構築

機械学習技術による特性選択と検出器のパラメータのチューニングを行う。

## 6. 検出結果の評価

精度評価、検出例目視、検出方式の特徴（実験条件による結果の違い等）把握を行う。

## 7. 検討ポイント考察、手法の一般化および実用性に関する議論

ただし、上記の手順は順に一通り実行するだけでなく、6の目視結果を3～5にフィードバックすることで、繰り返し精度向上を行うこととする。

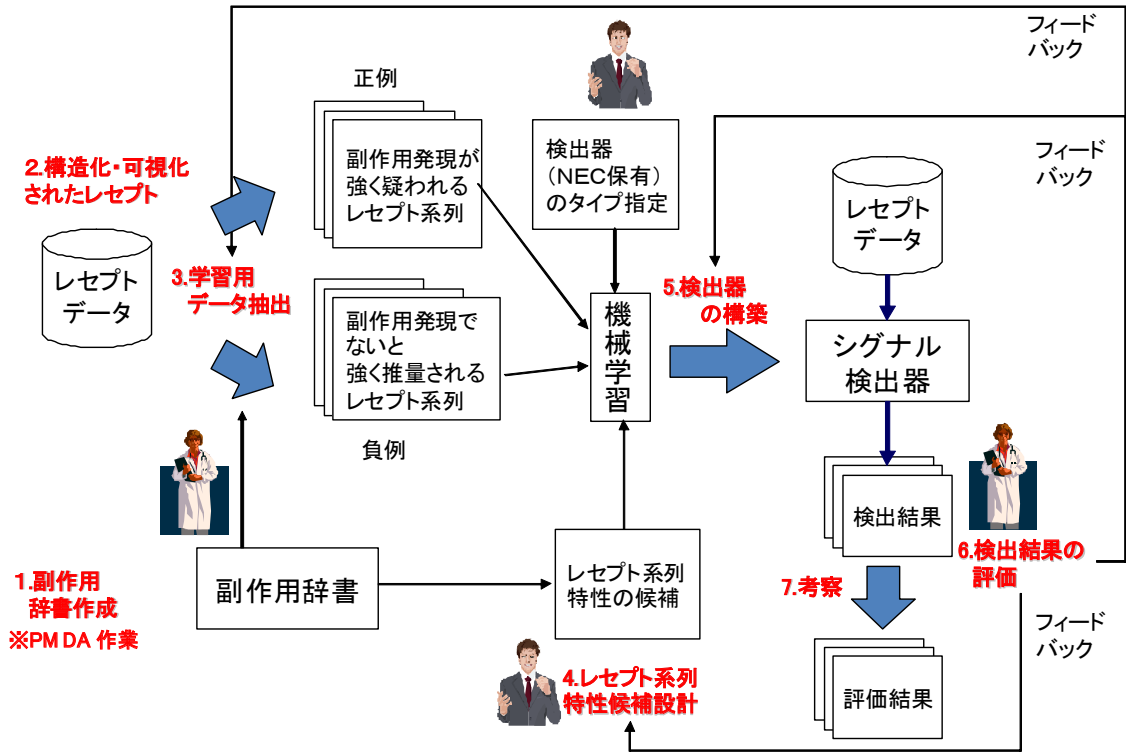


図3-1 課題1の検討手順

また、各検討過程におけるデータ処理と検討ポイントをまとめると、図3-2の通りとなる。以下、本章では課題1の各検討過程の内容を説明する。

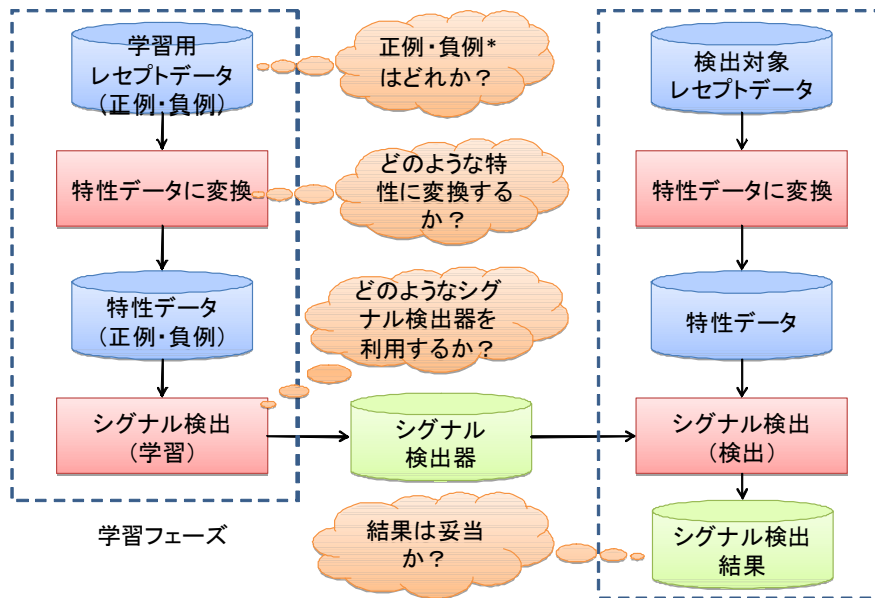


図3-2 課題1の検討過程における処理と検討ポイント

### 3-1 副作用辞書の作成

ターゲットとする、劇症肝炎・肝不全、間質性肺炎、間質性腎炎・急性腎不全、痙攣、意識障害、急性脳症の6種類の副作用に対し、副作用辞書をPMDAにて作成いただいた<sup>2</sup>。作成に当たっては、主に厚生労働省の事業により作成された「重篤副作用疾患別対応マニュアル」とPMDA専門委員である臨床医の意見が参考とされた。

これらの各副作用に対し、原因となる薬効群、治療薬、治療行為、副作用傷病名のそれぞれの項目をまとめたものが副作用辞書である（図3-3は副作用辞書の例）。学習用データの抽出等のためのレセプトデータとの各項目のマッチングは、それぞれ表3-1に掲げるレセプトデータのフィールドと一致するか否かで判別を行うものとした<sup>3</sup>。

表3-1 副作用辞書とマッチングをとるレセプトデータのフィールド

副作用辞書の項目	マッチングをとるレセプトデータのフィールド
原因薬	成分名と剤形大分類
治療薬	A T Cコードと剤形大分類
診療行為	標準化診療行為コード
副作用傷病	傷病名

<sup>2</sup> なお、検討を進める上で、シグナル検出課題1からシグナル検出課題2へ移行する際に副作用辞書を更新しており、両課題では副作用辞書が異なる。

<sup>3</sup> これ以外にも副作用辞書においては、原因薬にはA T Cコード、治療薬には成分名・一般名、診療行為には、標準化診療行為名・診療点数早見表・診療行為名・診療区分、副作用傷病にはI C D-10コードのフィールドが作成されたが、空欄の割合やどれくらい細かく区別するべきか等の観点から、マッチングするフィールドは表1の通りに決定した。

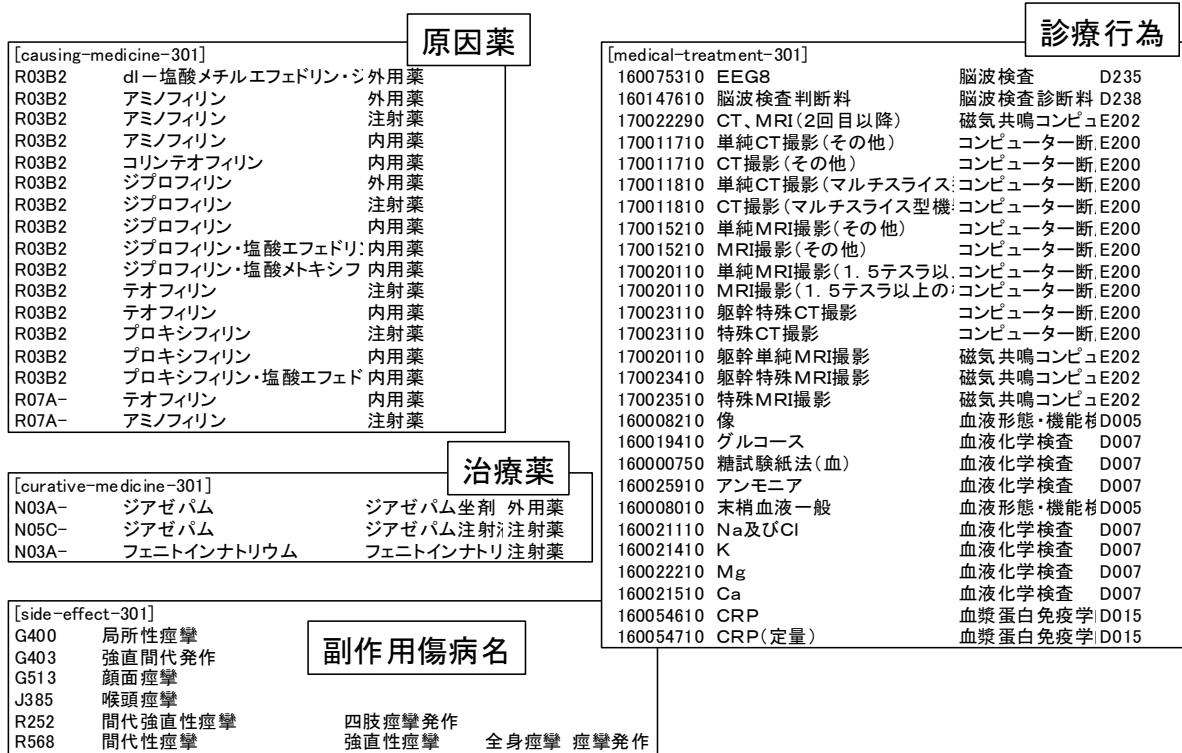


図 3-3 副作用辞書の例（痙攣の例を抜粋）

### 3-2 データ構造化／可視化

患者毎に医薬品、診療行為、傷病の発生を時系列で概観化した。各行がどのような医薬品、診療行為、傷病のレセプトがどの年月に発生したかを表しており、発生していなかった年月（以下、タイムスロット）では“-”（マイナス文字）、発生していた年月では副作用辞書にマッチする場合は“@”（アットマーク）、マッチしない場合は“+”（プラス文字）を表示した。医薬品や診療行為、傷病の内容は、名称や剤形、コードで表わし、同時に患者の性別・年代もみてとれるように可視化した（図 3-4 は可視化の例）。

このように、概観性を上げる可視化を行っておくことで、以降の検討手順における目視・考察、各種設計の効率を向上させた。

辞書	PID	2005	2006	2007	2008	種別	名称	剤形	コード	性別・年代
301A	50796	-----@-----				医薬品	テオフィリン	内用薬	R03B2	M_05-09_
301U	50796	-----@-----				医薬品	ジアゼパム	外用薬	N03A-	M_05-09_
301T	辞書記号			@	@	診療行為	CRP		160054610	M_05-09_
301T	A:原因薬			@	@	診療行為	末梢血液一般		160008010	M_05-09_
301T	U:治療薬			@	@	診療行為	Ca		160021510	M_05-09_
301T	T:治療行為			@	@	診療行為	グルコース		160019410	M_05-09_
301T	S:副作用傷病名			@	@	診療行為	Na及びCl		160021110	M_05-09_
301T	E:辞書マッチ終了			@	@	診療行為	K		160021410	M_05-09_
301t				@	@	診療行為	像		160008210	M_05-09_
301S	50796	-----@-----				傷病	痙攣発作		R568	M_05-09_
301E	50796									
	50796	-----+-----			++	医薬品	塩酸エプラシ	内用薬	R05C-	M_05-09_
	50796	-----+-----	+++	+++	+++	医薬品	レーカルボシ	内用薬	R05C-	M_05-09_
	50796	-----+-----	+++	+++	+++	医薬品	フマル酸ク	内用薬	R06A-	M_05-09_
	50796	-----+-----				医薬品	トラネキサム	内用薬	B02A1	M_05-09_
	50796	-----+-----			++	医薬品	セフテラムヒ	内用薬	J01D1	M_05-09_
	50796	-----+-----				医薬品	ロキタマイシ	内用薬	J01F-	M_05-09_
	50796	-----+-----				医薬品	ホスホマイシ	内用薬	J01X9	M_05-09_
	50796	-----+-----				医薬品	ドンペリドン	内用薬	A03F-	M_05-09_
	50796	-----+-----	++	++	++	医薬品	メキタジン	内用薬	R06A-	M_05-09_
	50796	-----+-----	+++	+++	+++	医薬品	ヒベンズ酸	内用薬	R05D1	M_05-09_
	50796	-----+-----				医薬品	リン酸オセル	内用薬	J05B4	M_05-09_
	50796	-----+-----			++	医薬品	クラリスロマ	内用薬	J01F-	M_05-09_

図3-4: レセプト系列可視化の例

### 3-3 マイニング用学習データの抽出

マイニング用学習データは、手順1) 副作用辞書にマッチするデータをPMDAが臨床的観点から目視確認し、以下のスクリーニングルールを作成、そのルールに従って副作用が疑われる患者を抽出、手順2) 当該患者のレセプト系列から年月タイムスロットごとに正例(副作用発現が強く疑われるレセプト系列例)、負例(副作用発現でないと強く推量されるレセプト系列例)を判定、という2段階の手順により行った。

手順1では、以下の四つのスクリーニングルール

- ルール①: 原因薬処方が中断、6ヶ月以上再開されず
- ルール②: ①の中断開始に同期(前後1ヶ月内)して該当治療薬発生
- ルール③: ①の中断開始に同期(前後1ヶ月内)して該当診療行為発生
- ルール④: ①の中断開始に同期(前後1ヶ月内)して、それ以前にない該当副作用傷病名が初めて発生

を作成し、①~④に合致する患者群を患者A(検討時はC1と定義)、①~③に合致し④に合致しない患者群を患者B(検討時はD1-C1と定義)とした。

手順2においては、患者A(C1)において上記①~④が成立しているタイムスロット群をAP(検討時はC1Pと定義)、そうでないタイムスロット群をAN(検討時はC1Nと定義)、患者B(D1-C1)において上記①~③が成立しているタイムスロット群をBP(検討時は(D1-C1)Pと定義)、そうでないタイムスロット群をBN(検討時は(D1-C1)Nと定義)と定義した。これらのタイムスロット群を目視したところ、AP(C1P)は正例が多数を占め、BP((D1-C1)P)は正例と

負例が混在、AN (C1N) とBN ((D1-C1)N) は負例が多数を占めることが観察されたので、正例をAP (C1P)、負例はAN (C1N) あるいはBN ((D1-C1)N) と暫定的に定義することとした。図3-5は全データ集合における正例・負例の範囲を図示したものである。

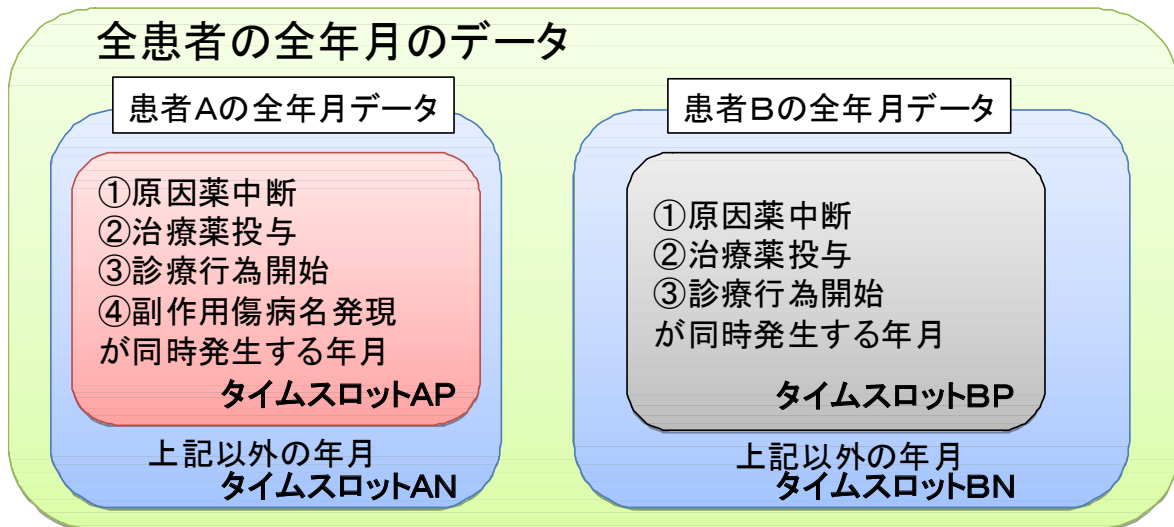


図3-5 全データ集合に対する正例負例の範囲

### 3-4 レセプト系列特性の候補設計

レセプト系列の特性として何を検出器に入力するかは、無数の候補が想定できるが、今回は「シグナル検出対象年月を中心とした12ヶ月において、原因薬、治療薬、診療行為の発生パターン毎の発生頻度」とする。12ヶ月間での発生パターンは全部で2の12乗 (= 4096) 通りあるが、それぞれ原因薬、治療薬、診療行為毎に頻度を集計した合計3×2の12乗 = 12288個の変数が、検出器に入力すべき変数の候補とする(次章「3-5 特性の同定と検出器の構築」において、これらの中からシグナル検出に有効な変数が機械学習技術によって選択される)。なお、上記で12ヶ月とされている発生パターンの長さについては6ヶ月から12ヶ月まで条件を変えながら同様の実験を行い評価したが、シグナル検出課題1においては、精度が最も高くパターンとしての表現力も十分に大きい12ヶ月を採用し、その結果を報告する<sup>4</sup>。

<sup>4</sup> なお、シグナル検出課題2では、パターンの長さを3ヶ月や6ヶ月へ変更する。これは、シグナル検出課題2では正例の数が減った事や、シグナル検出課題2では辞書の情報を利用する事が難しいため、4096通りのパターンが多くなりすぎるためである。(これは逆に、解くべき問題の性質に応じてパターンの長さを変えるべきという事を意味する。精度を追及する場合には、パターンの長さを交差検定法等によってデータから調整することが望ましい)

図3-6を用いて、上記の特性候補生成の様子を説明する。実際には12ヶ月における発生パターンを本検討では用いているが、説明が煩雑になるため、ここでは3ヶ月における発生パターンの例を説明する。

今回の検討に用いているレセプトデータは48ヶ月分であり、各年月において、ある原因薬や治療薬に関するレセプトが特定の患者に対して発生しているかは、「発生」を1、「発生せず」を0で表すと、1か0を48個並べたもの、すなわち48桁のビットパターンとなる。シグナル検出器に入力されるデータは、上記48桁のビットパターンから、シグナル検出の対象としているタイムスロットを中心とした12ビットを取り出して（図3-6の「ある患者のレセプトデータ」の部分。図では3ヶ月分の3ビットを取り出すとして説明）、頻度を原因薬、治療薬、診療行為ごと<sup>5</sup>に集計したものである。すなわち、3ビットのビットパターンは全部で8通り（図3-6の「全発生パターン」の部分）あるが、それぞれについて該当する原因薬の数、治療薬の数、診療行為の数を集計し（図3-6の「特性データの候補」の部分）、当該患者の当該タイムスロットの特性データ候補とする。

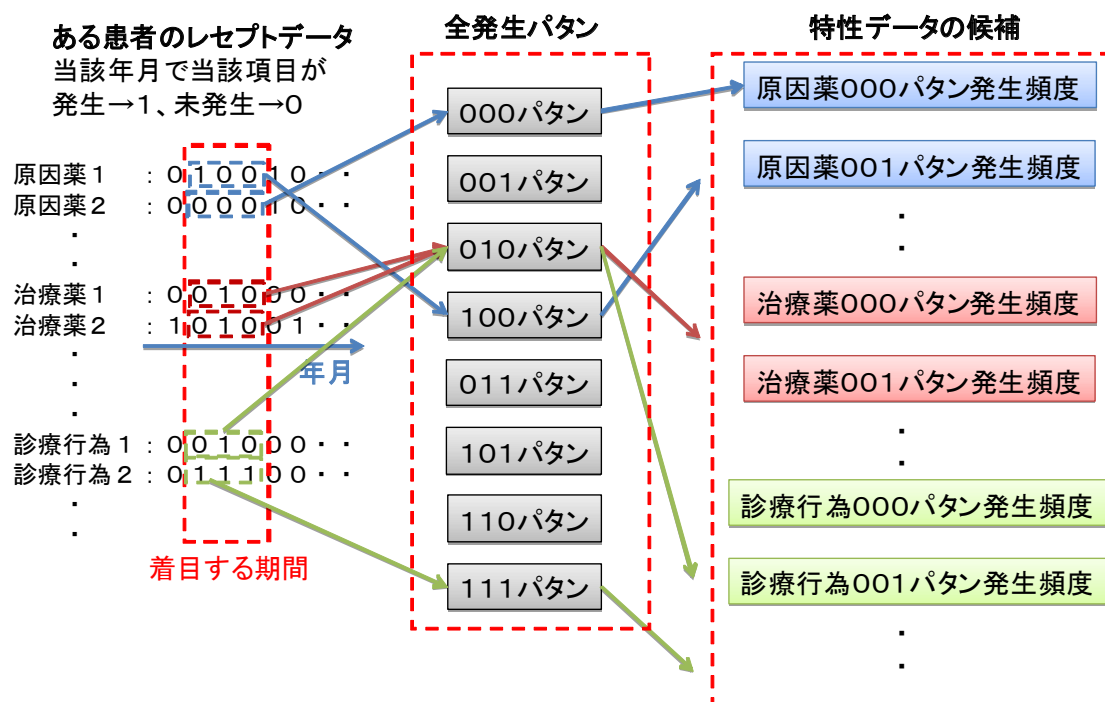


図3-6 レセプト系列特性データの候補生成

<sup>5</sup>副作用傷病名も重要な特性であるが、正例の作成時に副作用傷病名の情報を利用してゐるため、本検討では利用しない方針とした。



### 3-5 特性の同定と検出器の構築

前節で構成した、副作用シグナルの検出対象年月に対する12288個の系列特性候補から、入力変数としてシグナル検出に有効なものの選択と、当該年月（期間の中心）における副作用発現の疑いの強さをスコア（副作用シグナルスコア）として出力する「シグナル検出器」の構成を、機械学習技術により実施した。

今回の検討においては、NEC保有のシグナル検出器を用いた。本検出器は多数の調整パラメータ（加重値）を持ち、色々なシグナル検出基準を実現することが可能な柔軟な構造を有しているものであり、図3-7にその概要を掲げる（仕様詳細については付録を参照）。入力された変数に対してそれぞれ指定された加重値が乗ぜられ、それらをすべて合算したものをベースにスコアが計算されて出力されるという構造をしている。

この検出器に対し、12288個の候補の中のどれを入力データとして与えるべきか、それぞれの加重値をいくつにするべきか、さらに若干の挙動調整のパラメータをいくつにするべきかを、前々節で構成した正例・負例の両データから機械学習技術を用いて自動調整した。

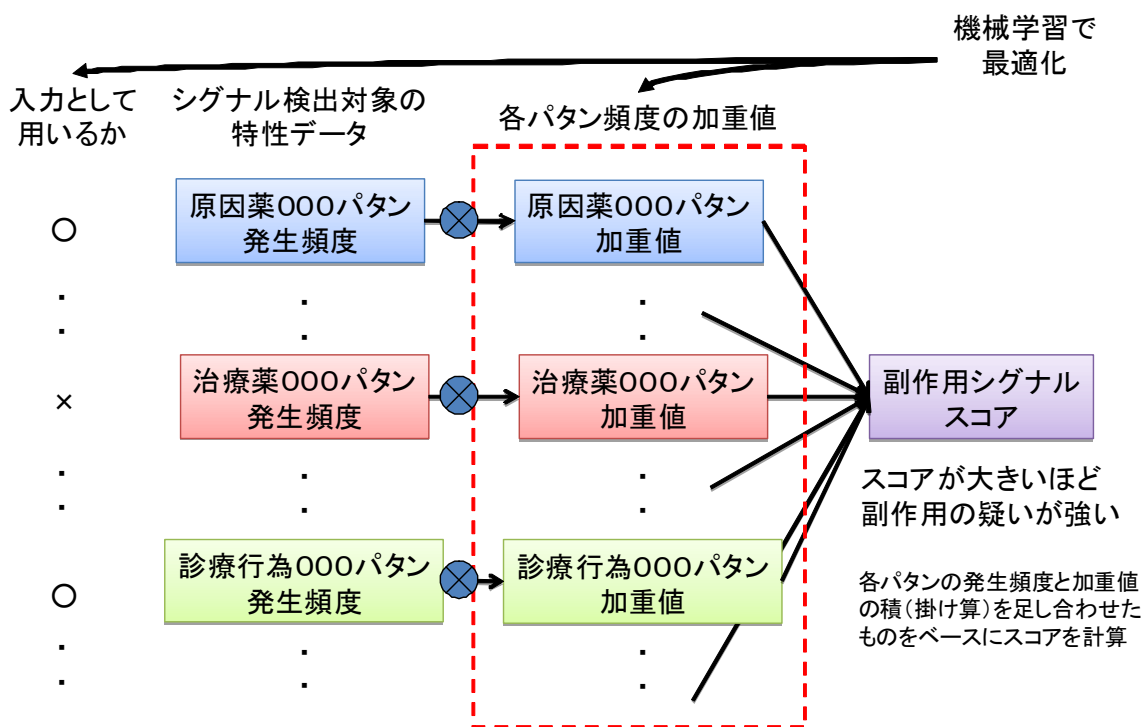


図3-7 シグナル検出器の仕様概要

### 3-6 シグナル検出結果の評価

シグナル検出の実験は、以下のように条件を変えて二通り行った(図3-8参照)。

**実験1**：正例＝タイムスロットAP (C1P)、負例＝タイムスロットAN (C1N)

**実験2**：正例＝タイムスロットAP (C1P)、負例＝タイムスロットBN (D1-C1) N)

実験1は同一患者群における副作用シグナル有(正例)と副作用シグナル無(負例)の判別をしているのに対し、実験2は異なる患者群における副作用有(正例)と副作用無(負例)の判別をしていることになる。

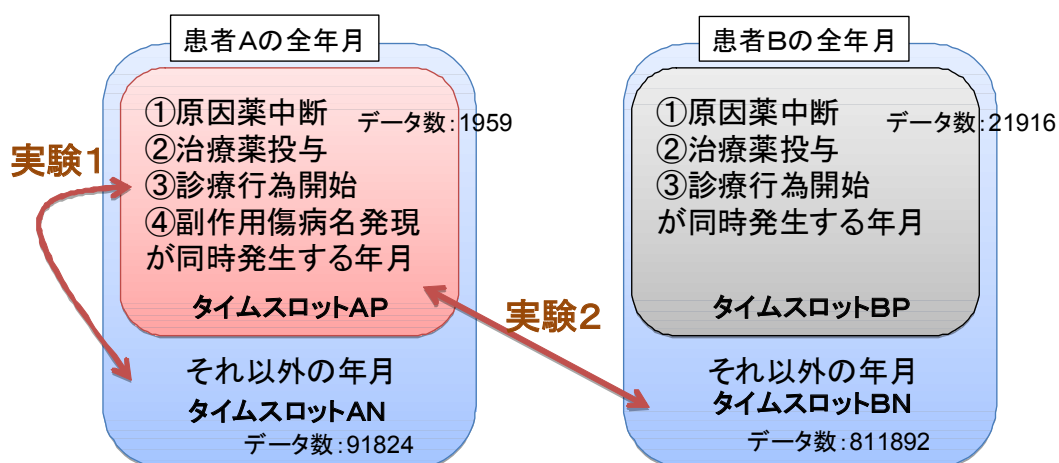


図3-8 評価実験1と実験2の学習データの範囲

それぞれの実験においては、正例データと負例データを学習用と評価用に分け、学習用データでシグナル検出器を構成し、評価用データに対して正しくシグナル検出が行われるかをみるクロスバリデーション(交差検定)によって、シグナル検出の精度評価を行った。特に本検討においては「検出器によって出力された副作用疑いの強さ(スコア)が大きいデータから、運用者が確認していく」という運用体制を想定して、スコアの上位何位まで確認すると、そのなかに副作用シグナル有のデータ(正例)が何件含まれるか、シグナル無データ(負例)が何件含まれるかという観点から評価を行った。すなわち、様々な分野で精度の評価によく用いられている基準を、スコアの上位X件のデータまで計算することによってシグナル検出器の評価を行うこととした。具体的には以下の3基準である。

**Positive Predictive Value (陽性的中率、PPV)**

計算式： スコア上位X位までの正例数 / X

**Sensitivity (感度)、True Positive Rate**

計算式： スコア上位X位までの正例数/スコア最下位までの正例数

**False Positive Rate (偽陽性率、FPR)、Specificity (特異度)**

計算式： スコア上位X位までの負例数 (誤報) / スコア最下位までの負例数

Positive Predictive Value は上位X位まで確認したときの精度、Sensitivityは検出の網羅性、False Positive Rateは誤報の多さを表している。

例えば、図3-9の仮想的なシグナル検出の出力(スコアの大きい順にソート)に対しては、第3位までの確認(X=3)で、 $PPV = 2/3 = 0.666$ 、 $Sensitivity = 2/4 = 0.500$ 、 $FPR = 1/3 = 0.333$ となり、第5位までの確認(X=5)で、 $PPV = 4/5 = 0.800$ 、 $Sensitivity = 4/4 = 1.000$ 、 $FPR = 1/3 = 0.333$ と計算されることになる。

副作用検出結果 (全データ数=7の場合の例の説明\*)

順位	PID, TIME	スコア	副作用
1位	12345, 2000/08	0.995	正例
2位	12121, 2001/10	0.992	負例
3位	3355, 2000/05	0.854	正例
4位	65432, 2002/04	0.841	正例
5位	12345, 2003/06	0.800	正例
6位	99999, 2000/05	0.741	負例
7位	22532, 2004/01	0.521	負例

\*実際のデータ数は、患者数×時刻数(48か月)

図3-9 シグナル検出器の出力例 (本実験の結果ではなく仮想例による説明)

副作用検出という目的においては、シグナル検出器のスコア上位は全てが正例で、下位は全てが負例というのが最も望ましいが、そのような完全分離ができない場合はPPV-Sensitivity、Sensitivity-FPRがそれぞれトレードオフの関係になる。よって、ここでは両者を対比した評価を実施することとした。

図3-10は実験1の結果である。赤い線がPPV-Sensitivity曲線、青い線がROC曲線を示している。例えば、Sensitivity=0.8でPPV=0.9を達成していることや、ROC曲線の立ち上がり非常快、FPRが極めて小さい場合でもSensitivityがほぼ1に近い値になっていることからスコア上位はほぼ全て正例で占められていることなどが見て取れる。また、図3-11は同実験の結果において、スコア上位何位までに正例が何件含まれていたかを表したグラフである。これにおいても、スコア上位1000件程度まで、ほぼ全てが正例で占められていることが見て取れた。これらの結果から、同一患者群内での判別に相当する実験1においては、マイニングによって高精度なシグナル検出が可能であることが観察された。

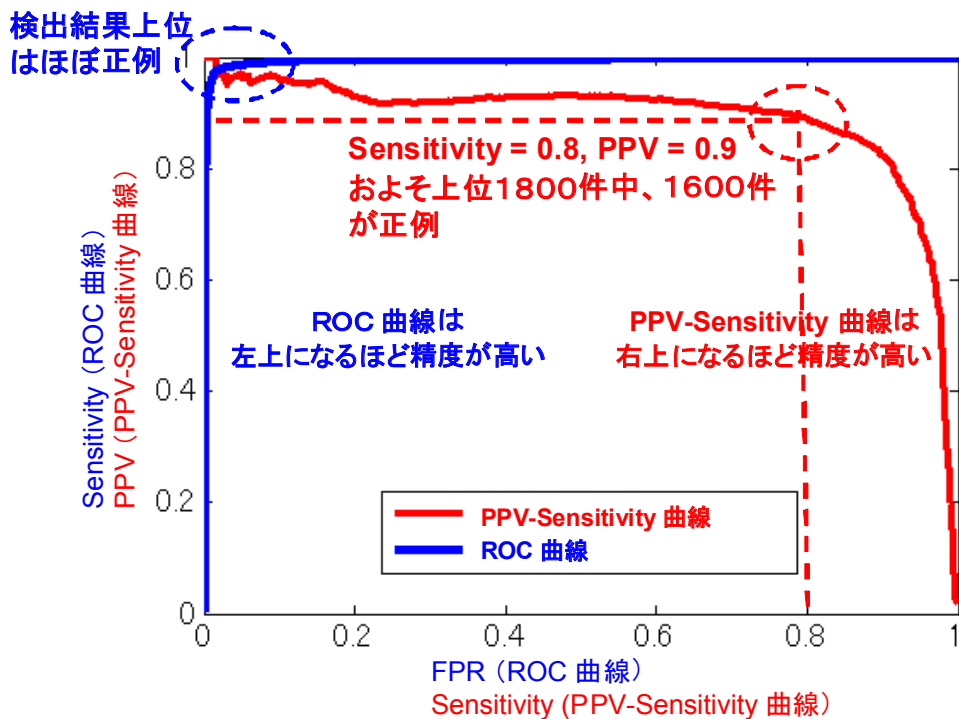


図3-10 実験1の結果 (PPV-Sensitivity 曲線、ROC 曲線)

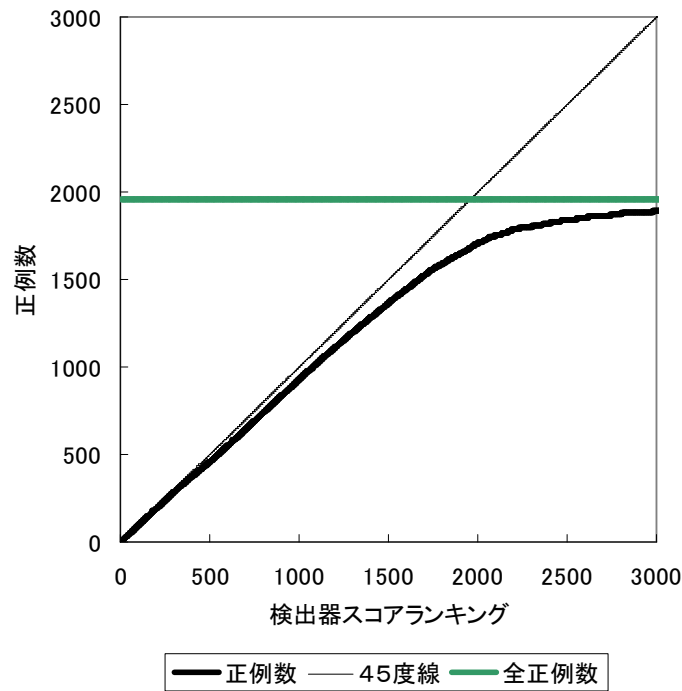


図3-1-1 実験1の結果（正例数－ランキング曲線）

図3-1-2は実験2の結果である。実験2においては、スコア上位200件に正例が94件含まれているが、スコア最上位50件には13件の正例しか含まれていなかった。図では見えにくいだが破線楕円で囲った部分に相当する。図3-1-3は同実験の結果において、スコア上位何位までに正例が何件含まれていたかを表したグラフである。スコア上位1500件程度で正例が1000件強程度含まれていることが見て取れる。これらの結果から、最上位の精度には課題があるものの、異なる患者群の間での判別に相当する実験2においても、マイニングによって一定の精度でのシグナル検出が可能であることが観察された。

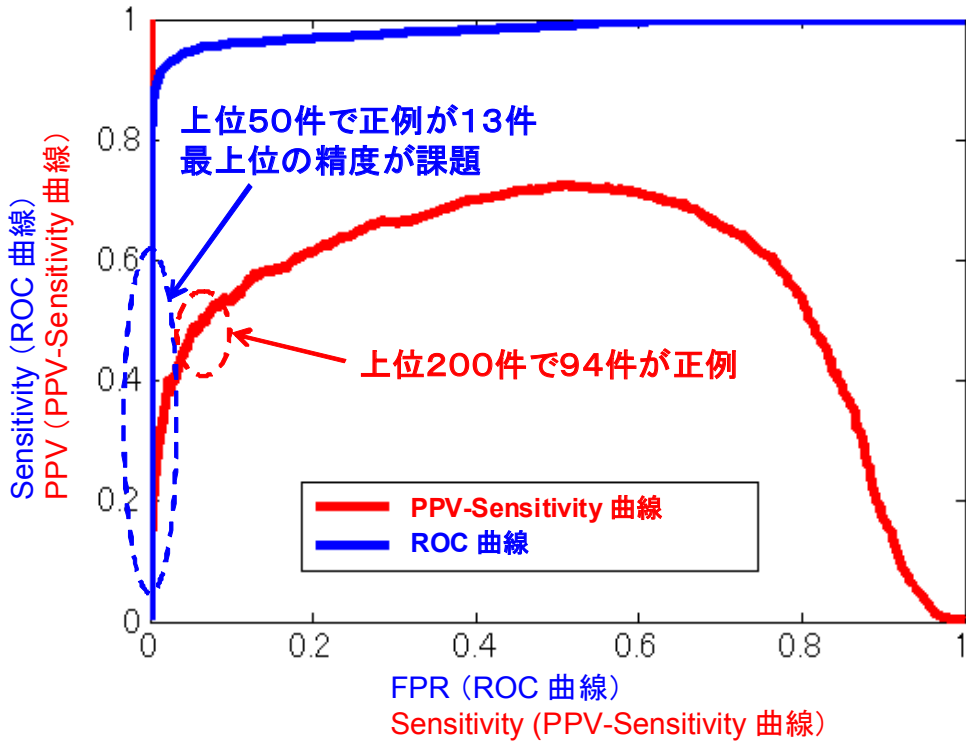


図3-12 実験2の結果 (PPV-Sensitivity 曲線, ROC 曲線)

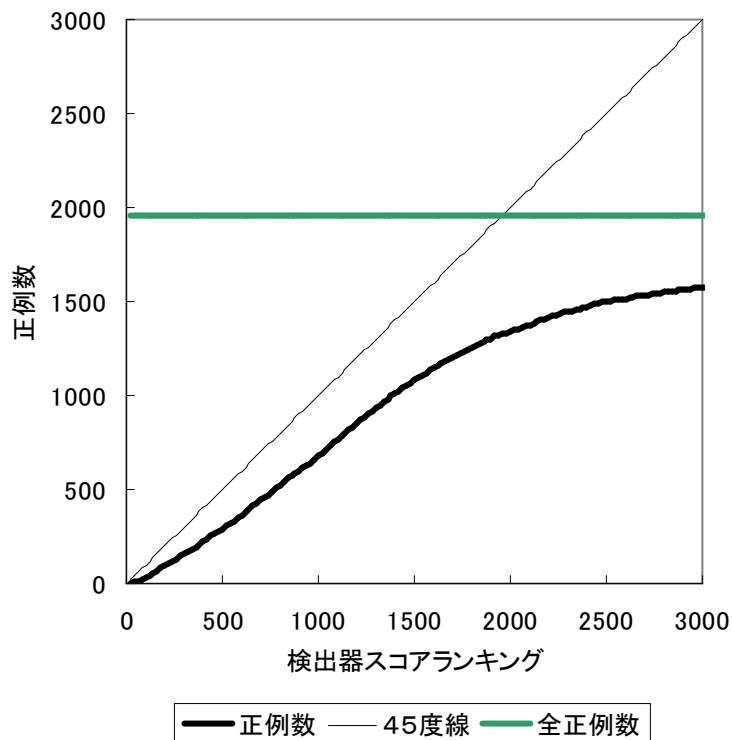


図3-13 実験2の結果 (正例数-ランキング)



「副作用辞書依存の副作用検出」の可能性を確認し、また副作用検出に有効な特性に関する知見が得られたことから、シグナル検出課題2へ進むこととした。一方、シグナル検出課題1についても、実運用における有効性判断のためにはさらなる検討が必要である。今後の検討ポイントとしては

- ・ 今回の結果の詳細分析と検討前段へのフィードバック
- ・ 正例・負例の定義の見直し、特性候補タイプの拡大、他のタイプの検出器の検討
- ・ 対象患者の拡大可能性（患者A、B以外でのシグナル検出）
- ・ 規模拡大した場合の、人的・計算機的負荷の見積もり

等が挙げられる。

#### **4. シグナル検出課題2： 副作用辞書に非記載の副作用検出方法の検討**

課題2においては、副作用辞書に記載されていない副作用のシグナルを網羅的に検出する方法を検討する。課題2の背景と問題構造、検討手順は以下の通りとなる。

課題1の副作用辞書に依存した副作用検出は、一定の精度で実現可能であることを3章で示したが、全副作用をターゲットにして網羅的に副作用辞書を作成するのは、膨大な費用・時間が必要となる。また、未知の副作用に関しては、副作用辞書を作成すること自体が原理的に不可能である。これらの理由から課題2で検討する副作用辞書に依存しないシグナル検出方法を構築する必要性は大きい。

そこで課題2においては、副作用シグナルを有するレセプト系列同士が（副作用の種類に依らず）何らかの観点での共通性質を持つと仮定し、課題1の検討等も含めてそれらを抽出・利用することによって、精度の高いシグナル検出方法を構成できるかを基本問題構造として設定する。特に

- ・ 副作用の種類に依存しない、副作用シグナル発現が疑われるレセプト系列の共通特性の同定
- ・ 同定した特性を組み合わせることで精度よく検出を行う方式の構成

が、極端なマンパワーの投入や、特殊な高性能コンピュータなどに頼ることなく、数ヶ月程度の期間で完了できるかが主たる検討ポイントになる。

課題2の具体的な検討手順は以下の通りである。また、その概要を図4-1に概観する。

##### **1. 副作用シグナル有のレセプト特性の一般化**

課題1の結果から副作用シグナル検出に有効だった特性を一般化する。

##### **2. レセプト系列特性の候補追加**



上記1だけでは必要な特性が足りないと想定されるので、臨床的観点による専門知識等から追加して特性を作成する。

### 3. 利用する特性の同定と検出器の構築

検出器の候補設計、機械学習技術による特性選択とパラメータ調整を行う。

### 4. 検出結果の評価

精度評価、検出例目視、検出方式の特徴（実験条件による結果の違いや、特に誤検出が大きくなる条件等）把握を行う。

### 5. 考察

課題1と同様、上記の手順は順に一通り実行するだけではなく、4の目視結果を1～3にフィードバックすることで、繰り返し精度向上を行うこととする。

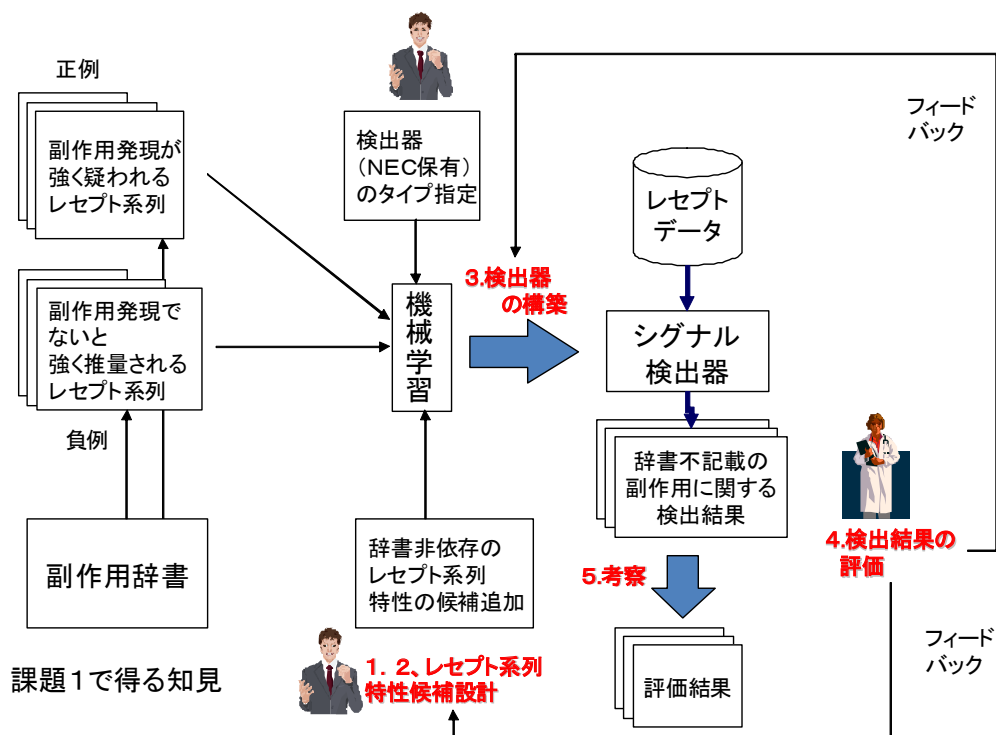


図4-1 課題2の検討手順

#### 4-1 マイニング用学習データの追加と再定義

マイニング用の学習データは、課題1と同様に課題2においても、

正例： 副作用発現が強く示唆されるレセプト系列例

負例： 副作用発現でないと強く推量されるレセプト系列例

である。課題2においては、学習用データ数の増加（および正例の精度向上）を目的とし、下記の3パタンの方法で学習用データを作成した。

##### パターン1： スクリーニングルールによる抽出 + PMDAによる目視

課題1の6種類の副作用に加え、脳梗塞、心不全、低血糖、血小板減少等副作用を拡大した副作用辞書を用いて、PMDA設計の

ルール① 3ヶ月以上の原因薬連続処方後に、6ヶ月以上連続して処方がなく

ルール② 原因薬中止月あるいは翌月に、治療薬の処方および診療行為が発生し

ルール③ ルール②にマッチする最初の診療行為の後、3ヶ月以内に再度診療行為が発生し

ルール④ 原因薬中止日より前6ヶ月間に、治療薬処方が発生しない

によりスクリーニングを行った。さらに、スクリーニングルールに合致する患者のレセプト系列に対してPMDAが目視を行い、合致した年月を正例として採用するかどうかの判断を行った。

その結果、正例のレセプト系列（正例とする「患者-タイムスロット」の組み合わせ）の数は57件となった。負例の定義は実験によって変えているので、対応する章で述べる。

##### パターン2： マイニング以外の試行調査で使用了副作用イベント

PMDAにより、マイニング以外の方法で試行調査された、

- ・NSAIDsによる光線過敏症
- ・ステロイドによる骨粗鬆症
- ・抗精神病薬による薬剤性パーキンソンニズム
- ・チアマゾールによる無顆粒球症
- ・ $\alpha$ -グルコシダーゼ阻害薬による肝障害

5通りに関して、特定された「患者-タイムスロット」の組み合わせを正例レセプト系列とし、当該患者で正例でないタイムスロットのレセプト系列を負例とした。正例数は

169件となった。こちらも負例の定義は実験によって変えているので、対応する章で述べる。

### パターン3： 「薬剤性\*\*\*」のように薬剤による副作用が明らかなイベント

薬剤性甲状腺機能障害や薬剤性溶血性貧血といった、薬剤による副作用が明らかな52タイプの傷病名をPMDAにより指定いただき、それを有する患者において当該傷病が初出するタイムスロットを正例とした。負例は当該患者の正例でないタイムスロットである。正例数は4711件となった。負例の定義は各実験の章で述べる。

課題2においても課題1と同様に、副作用の種類によって正例の区分けを行っている。パターン1であれば、副作用辞書に記載されている区分、パターン2では上記の5通り、パターン3では当該52タイプの区分けとなっている。また、同一患者の同一タイムスロットが、複数の正例区分けに該当する例も存在していたため、区分け別の正例数の総和は全正例数の総和と一致しない。

## 4-2 レセプト系列特性の一般化

---

シグナル検出課題1では、図3-6に示されるように、副作用辞書の情報に基づいた、原因薬、治療薬、診療行為の発生パターンを特性データとして利用し、図3-10や図3-12に示されるように、一定の精度で副作用シグナルを検出可能である事を示した。このアプローチでは、あらかじめ「どこに注目すればよいのか」という臨床的な専門領域知識に基づいた情報が副作用辞書情報から利用可能である点が、シグナル検出を可能とした一つの主要因と考えられる。

シグナル検出課題2では、副作用辞書に記載されていない副作用のシグナルを検出する事が目的のため、副作用辞書に記載された特定の副作用と紐づいた原因薬や治療薬などの分類を利用する事はできない（または、利用した場合にはその原因薬以外で発現する副作用は検出する事ができない）。

そこで、シグナル検出課題1で定義した発生パターンを以下の方法で一般化した。

### ・ 医薬品発生パターンの一般化

課題1においては副作用と紐づいている医薬品を原因薬及び治療薬に分類し特性データとして用いていたが、課題2においてはこの分類は使用せず、代わりに、「非除外原因薬」と「除外原因薬」の二つに分類しその発生パターンを特性とした。すなわち、外用薬や輸液のように、原因薬としては考慮の対象から外してもよいと思われるものを「除外原因薬」としてPMDAにご指定いただいた（ATCコードと剤形大分類の組み合わせで92種類）。これにより、医薬品から作成する特性データに対し「除外原因薬」と「非除外原因薬」の区別を与えた（除外原因薬も特性データ化の対象とした）。なお、

同様のフレームワークとして治療薬についても除外と非除外の区別をつける事が可能であるが、原因薬ほど区別が容易ではない点と、区別作業にかかる人的負荷軽減のため、本検討では治療薬の除外・非除外は扱わない方針とした。

- ・ **診療行為発生パタンの一般化**

医薬品と同様に、診療行為についても副作用との紐づきを利用する事はできないため、初診や処方料のような、シグナル検出をするための入力情報として不適切と思われるものを「除外診療行為」としてPMDAにご指定いただいた。医薬品の場合には除外原因薬が治療薬となる可能性があり、除外原因薬も特性データとして利用したが、診療行為の場合には除外診療行為は副作用とは関連がないと考えられるため、特性データとしては扱わない方針とした。除外診療行為は、標準化診療コードの頭3桁が特定の値を持つものや、名称に「加算」等の文字列が含まれるもの、および、それらの例外ルール等で指定される。それ以外の診療行為、すなわち特性データ化した診療行為を「非除外診療行為」と述べることにする。

#### 4-3 レセプト系列特性の追加

---

前節のとおりレセプト系列特性の一般化を行ったのに加え、課題2においては課題1で用いなかった特性の追加も行った。

追加した特性は、

- ・ 患者の年代
- ・ 当該年月の入院医科レセプト発生の有無
- ・ 入院外医科レセプトの発生の有無、（診療科不明以外の）診療科<sup>6</sup>
- ・ 傷病名

である。

ただし、患者の年代は2008年1月時点での年齢を、0歳～4歳から65歳～69歳までの5歳きざみの階層化および70歳以上としたものである。なお、傷病名については、シグナル検出課題1では副作用辞書との情報の重複が大きいため利用しなかったが、シグナル検出課題1では副作用辞書を利用しないため、特性データとして利用する方針とした。

---

<sup>6</sup> 内科、外科、小児科、放射線科、整形外科、歯科・口内外科、泌尿器科、産婦人科、皮膚科、眼科、精神・神経科、耳鼻咽喉科、脳神経外科、麻酔科

上記をまとめると、課題2において特性データ化される情報は、除外原因薬、非除外原因薬、非除外診療行為、傷病名、年代、入院医科レセプト発生、入院外医科レセプト発生、診療科（内訳は脚注5を参照）となる。図3-6と同様に、これらの情報に関する発生パターンを特性データとして利用した。

#### 4-4 シグナル検出課題2の評価方法

課題2では、「副作用辞書に記載されていない副作用」を検出する事が目的のため、副作用辞書をベースに正例を作成した本検討で、学習されたシグナル検出器の性能を厳密に評価する事はできない。そこで、本検討では「副作用辞書に記載されていない=学習データに含まれていない」と仮定し、図4-2に示される手順によって、課題2に関するシグナル検出器の評価を行った。

図4-2の手順では、まず副作用辞書に記載されている副作用イベント数に負例集合を分割する。次に、ある副作用イベントに該当する正例と、分割された負例集合の一つを選択し評価データとし（図中では赤にハイライトされたデータ）、残り（それ以外の正例と負例。図中では青でハイライトされたデータ）を学習データとする。このようにすると、学習データから評価データとしてマスクされた正例に関する副作用イベントがなくなり、学習していない副作用イベントを検出できるかを、評価する事が可能となる。最後に、各副作用イベントに対してマスクした評価を実行したら、全ての評価結果をシグナルスコアによってランキングし、精度を算出する。

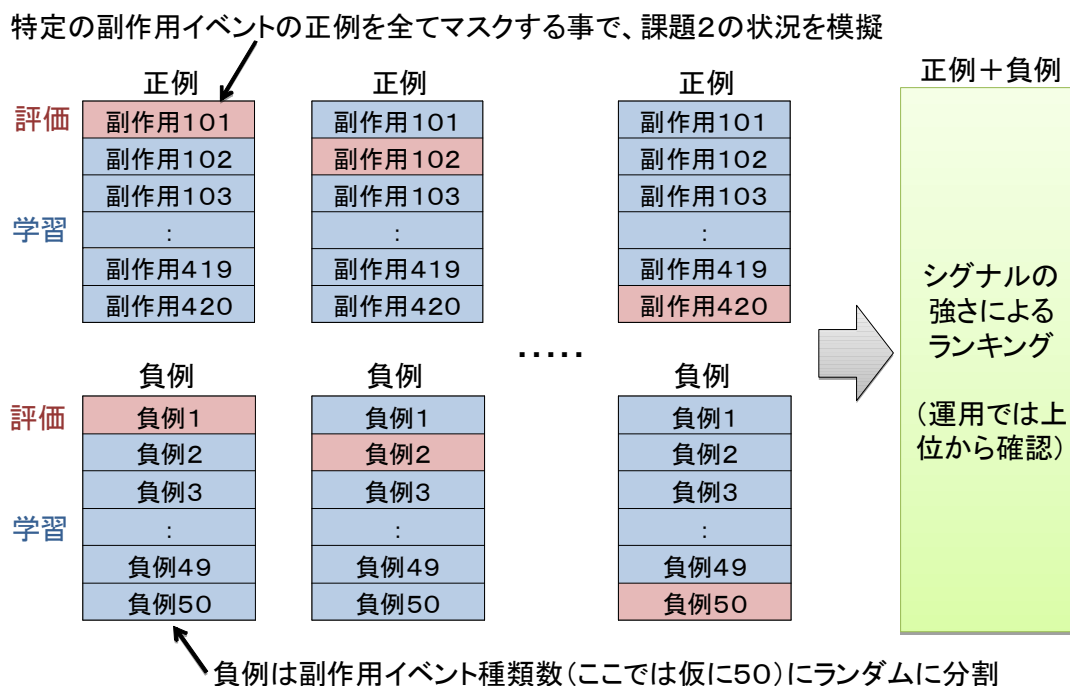


図4-2 課題2の精度評価方法

#### 4-5 シグナル検出結果の評価1 (課題1と課題2の比較)

---

課題2の評価として、まず課題1と課題2で、検出精度にどの程度の差があるのかを評価した。なお、本評価実験では、課題1と同様にパタン1に関する正例と負例を用いて評価を実施した(ただし、正例の作り方を4-1節のように修正したため、結果は3-6節とは異なる)。この実験では、正例数が57件、負例数が2775件で、正例数の割合は約2%である。また、特性データ生成の際の時系列長については、課題1よりも正例数が少なくなったため、12ヶ月から6ヶ月に変更とした。

図4-3に課題1と課題2の比較結果を示す。まず、PPV-Sensitivity曲線(上図)を観察すると、課題1ではSensitivity=0.5でもPPV=0.7程度であり、3-6節の結果と同様に正例数が少数であっても、高精度に辞書記載の副作用を検出できている事が確認できる。一方で、課題2では課題1と比較して同一のSensitivityに対するPPVは5分の1程度である。

次に、検出順位上位での正解数(下図)を観察すると、上位100位までで課題1では正解数が約40、課題2正解数が約8、上位300位までで課題1は正解数が約50、課題2は正解数が約20であった。課題1は上位300位で正例57件中50件が検出されている(Sensitivity=0.878, PPV=0.167)のに対して、課題2では57件中22件の検出(Sensitivity=0.386, PPV=0.073)で課題1と比較した精度低下が顕著である事が確認できる。一方で、課題2の場合であってもシグナル検出器を利用せずに、ランダムサンプリングで300件を調査した場合には、正例の数は約2%程度なので5~6件程度しか検出する事ができないため、精度は十分ではないが課題2の場合にもデータマイニングに基づくシグナル検出が一定の効果がある事が確認できる。

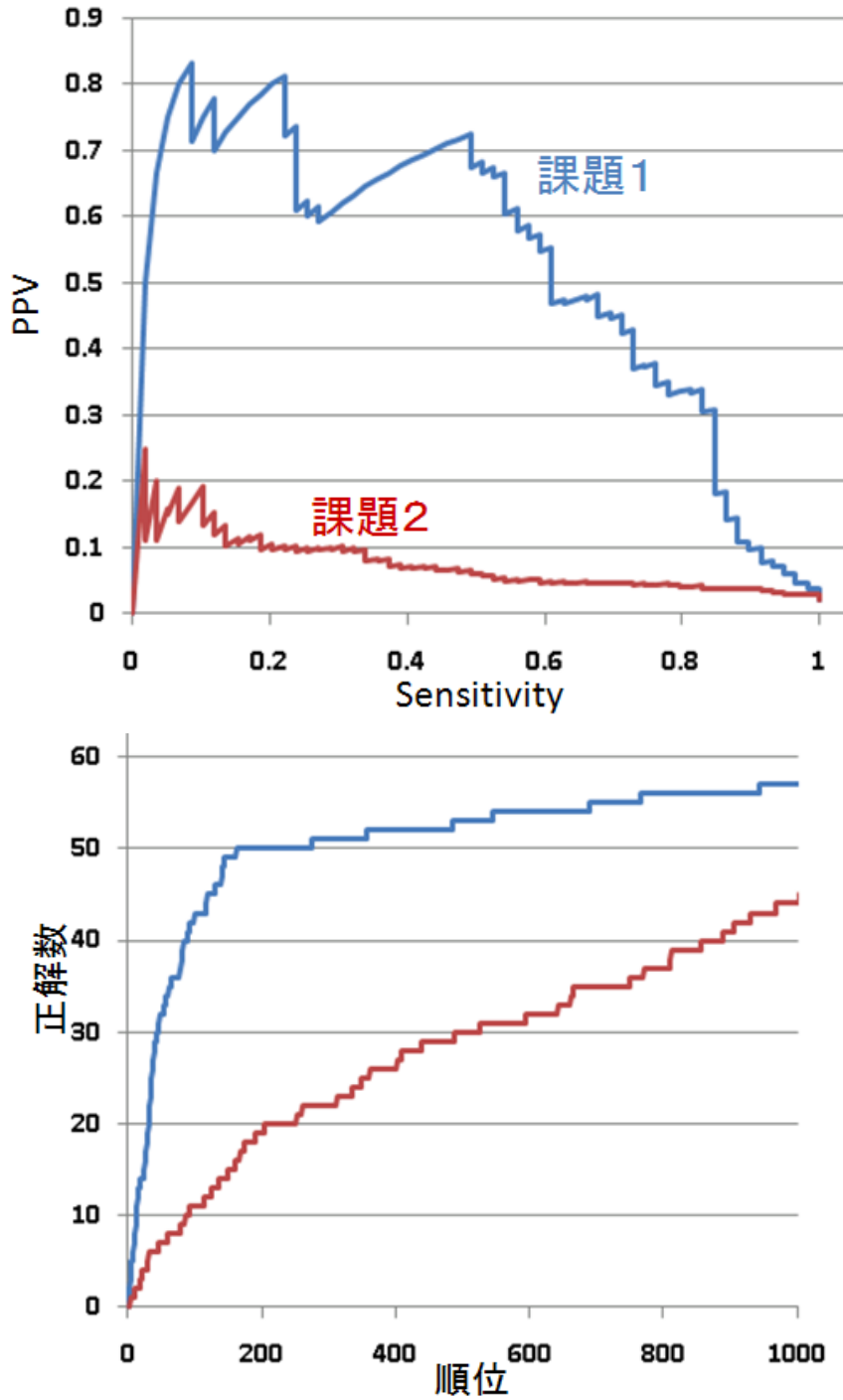


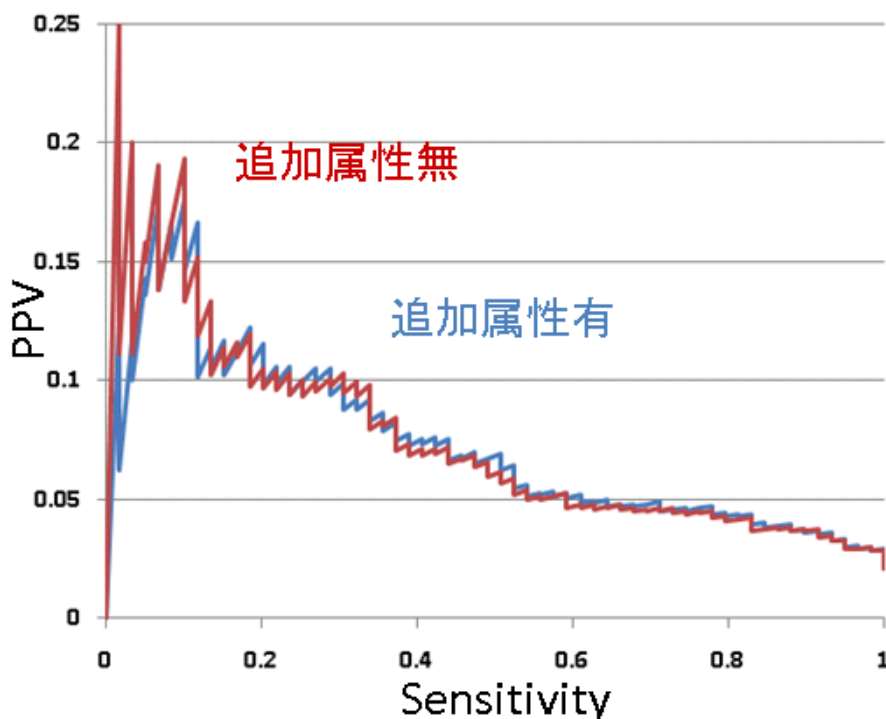
図4-3 課題1と課題2の精度比較

(上図：PPV-Sensitivity 曲線、下図：検出順位対する正解（正例）数)

#### 4-6 シグナル検出結果の評価2（特性の追加）

本評価では、医薬品、診療行為に加えて、傷病名、年代、入院医科レセプト発生、入院外医科レセプト発生、診療科の特性を追加した場合と、追加していない場合について精度の比較を実施し、特性を追加する事によってシグナル検出精度が向上するのかを評価した。この実験では、4-5での評価と同様に正例数が57件、負例数が2775件、正例数の割合は約2%である。また、特性データ生成の際の時系列長についても同様に6ヶ月とした。

図4-4にPPV-Sensitivity曲線（上図）と検出順位-正解数曲線（下図）を示す。特性の追加によって、1位から300位での検出精度にはほとんど変化はなく、300位から1000位の間では特性の追加によって若干の精度向上が観察された（件数として57件の正例に対して特性を追加した方が、正解数が1、2件多い。）統計的に有意な程の大きな差ではないが、適切な特性を追加する事でシグナル検出精度が向上する可能性がある事が示唆される。また、一般的には特性データを増やす事によってシグナル検出器が検出可能な副作用の種類が広がる事が期待され、評価データに対して同程度の精度であれば、将来の拡張を考慮し検出力の高い属性を利用する方がよいと考えられる。ただし、シグナル検出器を学習する観点からは、データ数が少数の際に特性の数が多すぎると、学習データに検出器が適合しすぎて汎用性が失われる「過学習」と呼ばれる現象が発生するため、学習データの数（特に本検討においては数が少ない正例の数）に応じて特性データを増加していく事が望ましい。





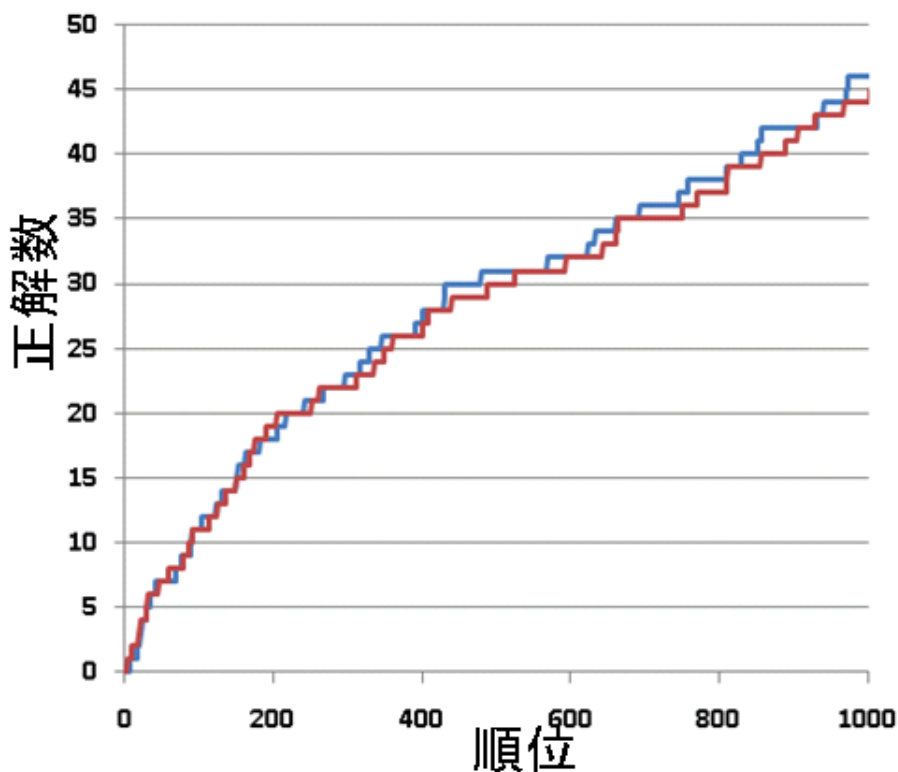


図4-4 追加属性の有無によるシグナル検出精度の比較

(上図：PPV-Sensitivity曲線、下図：検出順位対する正解（正例）数）

#### 4-7 シグナル検出結果の評価3（負例数増加）

本評価では、副作用パターン1について、正例を57件とし、4-5、4-6と同様に負例を2775件とした場合（正例の割合は約2%）と、PMDAが設定したスクリーニングルールによって抽出した負例も評価に利用し、負例を24618件とした場合（正例の割合は約0.23%）を比較し、負例が増加した場合にシグナル検出精度がどのように変化するかを観察した。一般的には、負例が増加すると相対的に正例の検出が難しくなるため、本評価はその傾向を調べることを目的としている。

図4-5にPPV-Sensitivity曲線（上図）と検出順位-正解数曲線（下図）を示す。負例追加無と負例追加有の場合では、前者の方が検出精度が高く、負例の割合が増加する事で相対的に正例の検出が難しくなる事が確認できる。一方、正例の割合が約十分の一になったが精度は二分の一から三分の一の低下にとどまっている。これは、負例には「明らかに負例である系列」と「正例と特性が類似した系列」があり、前者が大部分のため負例数の増加率よりも精度の低下率が低かったと考えられる（つまり、負例が10倍に増加すると単純にはスコア上位に現れる負例数も10倍に増加することが想定されるが、実際には上位に検出される負例数がそれよりも少なかった）。また、ランダムサンプリングによる方法と比較すると、上位300件で負例を追加しない場合にはシグナル検出器では正解数が23件、ランダムサンプリングでは5~6件で約4倍

に対して、負例を追加した場合にはシグナル検出器では正解数が18件、ランダムサンプリングでは1~2件で約10~20倍と、ランダムサンプリングと比較した場合の効率は負例を追加した方が高いという結果である。したがって、負例数が増加すると問題の難易度が高くなるため、検出順位上位での正解数は減少するが、マイニングによるシグナル検出器の効果が下がるわけではなく、ランダムサンプリングに比べマイニングでは、大規模なデータを効率よく判別することが可能であることが示された。

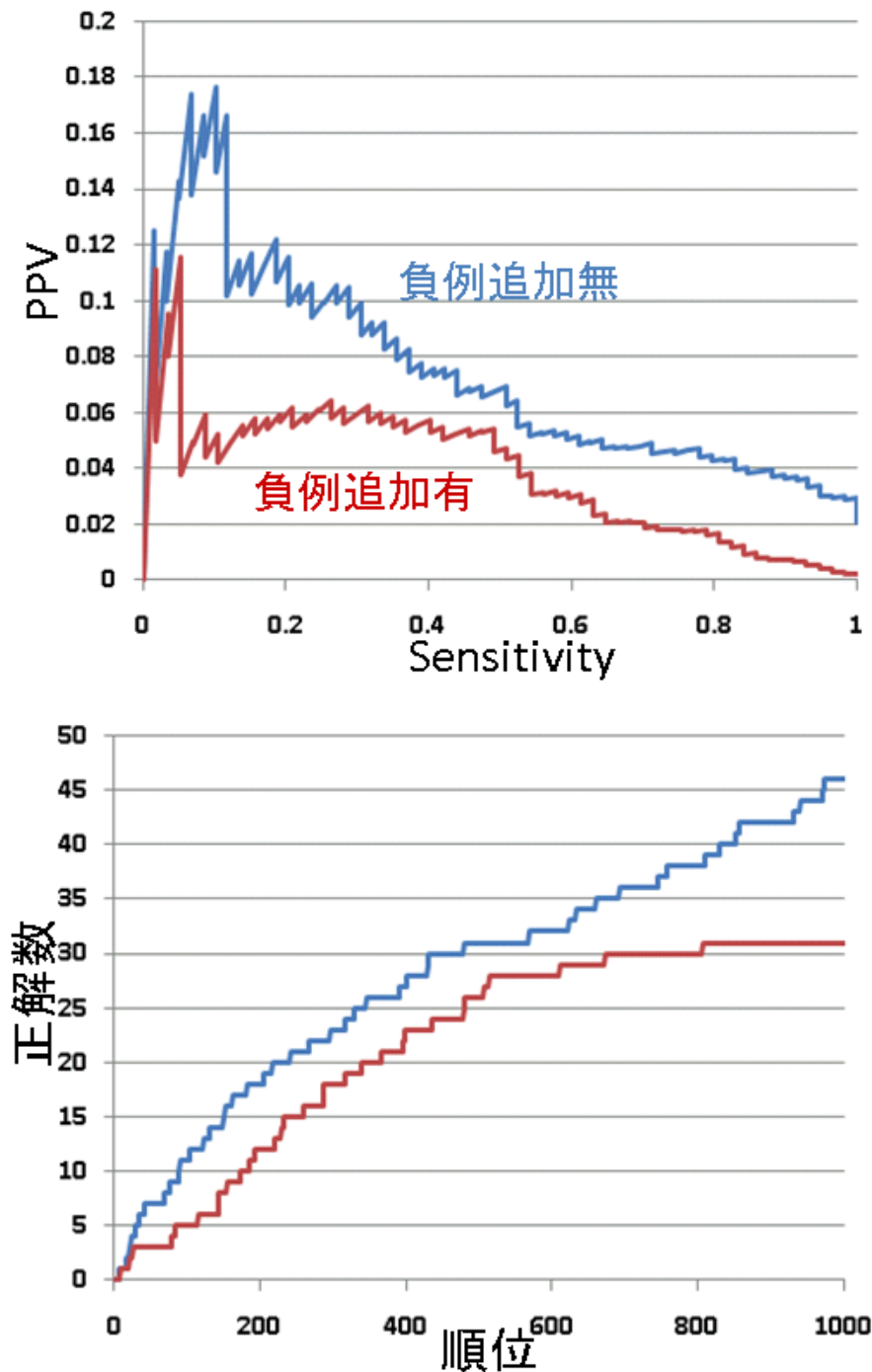


図4-5 負例の追加によるシグナル検出精度の比較

(上図：PPV-Sensitivity曲線、下図：検出順位対する正解（正例）数）

#### 4-8 シグナル検出結果の評価4（副作用パタンの比較）

本評価では、副作用パターン1から副作用パターン3に対するシグナル検出結果を比較するために、ケース1：副作用パターン1、ケース2：副作用パターン1＋副作用パターン2、ケース3：副作用パターン3、を正例とした3ケースについて、シグナル検出精度の評価を実施した。本評価では、パターンによるシグナル検出の難しさやその傾向を調べることを目的としている。ケース1は正例数が57件で負例数が24618件、ケース2は正例数が226件で負例数が30975件、ケース3は正例数が4711件で負例数が208409件である。

図4-6にPPV-Sensitivity曲線と検出順位-正解数曲線を示す。まず、パターン1のみの場合とパターン1＋パターン2の場合（上段）を比較すると、検出順位1000位程度までの上位ではパターン1＋パターン2の方が精度が高く、検出順位が高くなるとパターン1の方が同じSensitivityに対するPPVが高く検出精度が高い。ケース2の方が上位での検出精度が高くなった点は、ケース1とケース2ではケース2の方が学習する正例数が多いことや、パターン2にパターン1と類似する特性をもつ副作用があり、上位で検出できていた事が要因として考えられる。一方で、パターン2には異なる5種類の副作用イベントが混在しており（4-1節参照）、単一のシグナル検出器で学習するにはパターンが複雑で、検出順位が一定以上での検出精度低下につながったと考えられる。性質の異なる副作用イベントは、検出をするために重要な特性が異なる場合があり、領域知識によって副作用をグルーピングしたり、マイニングによるクラスタリングを併用したりする事で精度を改善できる可能性がある。パターン3（下段）は、パターン1やパターン2と比較してシグナル検出精度が高い事が確認できる。最大の要因は、正例数が4711件と多くシグナル検出器を安定して学習出来た点と考えられる。また、「薬剤性\*\*\*」という傷病がレセプトに現れるという臨床的イベントが、ビットパターンによる特性データによって表現し易いことも一因として考えられる。

ケース2について、図4-7でパターン1とパターン2の検出順位に対する正解数の内訳を示す。検出順位の最上位では、パターン1の正解数が多く、徐々に両者の正解数の差がなくなる傾向が確認できる。パターン1の正例数が57件、パターン2の正例数が169件で、パターン2の正例数が多い事を考慮すると、パターン1と比較してパターン2の検出が難しいと言える。これは、パターン1が発生パターンに基づくスクリーニングルールとPMDAの目視に基づいているため、パターン3と同様にバイナリが発生パターンによる特性データによって表現し易い事が一因として考えられる。逆に、パターン2は、スクリーニングルールとは独立に収集された正例のため、バイナリが発生パターンによる特性データだけでは正例の特徴をうまく捉える事ができていないと考えられる。

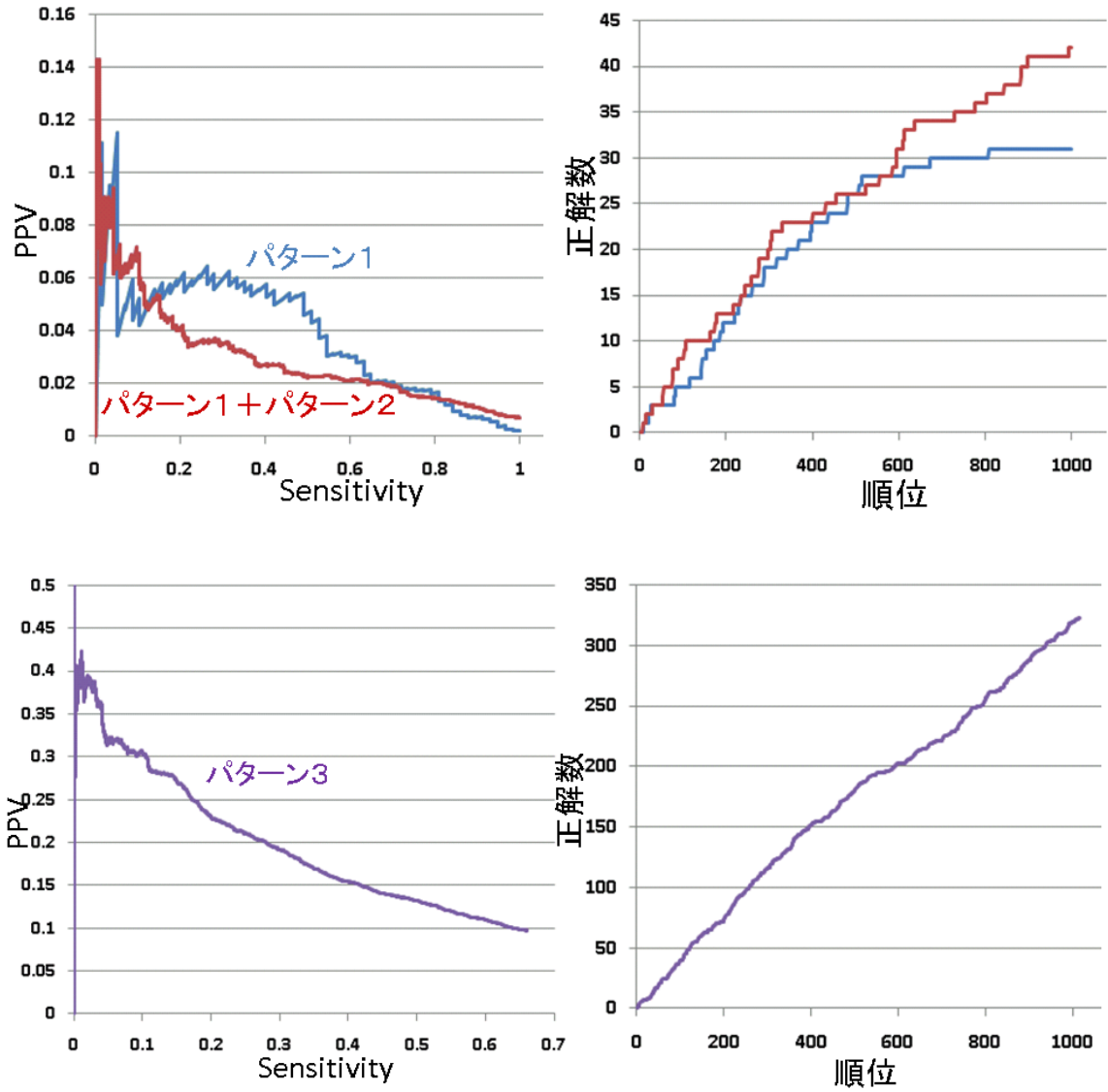


図4-6 パタンによるシグナル検出精度の比較  
(上段：パタン1、パタン1+パタン2、下段：パタン3)

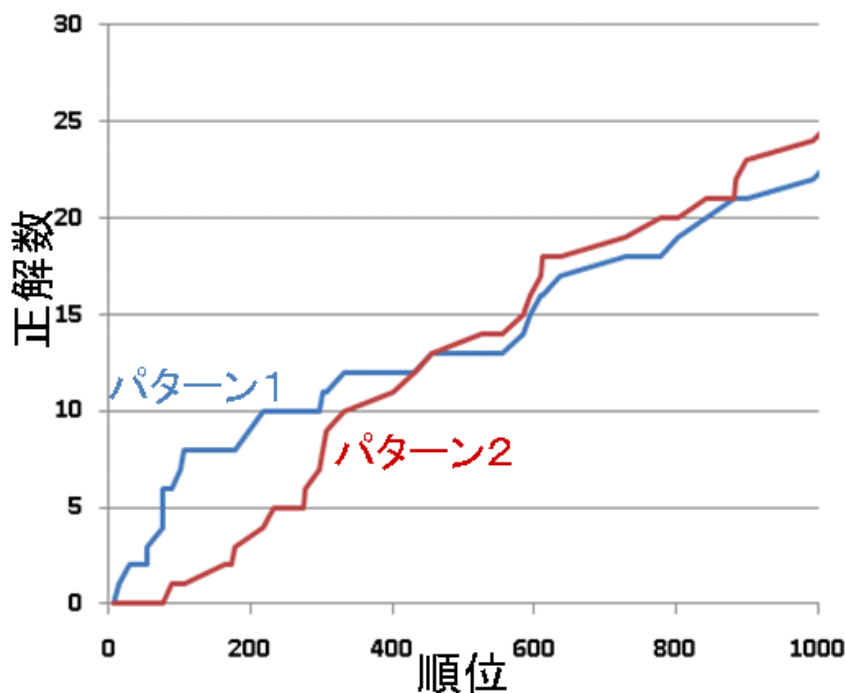


図4-7 パタン1+パタン2の評価における各パタンの正解数

#### 4-9 シグナル検出結果の評価5（特性データの発生パタン長）

本評価では、特性データを生成する際に発生パタンの長さを3ヶ月と6ヶ月と変え、シグナル検出精度を比較評価した。この実験では、特性データを生成するための重要なパラメータである発生パタン長の影響を調査することで、精度向上の可能性を議論することを目的としている。なお、実験の構成は、4-8節のケース1（パタン1）とケース2（パタン1+パタン2）とした。（なお、評価では2008年12月のレセプト系列がデータから除外されている。理由は評価6で述べる）

図4-8に、PPV-Sensitivity曲線と検出順位-正解数曲線を示す。ケース1（上段）では、検出順位最上位での検出精度に大きな変化はなく、500位以上の場合に3ヶ月とした方が、精度が高くなる傾向が確認できる。次にケース2（下段）では、検出順位上位において顕著に精度が向上し、検出順位下位では6ヶ月の場合も3ヶ月の場合もほとんど精度に差がない。検出順位上位で、前者は精度が向上せず、後者は精度が向上した要因は、以下の2点が考えられる。まず、パタン1は正例を生成する際に6ヶ月発生パタンに相当するスクリーニングルールが利用されているため、6ヶ月と3ヶ月でどちらも正例の特徴をよく表現できる点が考えられる。次に、正例数がケース1では57件、ケース2では226件と十分とはいえないため、発生パタンに必ずしも合致しない正例（パタン2に多いと考えられる）の傾向をつかむには6ヶ月では複雑で3ヶ月程度に単純化したほうがよい点が考えられる。なお、6ヶ月の発生パタンは2の6乗で64通りに対して、3ヶ月では2の3乗で8通りである。このように、データ

の性質や正例数が異なると適切な発生パターン長が変わるため、データに応じた調整を行う（場合によっては機械学習技術で最適化する）事が、精度向上に向けた一つの課題である。

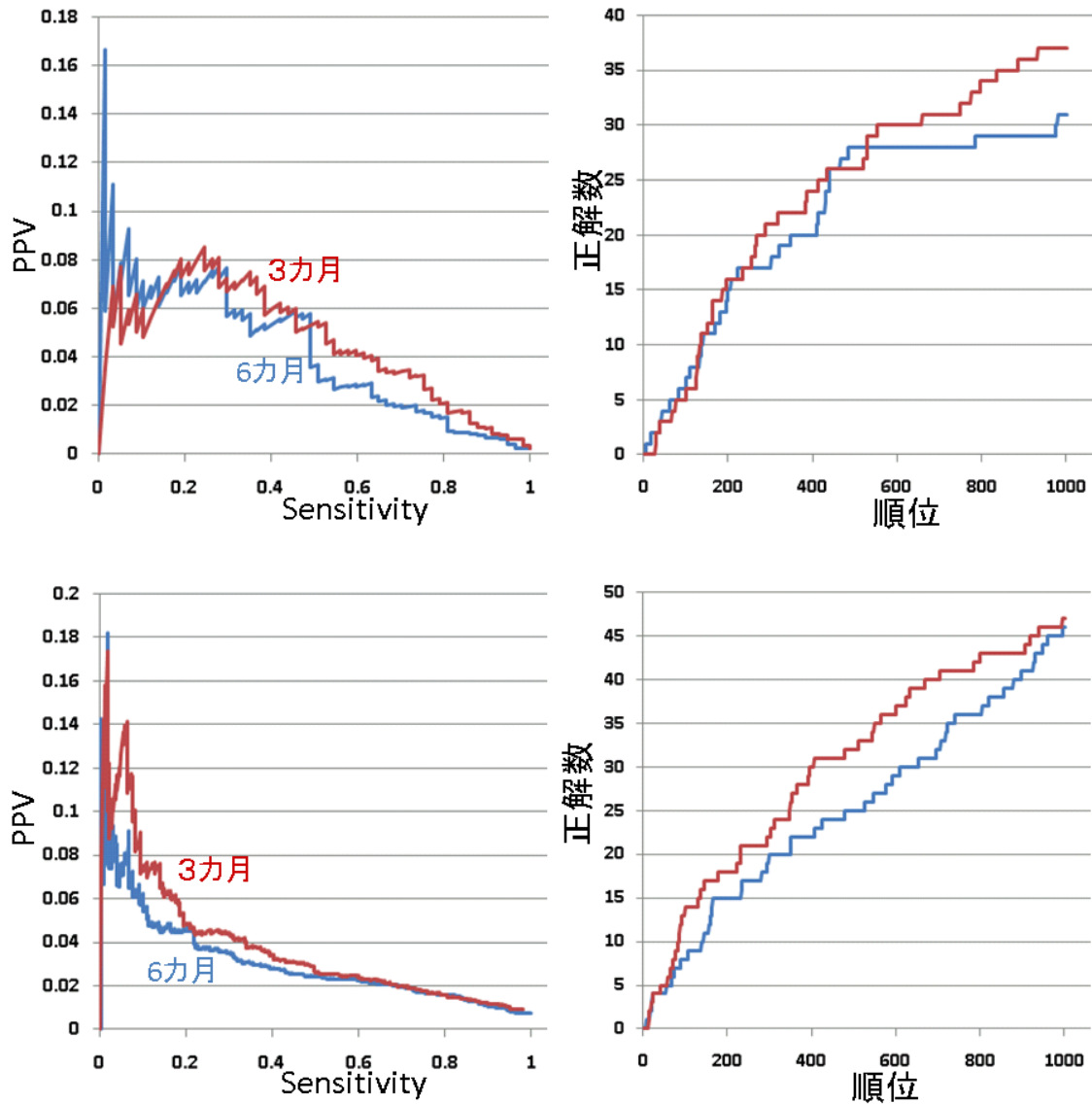


図4-8 特性データの発生パターン長の比較（6ヶ月と3ヶ月）

（上段：パターン1、下段：パターン1＋パターン2）

#### 4-10 シグナル検出結果の評価6（誤検出に関する考察）

評価1から評価5を通じて、検出順位上位で一定の精度で正例（副作用シグナル）を検出可能である事が確認されたが、負例も一定数は誤検出されることが確認された。その中で、目視による確認では、特定の年月に関するレセプト系列が検出順位の上位に現れる傾向が確認されたため、定量的な評価と考察を行った。

検出順位に対して、どの年月のレセプト系列が何件出現したかを図4-9にグラフ化した。図4-9では、time 0が2005年1月、time 1が2005年2月というようにtimeの引数が2005年1月からの月数を表し、time 47が2008年12月となる。

まず、ケース2（パタン1+パタン2、上図）をみると、検出順位上位におけるtime 47（2008年12月）の出現頻度が他の年月と比較して突出して大きい事が確認できる。逆に、ケース3（パタン3）ではtime 0の出現頻度が突出して大きい。まず、ケース2については、パタン1の正例は「原因薬の中断後に副作用が発生する」という、医薬品の発生パタンが重要な特性となる。これは、図3-6の0-1のビットパタンでいえば、111000（6ヶ月で3ヶ月処方後3ヶ月中断）というような、「パタンの後ろ側が0」という特性が重要である事を意味する。本検討では、2005年1月以前と2008年12月以後のデータがないため、その前後は0で補完されるため、2008年12月を中心とする発生パタンは、評価の都合上で「パタンの後ろ側が0」という特性をもってしまい、正例と類似してしまったためと考えられる。同様に、パタン3では「薬剤性\*\*\*」という医薬品が出現する事が重要な特性となり、これは「パタンの前側が0」という特性が重要である事を意味する。したがって、ケース3では、2005年1月（time 0）が正例と類似してしまい、検出順位上位に出現しやすくなったと考えられる。

これらの傾向は、本検討のようなデータ期間が限定された評価データを利用した場合に発生してしまう誤検出であり、より長期間のレセプトデータでの運用を考えた際には本質的に重要な誤検出とはならない。そこで、time 0とtime 47に対応するレセプト系列をデータから除外して、シグナル検出精度を評価した。

図4-10に、time 0とtime 47を除去した場合としない場合のシグナル検出精度の比較を示す。検索順位下位では、除去しない方の精度が高い。しかし、上位1000程度では、time 0やtime 47を除去する事で上位の（評価のために発生する、ある意味人工的な）負例が検出されなくなるため、精度が向上する。この結果は、評価データの実際には発生しない「端」の誤検出を除いた結果精度が向上したという事と、学習データからこのようなデータの「端」を除く事で精度の向上が見込めるという事の二つを示唆している。

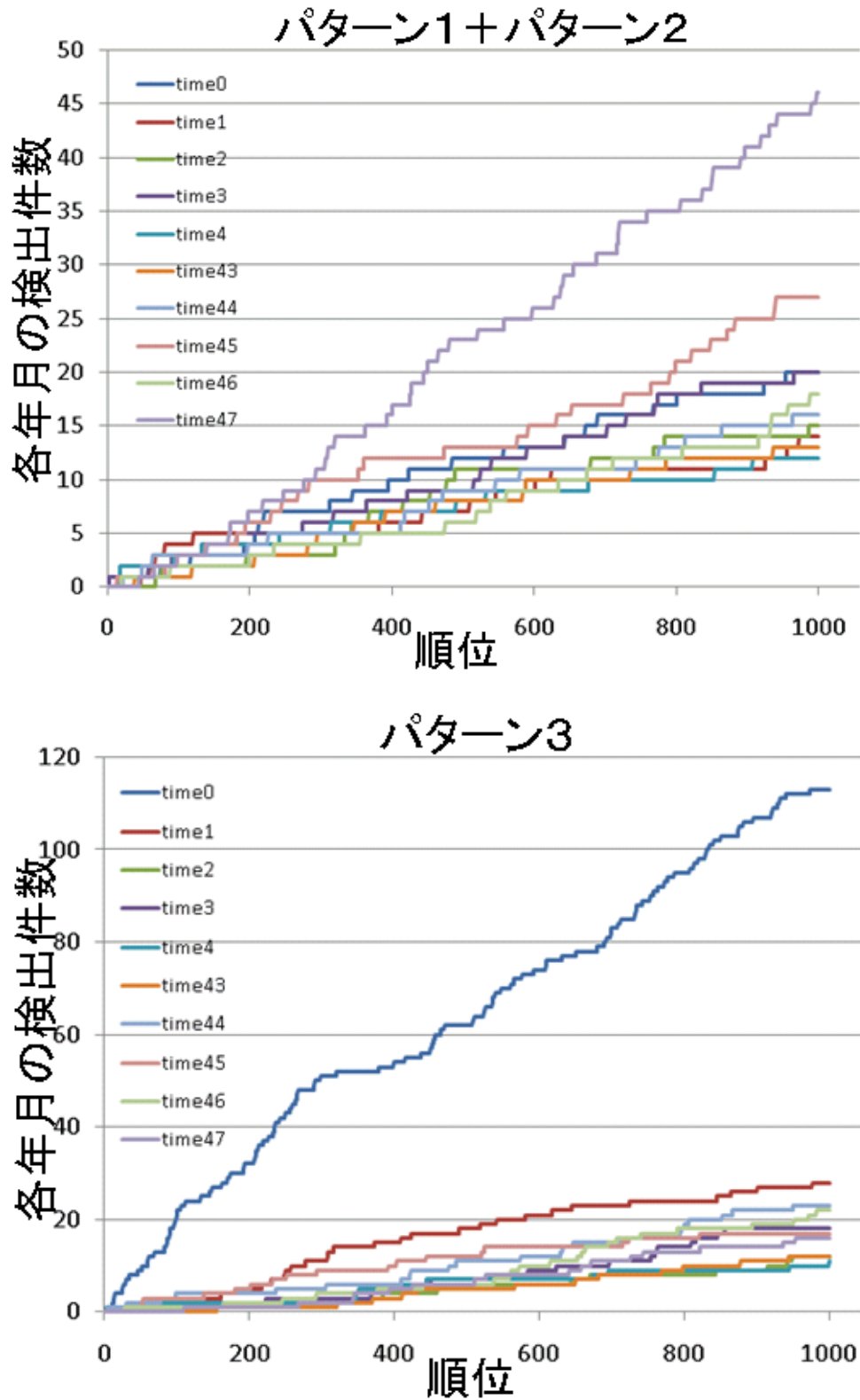


図4-9 検出順位上位に現れる年月数

(time0が2005年1月でtime1、time2と1ヶ月ずつ増え、time47が2008年12月)

(上段：パターン1+パターン2、下段：パターン3)



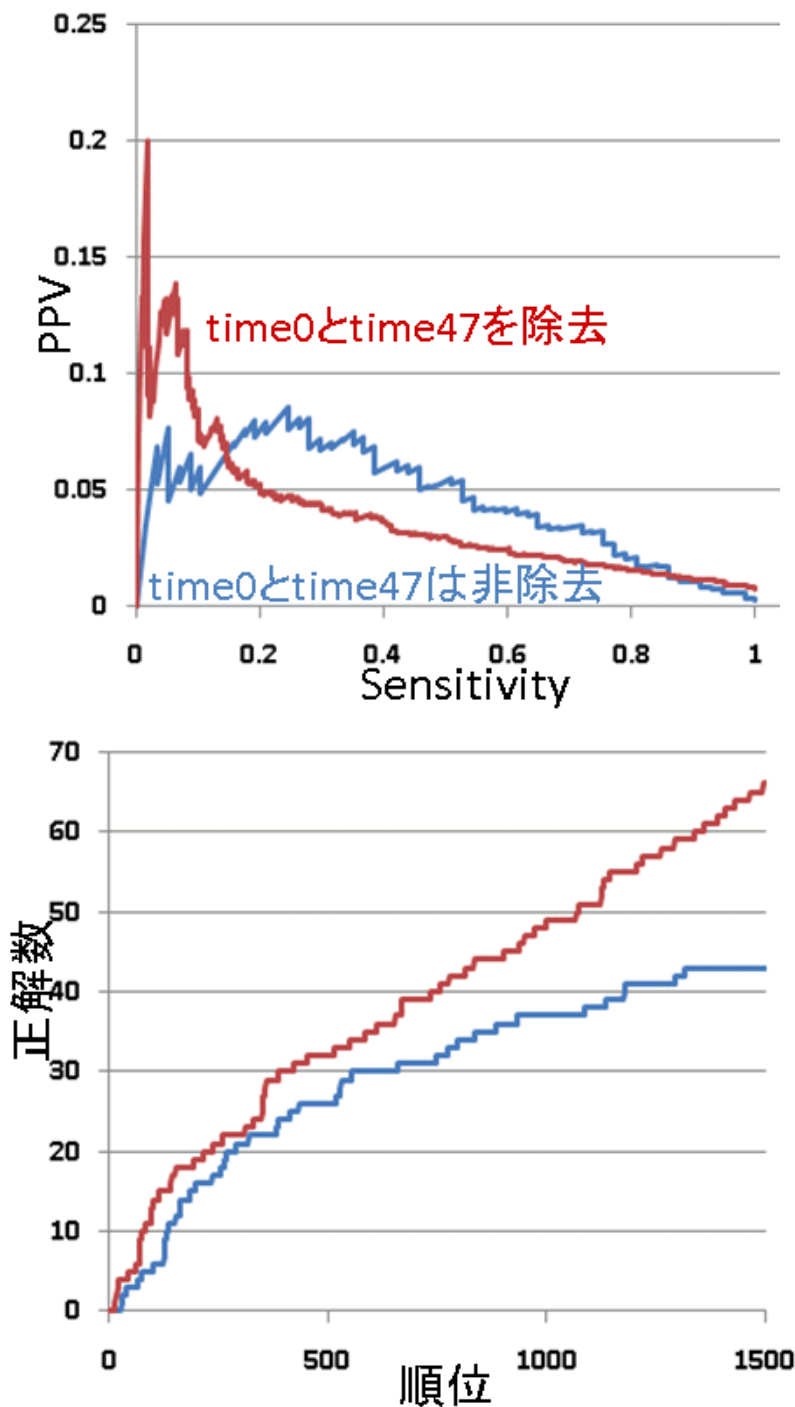


図4-10 time0とtime47を除去した場合と非除去の場合の比較  
 (上図：PPV-Sensitivity曲線、下図：検出順位対する正解（正例）数)

#### 4-11 まとめと考察

シグナル検出課題2として、副作用辞書に記載されていない副作用のシグナルを網羅的に検出する方法を検討した。副作用辞書に記載された副作用を検出するシグナル検出課題1（第3章）で得られたレセプト系列の発生パターンに基づく特性データを一般

化する事で、データマイニングを用いて副作用辞書非記載の副作用を検出できる可能性がある事を確認した。

シグナル検出精度の向上に向けては、第一に特性データを高度化する事が挙げられる。課題1では副作用辞書の情報を使い重要な情報を特定した上で特性を生成した結果、非常に高い精度でシグナルを検出できたが、課題2では副作用辞書の情報を利用していないため、例えば医薬品について原因薬になりうるか否か、という非常に大きなくくりで特性データを生成した結果、診断精度の低下につながった。医薬品や診療行為は、ある程度臨床領域知識からの分類が可能であったり、注目すべき薬剤情報のようにターゲットを絞る情報が入力可能であったりすると考えられ、そのような情報を特性データの生成に活用する事が重要である。また、機械学習技術を用いて、医薬品の組合せや非線形な特性を生成する事も考えられる。ただし、非線形属性を利用する場合には、検出結果の根拠を解釈する事が難しくなるという精度とは別の課題が出る事に注意が必要である。

次に、シグナル検出器を高度化する事が課題として挙げられる。本検討では、単一のシグナル検出器によって検出を試みたが、副作用パターンによって重要な特性の傾向が異なるため、領域知識やクラスタリング技術に基づき、副作用パターンをグルーピングし、複数のシグナル検出器を学習し検出をする事で、シグナル検出精度が改善する可能性がある。また、評価データ中の正例の割合が低下すると、相対的に誤検出が多くなり上位での検出が難しくなるため、明らかに副作用ではないレセプト系列はスクリーニングルールによってフィルタし、ルール化しにくい部分をマイニングによって判定するなど、ルールとマイニングの併用も重要である。

本検討では、シグナル検出器を学習する目的とその評価の目的で、正例と負例を準備することが最も困難であった。これは、レセプトデータには副作用イベントの情報が明示的に記載されていないため、レセプト系列だけで副作用を判断することは多くの場合専門家にも不可能で、最終的には相当数の生データの目視と、複数人による確認作業などが発生するためである。また、結果の評価が難しいため、能動学習や半教師学習など、評価者とのインタラクションを前提とする機械学習手法の適用が難しいという問題も発現した<sup>7</sup>。シグナル検出器の精度向上と改善ロジックの評価には正例と負例が必要であり、その効率的な取得の仕組みが今後の重要な課題として挙げられる。

---

<sup>7</sup> 能動学習や半教師学習は、正例や負例のラベルが付いていないデータを活用する手法であるが、結果に対して人間に正例か負例かの判定をしてもらい、再学習を繰り返しながら精度を向上させていく枠組みのため、「ラベルの付いていないデータを正例か負例かに判定する事が難しい」という状況では適用が困難となる。

## 5. 特徴抽出による副作用仮説抽出の検討

---

本章では、シグナル検出の結果から「疑いが強い」とされたレセプト系列で、どのような臨床イベントが起きているかを機械的に分析する「特徴抽出」に関して検討した結果を述べる。

前章で論じたシグナル検出は「どの患者でいつ副作用がおきていると疑われるか」を分析するものであった。シグナル検出の結果をうけて「その患者において、そのタイミングでどのような臨床イベントが起きているか」や「その臨床イベントは副作用が疑われるレセプト系列に特有なものなのか」を、元のレセプトデータにさかのぼって人手で確認するのは、目視や比較の対象となるデータが非常に多くなるため、現実的ではない。特徴抽出はそれを自動的に行う分析である。また本検討では、シグナル検出結果の把握という使用方法を越えて、特徴抽出分析による未知の副作用発見の可能性についても検討を行う。

### 5-1 特徴抽出と副作用仮説抽出

---

レセプト系列データからの臨床イベントの特徴抽出には様々な方法が考えられるが、本検討においてはシグナル検出結果を特徴付けるイベントの抽出を行うものとした。すなわち、シグナル検出器が大きいシグナルスコアを発するレセプト系列において、検出器が出力するスコアを増加させる発生パターンを見せた薬剤、診療行為、傷病名の共起組み合わせを抽出し、それらの頻度を、シグナルスコアの大きい／小さいレセプト系列群で比較するという特徴抽出方式を検討の対象とした。

2-3-2節でも述べたように、シグナル検出器が「いつ副作用が起きているか」を判定するのをうけて、「そのとき何が起きているか」を特徴抽出で確認するというのが本来の分析の目的であるが、本検討ではこれをさらに副作用仮説の発見、すなわち「副作用の原因薬、治療薬、診療行為、傷病名の発見」に利用できないかの検討も行った。これは、シグナルスコアが高いレセプト系列における固有の臨床イベントとして、そのような組み合わせが抽出されるのではないかという期待に基づいた検討である。

### 5-2 特徴抽出手順

---

特徴抽出に用いる入力情報は、

- ・レセプト系列集合
- ・当該レセプト系列に対する検出器の出力（シグナルスコア）

・検出器が学習した薬剤、診療行為、傷病名の各ビットパタンの加重値である。

これらの情報を用いて、特徴抽出は以下の手順で「シグナルスコアが大きいレセプト系列に固有の臨床イベント」を抽出する。

手順1：シグナルスコアが大きいレセプト系列に対して、加重値の大きいパターンに該当する薬剤、診療行為、傷病名を抽出し集計

手順2：シグナルスコアが小さいレセプト系列に対して、上記パターンに該当する薬剤、診療行為、傷病名を抽出し集計

手順3：薬剤、診療行為、傷病名間の共起に関しても、同様に抽出し集計

手順4：これらの集計結果から、シグナルスコアが大きいレセプト系列に多く、小さいレセプト系列に少ない薬剤、診療行為、傷病名、あるいはそれらの共起を特徴として出力

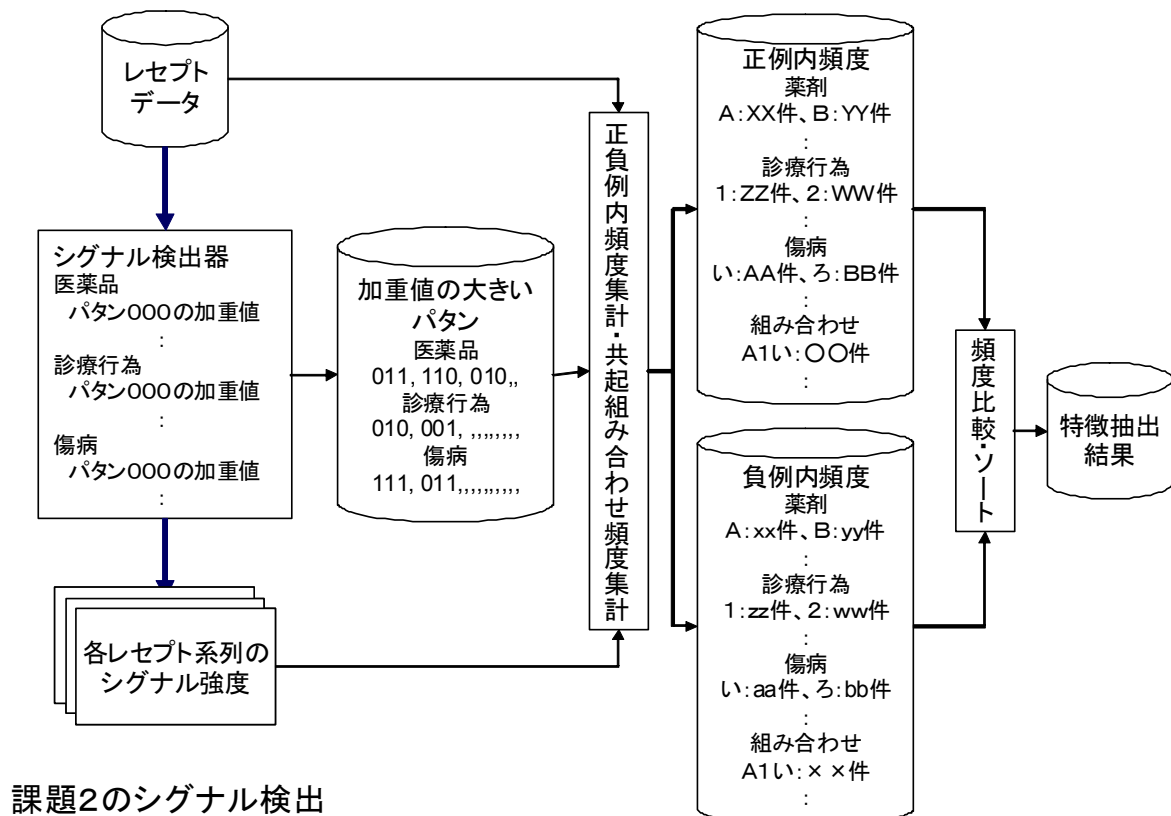


図5-1 特徴抽出手順

ただし、抽出結果を理解しやすくするため、薬剤に関しては、レセプト系列前期に発生が多いパタンのもを原因薬、後期に多いものを治療薬として別々に処理し、出力時にも区別することとした。また、上記に記した手順は概念的なものであり、実際には、大きい／小さいとみなすシグナルスコアの閾値をいくつに定めるかで結果が大幅に変わってしまうことを避ける工夫や、計算量が爆発的に増えるのを避けるための工夫などを行っている。詳細な抽出手順は付録に記述することとする。

図5-1は、上記の特徴抽出手順のイメージを表したものである（本図中ではスペースの関係で「シグナルスコアの大きいレセプト系列」を正例、「シグナルスコアの小さいレセプト系列」を負例と記述している）。

### 5-3 特徴抽出結果の評価1（シグナル検出結果／検出器の挙動把握）

前節の特徴抽出手順を、第4-8章におけるパタン1+パタン2のシグナル検出の結果に適用し、原因薬・治療薬・診療行為の組み合わせについて特徴として抽出されたものを、表計算ソフトで表示した結果の例を図5-2に掲げる。

1	A	B	D	E	G	F	Y					
1	高シグナル内頻度	低シグナル内頻度	頻度比	該当P.D.	TimeSlot	シグナル値	原因薬	治療薬	診療行為			
9401	1.19	3.82	0.24	230159-30	0.28486282	1002-17	0	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	トレン法(トレンジ)C		
10619	1.19	5.81	0.17	230159-30	0.26486262	71060-22	0	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	超音波(UCG)(断層掃		
10691	1.52											
11926	1.25											
11883	1.52											
12022	1.4											
13618	1.61											
13695	1.15	10.85	0.1230159-29	0.01243219	183025-27	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	MG	160020910	診療行		
17590	1.7	16.3	0.09	60290-35	2.93e-06	202105-24	0.7	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	CRP	160054610	診療行
13628	1.57	15.43	0.09	195119-10	0.00076186	254824-10	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	点滴注射	130003810	診	
14017	1.67	17.33	0.09	254824-10	0.00057454	80290-35	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	コンピュータ	断層診断		
14203	1.51	16.49	0.08	254824-10	0.00057454	80290-35	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	Amy	160020310	診療行	
14295	1.74	19.26	0.08	254824-10	0.00057454	80290-35	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	TP	160017410	診療行	
14271	1.49	16.51	0.08	254824-10	0.00057454	202105-24	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬		血液学的検査判断科	1	
14286	1.4	15.6	0.08	254824-10	0.00057454	248581-43	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	腫	160008210	診療行	
14291	1.07	11.93	0.08	230159-29	0.01243219	71060-22	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	ChE	160020210	診療行	
14305	1.55	17.45	0.08	254824-10	0.00057454	202105-24	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	ALP	160020010	診療行	
14375	1.68	19.32	0.08	254824-10	0.00057454	202105-24	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	クシアチニン	160019210	診療行	
14397	1.51	17.49	0.08	254824-10	0.00057454	80290-35	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	免疫学的検査判断科	1		
14405	1.74	20.28	0.08	195119-10	0.00076186	202105-24	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	デジタル映像化処理(単			
14426	1.73	20.27	0.08	254824-10	0.00057454	248581-43	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	単純撮影(イ)の写真診断			
14445	1.56	18.44	0.08	195119-10	0.00076186	254824-10	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	CPK	160020810	診療行	
14462	1.65	19.45	0.08	254824-10	0.00057454	80290-35	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	UCG	160020910	診療行	
14517											110	診療行
14560											110	診療行
14564											110	診療行
14565											110	診療行
14580											1)	判断#
14603											17010	巨
14604											110	診療行
14700												診療行
14701	1.58	20.42	0.07	254824-10	0.00057454	202105-24	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	Na及Cl	160021110	診	
14701	1.60	21.41	0.07	254824-10	0.00057454	248581-43	ベシル酸アムロジピン 内用薬	ヘパリンナトリウム 注射薬	単純撮影(イ)	160020810	診	

出力される組み合わせ数は膨大  
→結果の目視には表計算ソフトの検索／抽出機能の活用が必要

- ・「あやしい」とされたレセプト系列の膨大な情報を、薬剤(原因薬、治療薬に区別)、診療行為、傷病の発生パターンの観点から整理  
→シグナル検出結果の目視確認方法の一つとなり得る。
- ・シグナル検出器の挙動把握方法としては改良の余地あり

図5-2 特徴抽出の結果を表計算ソフトで表示した例

ここで掲げた例は、膨大な抽出結果に対して、表計算ソフトの検索機能等を用いて特定の原因薬と治療薬についてのみ表示し、シグナルスコアの大きいレセプト系列における頻度とシグナルスコアの小さいレセプト系列における頻度の比(頻度比)が大きい順にソートしたものである。

図からも見て取れる通り、特徴として抽出される臨床イベントの組み合わせ（この例では原因薬と治療薬、診療行為の組み合わせ）は膨大な数にのぼり（次節も参照）、結果の目視には表計算ソフトの検索や抽出、ソートに相当する機能が必須となる。

抽出されるものが膨大な数になるということは、事実として「シグナルスコアが大きくなるときでも、必ずしも少数の副作用仮説に相当する臨床イベントのみが固有に発生しているのではなく、様々な臨床イベントも同時におきている」ということを示している。そのような状態に対して「何がおきているか」を見て取るという目的に対しては、このように「特徴抽出分析を行った結果を表計算ソフトにより目視する」というのは、ひとつの方法となりえる。一方、あまりにも提示される情報が多いため、「どのような入力データのときにシグナル検出器は大きな値を出力するのか」を読み取ることは簡単ではなく、シグナル検出器の挙動を把握するという目的に対しては、それを達成しているとは言い難い。検出器の挙動を把握するという目的に対しては方式の改良が必要である。

#### 5-4 特徴抽出結果の評価2（マスクされた副作用仮説の再現）

---

課題2のシグナル検出器は、学習時に提示されなかった種類の副作用についても、シグナルを一部検出することができることを前章で観察した。これは、当該シグナル検出器の出力をうけて特徴抽出分析を行うことで、学習時に提示されなかった副作用に関する「薬剤、診療行為、傷病」等の組合せ、すなわち副作用に関する仮説を発見することができる可能性があることを示唆している。

本節では、一部の副作用辞書の内容をマスクして実験を行った第4－8章のシグナル検出器の出力を入力とし、第5－2章で示した手順で特徴抽出を行うことで、学習時にマスクされていた辞書の内容が再現されるかを観察し、上記のような副作用発見の可能性に関する検討を行う。

表5-1 抽出された特徴数とマスク内容該当数

<b>原因薬</b>	
抽出された特徴の数	731
マスク内容に該当する数	99
<b>原因薬・治療薬の共起</b>	
抽出された特徴の数	19372
マスク内容に該当する数	458
<b>原因薬・治療薬・診療行為の共起</b>	
抽出された特徴の数	47797
マスク内容に該当する数	241
<b>原因薬・治療薬・診療行為・傷病の共起</b>	
抽出された特徴の数	61604
マスク内容に該当する数	0

実験条件は第4-8章と同じである。このシグナル検出の結果に対して「原因薬」、「原因薬、治療薬の組み合わせ」、「原因薬、治療薬、診療行為の組み合わせ」、「原因薬、治療薬、診療行為、傷病名の組み合わせ」の4通りの特徴抽出を行い、その結果に対して「学習時にマスクされていた辞書内容に合致するものが何件あるか」の集計を行った。表5-1は抽出された特徴数と、マスクされていた辞書内容（＝マスク内容）に合致していたものの件数をまとめたものである。

「原因薬」、「原因薬、治療薬の組み合わせ」、「原因薬、治療薬、診療行為の組み合わせ」に関しては、マスク内容の再現はされているが、膨大に抽出される大量の特徴に比して多くはない。これは、特徴抽出結果から副作用仮説の発見を行おうとしても、このままでは効率は低くなることを意味している。また、第5-3章に記載した頻度比や、独立性の検定統計量、あるいは単純に「シグナルスコアの大きいレセプト系列内の頻度」などを用いて抽出される特徴をソートしてみたが、特にマスク内容に相当する特徴が上位に集まるようなことは観察されなかった。これは、そのような自然なソート方法程度では、副作用仮説の発見効率を向上させることは難しいことを示している。

また「原因薬、治療薬、診療行為、傷病名の組み合わせ」に関しては、マスク内容に相当する特徴は抽出されなかった。この四つの組み合わせの特徴抽出から副作用仮説を発見することは、非常に難しいことを示している。

### 5-5 特徴抽出結果の評価3（そのほかの副作用仮説の抽出）

上記の実験でマスク内容に相当するものが抽出されなかった「原因薬、治療薬、診療行為、傷病名の組み合わせ」の特徴に関して、臨床的観点から副作用に該当するものが含まれていないかをPMDAにより目視で確認を行った。

結果としては抽出された特徴の総数61604件中、9件のみが副作用に相当していた。これは非常に低い割合であり、この四つの組み合わせの副作用仮説を本特徴抽出分析によって発見しようとしても、極めて効率が低くなることを示している。

## 5-6 副作用仮説発見効率の向上可能性

---

第5-4章のマスク内容再現実験の結果も、第5-5章のその他の副作用仮説の目視確認の結果も、「シグナル検出結果の特徴抽出による副作用仮説発見」は効率が低くなることを示している。仮説発見の目的には、シグナル検出結果からの特徴抽出は適していないといえる。

今回はシグナル検出を主目的とする検討において、副次的な分析として特徴抽出による仮説発見を検討したが、もし仮説発見を主目的とするならば、最初から「薬剤、診療行為、傷病名」等の組み合わせが副作用に相当するか否かを同定するための学習を行う方が自然である。そのような学習方法としては関係性学習などの技術が存在し、学習された関係性を満たす組み合わせを出力する分析は情報抽出と呼ばれて研究が行われている。

副作用発見のために関係性学習を行い情報抽出をするとすれば、今回のシグナル検出結果からの特徴抽出と比較して「どのような条件を満たす組み合わせを出力するか」が異なってくる。今回の特徴抽出が出力する組み合わせが満たす条件は「シグナルスコアが大きいレセプト系列で」「シグナルスコアを増加させる発生パターンをもつ臨床イベント」である。

副作用発見を目的とした関係性学習等を行った際に、発見効率が向上する可能性があるかを評価するために、今回の特徴抽出における上記の出力条件の後者を変更、すなわち特徴として出力する臨床イベントの発生パターンを変更して第5-4章と同様な「マスクされた内容が再現されるか」の実験を行ってみた。

表5-2は発生パターンに関する抽出条件を変更した実験の結果である。発生パターンの抽出条件以外は、第5-4章の実験と同じ条件で行っている。



表5-2 抽出用のパタンの変更と辞書内容の再現性

(数値上段が抽出特徴数、下段がマスク内容該当数)

抽出する特徴	抽出に用いたパタン			
	元のまま	Strict	Typical	人手設計
原因薬	731	393	440	800
	99	79	84	88
原因薬・治療薬	19372	2346	3357	28861
	458	127	183	416
原因薬・治療薬・診療行為	47797	34538	65075	76353
	241	456	893	100
原因薬・治療薬・診療行為・傷病	61604	95316	90727	66124
	0	131	128	0

表中の抽出に用いたパタンは

- ・元のまま シグナル検出器で加重値の大きいパタン (第5-4章と同じ)
- ・S t r i c t 原因薬は期間前半での頻度が後半より2以上大きい全てのパタン、治療薬、診療行為、傷病はその逆
- ・T y p i c a l 原因薬は期間前半での頻度が後半より1以上大きい全てのパタン、治療薬、診療行為、傷病はその逆
- ・人手設計 副作用仮説の抽出がうまくいくことを狙って人手で工夫したパタン。シグナルスコアの大きいレセプトデータを目視しながら、不要と思われるパタンを上記T y p i c a l から削除することにより調整 (体系的に何かの基準に基づいて調整を行ったわけではない) である。

S t r i c t やT y p i c a l に関しては、そのようなパタンを用いることによりマスク内容の再現性がよくなる場合があることが見て取れる。特に、シグナル検出結果から特徴抽出するだけでは全く再現できなかった、「原因薬、治療薬、診療行為、傷病名の組み合わせ」に関するこれらのパタンでは再現できるものがあることが見て取れる。一方、人手によるパタンの設計では本実験では再現性を高めることに失敗していることも観察された。

本実験においては抽出するパタンの条件を単純に別のものに変えてみただけであるが、それだけでも再現性があがることが観察された。実際に「仮説発見を目的とした分

析」をする際には、仮説発見がうまくいくように抽出するパターンの条件等が学習によって調整されるので、再現性はさらに向上すると考えられる。

## 5-7 まとめと考察

---

シグナル検出結果を入力として、シグナルスコアが大きいレセプト系列において、シグナルスコアを増加する臨床イベントの組み合わせを出力する特徴抽出分析について、シグナル検出結果の整理やシグナル検出器の挙動の把握、副作用仮説の発見の可能性の観点から検討を行った。

抽出される特徴（臨床イベントの組み合わせ：原因薬と治療薬、診療行為の組み合わせ）の数は膨大で、結果を目視するには表計算ソフトにおける検索やソート等に相当する機能が必要である。シグナル検出結果の整理という意味では一つの手法として位置づけられるが、シグナル検出器の挙動を把握するのは困難である。

副作用仮説の発見可能性の検討のためには、学習時にマスクされた副作用辞書内容の再現性を評価する実験と、その他の副作用に相当する内容の目視確認を行った。再現性実験では、膨大に抽出される特徴の一部のみしかマスク内容に合致しないことが観察された。また、その他副作用の目視確認においてはさらに極めて低い割合でしかそのような特徴は含まれていないことが観察された。これは、このような特徴抽出をそのまま利用しても、副作用仮説の発見効率は低くなることを示している。

一方、参考実験として、シグナルスコアが大きいレセプト系列において抽出すべき対象を選定する条件を変えてみたところ、マスクされた副作用辞書内容の再現性が向上することが観察された。これは、仮説発見自体を目的とした関係性学習等による情報抽出分析を行うことにより、副作用仮説の発見効率が向上するであろうことを強く示唆するものである。

## 6. NRDB への適用に向けた課題

本検討作業では、クレンジングやコーディング等の整形作業を済ませてあるレセプトデータ約40万人分を分析の対象とした。そのため、データの不備やサイズが主たる問題となることはなかった。一方、ナショナルレセプトデータベース（NRDB）を分析の対象とする場合には、このデータ整形と莫大な人数へ対応は非常に大きな問題となることは明らかである。本章では、本検討と同様な分析をNRDBに対して行う際に想定される課題と、その解決案に関して述べる。

### ・クレンジングの問題に起因する課題

NRDBにおいては、入力規則の違反や値の欠損、誤入力を含むデータレコード（不全レコード）が一定の割合で含まれると想定される。このようなデータレコードが入力データに含まれていると、本検討の対象としたシグナル検出や特徴抽出といった分析においてノイズとして働き、各種精度の低下の重大な要因となる。

この解決方法としては、一定のルールに従って機械的にクレンジング処理を行うことが第一に挙げられる。特に「副作用が発現しているレセプト系列を全て検出する」等の、全件抽出を主目的としないのであれば、データの量自体は多大であるので、不全レコードを修正する方法を工夫するのではなく、機械的に入力データから取り除くという方式が適切であると想定される。そのような処理は既存の技術である程度までは実現可能である。

### ・コーディングの問題に起因する課題

コーディングの主眼は「同一とみなすべきものの表現を同一にする」「区別すべきものの表現を違うものにする」こと（一意性の保証）である。本検討においては、コードの値自体は表面的には分析に使用していないが、辞書のマッチングや、特徴データの生成、臨床イベントの発生回数の集計等のいたるところで、一意性の保証に依存した処理を行っている。NRDBにおいては、同一とみなすべきものが違う表現で入力されていたり（表記ゆれ）、区別すべきものも同一の値を入力されていたりする（低分解能）レコードが無視できない割合で含まれると想定される。表記ゆれはシグナル検出器の学習効率（一定の精度を獲得するのに、どれくらいの学習データが必要か）の低下や、特徴抽出結果の可読性の低下を引き起こす。低分解能は各種分析の精度を低下させる。

表記ゆれへの対処方法は、機械的に表記ゆれを修正する処理が第一に挙げられる。表記ゆれ修正処理は、基本的に表記ゆれ辞書に基づいて、同一表現すべきものをレセプトデータからマッチング抽出し、特定の表現に揃えていくという動作をする。どれだけ充実した表記ゆれ辞書を作成できるかが、表記ゆれ修正処理の精度を決定するが、表記ゆ

れ辞書作成を支援するシステムの研究も行われており、これも既存の技術である程度までは可能であると想定される。また、副作用辞書などのような突合すべき対象データが、レセプトデータと異なるコード体系で表現されている場合も、問題は表記ゆれと同じであり、ここで述べた対処方法が第一に挙げられるものである。

低分解能への対処は既に失われている情報を復元することに相当するので難しい。汚いレコードとみなしてクレンジング処理により入力データから削除する等が考えられる。

#### ・データサイズの問題に起因する課題

NRDBのデータサイズは、今回対象としたデータと比べ物にならないほど大きい。これら全件に対して、本検討における分析を素朴に適用することは、必要な処理時間・メモリ量などの意味から非現実的である。

データサイズへの対処は、分析の種類や目的に依存して、サンプリングや処理のオンライン化、分解・並列化などが考えられる。例えば、シグナル検出器の学習であれば、もとより全データを入力する必要はなく、各加重値の値を精度よく決定するのに必要なだけのデータ数をサンプリングして用いれば足りる。また、特徴抽出における各種イベントの頻度集計は、全データを一度にシステムに入力して行う必要はなく、データを入力するたびに過去の集計結果を利用して集計結果を更新していくような方法（オンライン化）が可能である。また、学習が済んだシグナル検出器で、色々なレセプト系列に対するシグナルスコアを計算する際も、一つのシグナル検出器に全データを入力するのではなく、複数のシグナル検出器を準備しておいて、それぞれが手分けしてシグナルスコアの計算を行う（分解・並列化）等の方法が可能である。

#### ・運用フロー関係の課題

本検討においても、分析結果や学習されたシグナル検出器の加重値に対する人手による評価結果を、さらに分析アルゴリズムの調整にフィードバックするというフローが重要な役割を果たした。NRDBにおいては各種イベントの発生頻度の経年的変化や、新薬の追加などがおこるので、分析アルゴリズムの調整は継続的に必要となり、さらにこのフローの重要性が大きくなると想定される。NRDBに対する分析運用においては、分析結果が一方的に出力されるだけでなく、出力された分析結果に対する何らかのフィードバックが、分析アルゴリズムの調整のために利用されるようなフローが存在することが重要であるといえる。

## 7. まとめ

---

レセプト系列の医薬品や診療行為パターンを特性とし、「副作用発現の疑いの強さ」を出力するシグナル検出器を機械学習技術によって構成し、副作用辞書に記載された副作用と記載されていない副作用の両者についてシグナル検出の可能性を評価した。実験結果からは、学習した副作用イベントだけでなく、類似の特性をもつ未学習の副作用イベントに関しても、副作用が発現しているレセプト系列を特定できる可能性がある事が示唆された。一方で、正例数の不足、正例と類似の特性を示す負例の存在等に起因し、直接実務で利用するに十分な検出精度は得られなかった。

また、シグナルスコアが大きいレセプト系列固有の臨床イベントを、機械的に抽出する「特徴抽出」分析に関しても検討を行った。実験からは、シグナルスコアが大きくなるレセプト系列の特徴を抽出するだけでは、副作用仮説（薬剤、診療行為、傷病名等の組み合わせ）発見の効率は低くなるが、仮説発見を目的とした分析手法で効率向上できる可能性は十分に期待できることが確認された。

シグナル検出精度向上にむけては、特性データの高度化（組合せ、外部情報、非線形等）やシグナル検出ロジックの高度化（ルールとの融合、学習データ獲得の効率化等）等が今後の重要な課題として挙げられる。また、特徴抽出については、本検討は副作用仮説の抽出という視点に関しては、重要でありながらも、副次的な取り組みによって可能性を示したにとどまっている。「副作用仮説の抽出」という視点からの本格的な取り組み、および今回の結果との融合方法の検討は重要な今後の課題である。

## 付録

---

### A. シグナル検出アルゴリズム

---

本節では、機械学習に基づくシグナル検出アルゴリズムの詳細について説明する。

シグナル検出器は、正例と負例を学習し、評価対象のレセプト系列に対する副作用の疑わしさの順位付けをする機能が要求され、これは機械学習分野でいえば教師付学習問題と捉える事ができる。特に、本検討では以下の理由から、L1正則化ロジスティック回帰モデルを利用する方針とした。

- ・正例と負例が与えられる事から、二値判別問題として捉える事が可能であり、順位付けをするために正例（副作用）である確率（副作用シグナルスコア）を出力可能なモデルである事が望ましい
- ・学習されたモデルがシグナルスコアに基づく順位付けをした際に、シグナルスコア算出の根拠を利用者が理解できる（どの特性がシグナルスコアの増加（減少）にどの程度寄与しているのか）事が望ましい（根拠導出が難しい非線形分類器は精度が良かったとしても本検討には適さない）
- ・ロジスティック回帰モデルは医療分野でも頻繁に利用されるため、検討内容に対する親和性が高い
- ・レセプト系列は多数の特性から構成されるため、シグナル検出器が自動的に重要な特性を選択できる必要がある

以下、L1正則化ロジスティック回帰の数理的構造について説明する。

Xを特性データに対応するD次元の説明変数とし、Yを副作用発現の有無を表す確率変数とする（Y=+1は副作用が発現、Y=-1は非発現を表す）。また、WをD次元の重みベクトルとする。この時、ロジスティック回帰モデルは、

$$P(Y = +1|X; W) = \frac{1}{1 + \exp(W^T X)}$$

$$P(Y = -1|X; W) = \frac{\exp(W^T X)}{1 + \exp(W^T X)}$$

と表わされる。ただし、 $P(\bullet|\circ; \star)$  は $\star$ をパラメータとし、 $\circ$ が与えられた場合の $\bullet$ の条件付確率を表すとする。また、上付きのTはベクトルの転置を表す。

学習データとして正例と負例  $\{x_n, y_n\}$  ( $n = 1, \dots, N$ ) が与えられた場合に、L1ロジスティック回帰では、以下の目的関数を最適化する事によって、パラメータWの値を算出する。ただし、 $x_n$  と  $y_n$  はそれぞれXとYの実現値とする。

$$L(W) = \sum_{n=1}^N \log P(y_n|x_n; W) - \lambda|W|_1$$

ただし、 $\lambda$  は尤度（右辺第1項）と罰則項（右辺第2項）のバランスを調整するパラメータであり、 $|W|_1$  はWのL1ノルムである。なお、 $L(W)$  はWに関する凸関数であり、勾配法に準じた方法によって最大化をする事が可能である。最大化のアルゴリズムの詳細については割愛する。

$L(W)$  を最大化するパラメータの値を $W^*$  とすると、評価データ  $x$  に対するシグナルスコアは $P(Y = +1 | x; W^*)$  として計算される。この時、WとXのd次元目の要素をそれぞれ $W_d$ 、 $X_d$  とすると、 $W^T X = \sum W_d X_d$  であるため、 $W_d X_d$  の値を確認する事で、特性のうちどの要素がシグナルスコアの増加（減少）に重要であったかを確認する事が可能である。また、詳細は割愛するが、L1正則化によってWを最適化した場合には、D次元のWの要素の多くが0となる（つまり $W_d = 0$  となる特性はシグナルスコアの計算に寄与しない）事が知られており、本検討においても重要な特性の自動抽出が可能である。

## B. 特徴抽出アルゴリズム

ここでは、本検討における特徴抽出アルゴリズムの詳細を述べる。

特徴抽出アルゴリズムの入力データは、

- ・ レセプト系列集合
- ・ 当該レセプト系列に対する検出器の出力（シグナルスコア）
- ・ 検出器が学習した薬剤、診療行為、傷病名の各パターンの加重値

である。これらのデータに対し、以下の処理を行うことで「シグナルスコアが大きいレセプト系列に固有の臨床イベント」を抽出する。

#### ステップ1：抽出対象パタンの構成

検出器が学習した加重値が正となっている薬剤、診療行為、傷病それぞれの発生パターンを全てピックアップする。各発生パターンは入力レセプト系列の期間長と同じ桁数の0か1のビットパターンで表わされている（第3章参照）が、期間の前半と後半のどちらに1が多いかで、次のように抽出対象パターンを定義する。

- ・原因薬抽出パターン 薬剤に関してピックアップされた発生パターンで期間の前半に1が多いもの（前半後半同数の場合も含む）
- ・治療薬抽出パターン 薬剤に関してピックアップされた発生パターンで期間の後半に1が多いもの（前半後半同数の場合も含む）
- ・診療行為抽出パターン 診療行為に関してピックアップされた発生パターンで期間の後半に1が多いもの（前半後半同数の場合も含む）
- ・傷病名抽出パターン 傷病に関してピックアップされた発生パターンで期間の後半に1が多いもの（前半後半同数の場合も含む）

例えば、期間長6のレセプト系列に対して、シグナル検出器における薬剤の発生パターン「101001」の加重値が正であれば、これは原因薬抽出パターンとなる。

#### ステップ2：シグナルスコアが大きいレセプト系列における集計

ステップ1で定義された原因薬、治療薬、診療行為、傷病名の抽出ルールに従って、その発生パターンに該当する薬剤や診療行為、傷病に対して、入力されたレセプト系列で検出器の出力（シグナルスコア）が大きいものの中における頻度を集計する。ただし、シグナルスコアが大きいレセプト系列における頻度とは、シグナルスコアが特定の閾値を越えるレセプト系列内での頻度ではなく、シグナルスコアが1のレセプト系列における頻度は1件、シグナルスコアが0のレセプト系列における頻度は0件、シグナルスコアが0.5のレセプト系列における頻度は0.5件というように、スコアに応じてソフトカウントしたもの（シグナルスコアpのものをp件とカウントしたもの）を集計するものとする。

例えば、シグナルスコアが0.8のレセプト系列において、薬剤Aが上記の原因薬抽出パターンに合致していれば、この薬剤は当該レセプト系列において原因薬として0.8件発生していたとカウントされることになる。



上記の手順で、レセプト系列集合に存在する全ての薬剤、診療行為、傷病名に対して、ステップ1で定義された抽出ルールに従って、シグナルスコアが大きいレセプト系列における原因薬としての発生件数、治療薬としての発生件数、診療行為としての発生件数、傷病名としての発生件数として集計される。

#### ステップ3：シグナルスコアが小さいレセプト系列における集計

ステップ2と同様な方法で、原因薬、治療薬、診療行為、傷病名の抽出ルールに従って、その発生パターンに該当する薬剤や診療行為、傷病に対して、入力されたレセプト系列で検出器の出力（シグナルスコア）が小さいものの中における頻度を集計する。ここでも、シグナルスコアに閾値を設定して条件を見たすものを数えるのではなく、シグナルスコアが0のレセプト系列における頻度は1件、シグナルスコアが1のレセプト系列における頻度は0件、シグナルスコアが0.5のレセプト系列における頻度は0.5件というように、スコアに応じてソフトカウントしたもの（シグナルスコア $p$ のものを $1-p$ 件とカウントしたもの）を集計するものとする。

#### ステップ4：組み合わせに関しても同様に集計

原因薬、治療薬、診療行為、傷病名の組み合わせに関しても、ステップ2、3と同様にシグナルスコアの大きいレセプトにおける頻度、シグナルスコアの小さいレセプトにおける頻度の集計を行う。ここで組み合わせというのは、特定のレセプト系列において薬剤Aが原因薬抽出ルールに合致し、同じレセプト系列（の同じタイムスロット）において薬剤Bが治療薬抽出ルールに合致したときに、原因薬Aと治療薬Bの組み合わせが当該レセプト系列で頻度集計の対象になるということである。

原因薬、治療薬、診療行為、傷病名の組み合わせ方は多数あるが、ここでは副作用仮説発見の可能性検討という目的を鑑み、「原因薬、治療薬」、「原因薬、治療薬、診療行為」、「原因薬、治療薬、診療行為、傷病名」の3通りの組み合わせ方だけを集計の対象とした。

また、組み合わせに関して集計を行う際は、集計対象となる薬剤や診療行為等の組み合わせ数が膨大な数にのぼるため、計算量や必要な記憶容量を節約するための処理も行った。具体的には、集計対象となる組み合わせの数が10万を超える場合には、シグナルスコアが大きいレセプト系列における頻度が一定値を下回るものを無視するという処理である。これにより、シグナルスコアが大きいレセプトで稀な臨床イベントの組み合わせは集計処理の対象外となり、特徴抽出の結果にも現れなくなる。

#### ステップ4：特徴度によるソートと出力

ステップ2、3あるいは4の結果を用いて、各原因薬、治療薬、診療行為、傷病名あるいは、それらの組み合わせの特徴度を計算し、それを用いたソートと結果の出力を行う。特徴度としては、集計された頻度や入力データ数などを用いた任意の数式を採用することができるが、本検討においては「シグナルスコアが大きいレセプト系列に、どれぐらい偏って現れるか」を直感的に定式化した頻度比（シグナルスコアが大きいレセプト系列中の頻度を、全レセプト系列中の頻度で除したもの）を基本的に採用した<sup>8</sup>。

例えば、原因薬Aと治療薬Bと診療行為Cの組み合わせの、シグナルスコアが大きいレセプト系列における頻度が8.3件、シグナルスコアが小さいレセプト系列における頻度が1.7件であったとすると、頻度比は $8.3 / (8.3 + 1.7) = 0.83$ となる。この値で原因薬や治療薬など、あるいは組み合わせそれぞれの抽出結果をソートして出力する。

以上

---

<sup>8</sup>頻度比以外にも、シグナルスコアの大小と頻度の高低に関する独立性検定統計量や、原因薬や治療薬単独の頻度と、それを組み合わせた頻度の独立性に関する検定統計量なども採用してみたが、副作用仮説発見効率などの面で定性的にとりたてて大きな差はなかった。逆に単純にシグナルスコアが大きいレセプト系列における頻度そのものを特徴度とソートに使うことも行ったが、高頻度な臨床イベントが上位にランキングされるにとどまった。