

1, ..., M) と更新する。

3. $G(x) = \text{sgn} \left[\sum_{m=1}^M \alpha_m G_m(x) \right]$ を求める。

シグナル検出では、 $G_m(x)$ として PRR、BCPNN、GPS など考えることになる。この定式化では、 N 個のトレーニングサンプルにより、 $G(x)$ の係数を逐次的に決定していくことになる。シグナル検出においては、まず、シグナルかどうかの分かっているサンプルから統合した判別関数 $G(x)$ を作り、その後、 $G(x)$ を用いた判別の結果シグナルとなったものおよびならなかったものの情報を取り込みながらシステムを改善していくことになる。

D. 考察

シグナル検出のデータマイニング的アプローチを語る前に、データマイニングが既存の統計手法とどのように異なるのかあるいは同じであるのかを見ておきたい。

データマイニングは、データベースからの知識発見のための方法論であり、統計学、データベース技術、人工知能、機械学習、パターン認識、情報処理などの学問分野、そして金融、マーケティング、医学など種々の応用分野の境界に位置する。それらの研究者や実務家間それぞれの専門的な立場からの総括があり得るが、以下統計の立場から考察する。

まず、統計解析には次の 3 つのスタイルがあることを確認する：

(1) 計画された実験 (Designed

Experiments)

臨床試験、前臨床の動物実験、工業実験

(2) 観察研究 (Observational Study)

疫学研究、社会学的調査、経済・経営分野

(3) データマイニング (Data Mining)

薬剤市販後調査、ゲノム、POS データ等

これまでの統計解析の流れは次のように要約される：

1. 研究目的の明確化
2. 標本調査や実験の計画
3. 注意深いモニタリングによるデータの収集
4. コンピュータへのデータの入力
5. データの予備的な解析
6. 評価 (統計的推論)
7. プレゼンテーション
8. 実践

それに対し、データマイニングの流れは次のようである：

1. 研究目的の明確化
2. データベースの選択
3. データの予備的な編集作業
4. データの予備的な解析
5. 評価 (知識発見)
6. プレゼンテーション
7. 実践

これらから次のようにまとめることができる：データマイニングは、これまでの歴史と伝統のある統計分析に代わる方法論では断じてない。統計家の目から見れば、統計手法の適用範囲が広

がったともいえる。しかし、これまでの統計のセンスでは立ち行かない問題も多くはらんでいて、きわめてチャレンジングなトピックである。

○ 統計解析とデータマイニングの類似点

・データベースから有用な情報を取り出す

- ・適切な（統計）理論に基づく
- ・適切なソフトウェアの利用
- ・当該分野の知識がきわめて重要

○ 統計解析とデータマイニングの非類似点

・データの量（伝統的な統計解析では、多変量といっても高々10変数、大標本も数百程度のデータ数）

・データの収集法（実験計画法と標本調査法はよい統計解析のためにきわめて重要とされる）

・仮説（データマイニングは仮説の創生を目的とする）

データマイニングが既存の統計手法と異なる最大の点は、変数数およびデータ数の量である。また、統計解析では重要な位置を占めるデータ取得の計画段階がデータマイニングにはないことも忘れてはならない特徴であるといえる。さらには、統計解析ではデータは既にある程度の解析に耐え得る形式で提供されるのに対し、データマイニングでは、データの前処理に多大に時間とエネルギーがかかるのが普通である。これらを念頭においてシステム作りをしなくてはならない。また、マイ

ニングツールの構築は、とても素人ではできないことも重要なファクターである。

E. 結論

シグナル検出のシステム作りにおける考慮点・留意点は以下のようにまとめることができるであろう。

(1) Global vs. Local

方法論は world-wide である必要がある。しかし、日本には日本特有の事情がある。日本では製薬メーカーを通じた報告が主であり、それは詳細で正確な情報をもたらしている。また、医師の所見を始めとするテキスト情報も重要な位置を占める。ところが、メーカー間に濃淡があることも否めない。

(2) 表示法

まず：グラフ化を考えるべきである。見易いというのが重要なファクターである。次に数値的情報が必要となる。グラフよりも詳細情報を含み、グラフによって大雑把に得られた情報がここで精緻化されるであろう。さらに文書情報も必要となる。ここではきわめて詳細な情報がもたらされる。

(3) 改訂と評価のタイミング

副作用情報は日々集積されるものである。ところが、データが増えるのを待つと時間がかかってしまう。なるべく早期に情報提供したいが、不正確な情報提供は却って混乱のもととなる。したがって、どのようなタイミングで

評価すればいいのかは重要なポイントとなる。

データマイニングは、これまでの歴史と伝統のある統計分析に代わる方法論では断じてない。統計家の目から見れば、統計手法の適用範囲が広がったともいえる。しかし、これまでの統計のセンスでは立ち行かない問題も多くはらんでいて、きわめてチャレンジングなトピックである。

これらの諸点を踏まえ、真に有用なシステム作りを目指さなくてはならない。

参考文献

- DuMouchel, W. (1999) Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *American Statistician*, **53**, 177-90 (with discussion).
- Evans, S.W., Waller, P. C. and Davis, S. (2001) Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, **10**, 483-486.
- 藤田利治 (2004) 医薬品の副作用自発報告によるシグナル検出の実用化に向けての検討. 2003 年度厚生労働科学研究費補助金(医薬品等医療技術リスク評価研究事業) 分担研究報告書.
- Freund, Y. and Schapire, R. (1997) A decision- theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, **55**, 119-139.
- Hand, D. J. (1998) Data mining: statistics and more? *American Statistician*, **52**, 112-118.
- Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, New York.
- 岩崎 学 (1999) データマイニングと知識発見—統計学の視点から—. 行動計量学, **26**, 46-58.
- Iwasaki, M. (2004) Signal detection : Statistical approaches. The DIA 11th Workshop on biostatistics held in Tokyo in August 2004.
- 久保田潔 (2001) 自発報告からのシグナル検出 : 英国 MCA、米国 FDA,WHO の新しい手法. 薬剤疫学, **6**, 101-108.
- 渡邊裕之・松下泰之・渡辺 篤、他 (2004) 重要な安全性情報を早期に検出する仕組み : シグナル検出の最近の手法について. 計量生物学, **25**, 37-60.

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表等

岩崎 学 (2004) 欠測データの処理. 品質, **34**, 2, 46-50.

岩崎 学 (2005) 医薬品の市販後安全性情報の分析—データマイニング

的アプローチ. 理大科学フォーラム. No. 248, 30-35.

H. 知的所有権の取得状況
該当なし

2. 学会発表

藤田利治・岩崎 学：医薬品の副作用
自発報告によるシグナル検出の実
用化に向けての課題. 科研費シンポ
ジウム「医薬品の市販後安全性情報
の集積とシグナル検出」. 2004. 7.

Iwasaki, M. : Signal detection - Statistical
approaches. The 11th DIA Annual
Biostatistics Workshop. 2004. 8.

岩崎 学：医学データの統計解析とデ
ータマイニング. BioJapan 2004 特
別セミナー講演. 2004. 9.

岩崎 学：データマイニングの過去・
現在・未来. 第 18 回日本計算機統
計学会シンポジウム特別講演.
2004.10.

岩崎 学：医薬品の市販後安全性情報
の集積とシグナル検出. 科研費特別
セミナー. 2004.10.

岩崎 学：医薬データの統計解析の最
近の話題. 科研費生物統計セミナー.
2004.11.

岩崎 学：Signal detection - Statistical
approaches. 科研費生物統計セミナー
「医薬品の市販後安全性情報のデ
ータマイニング」(東京理科大学・
森戸記念館). 2005. 1.

岩崎 学：自発報告データベースから
のシグナル検出. データマイニング
に関する検討委員会(医薬品医療機
器総合機構). 2005. 1.

添付資料：アンケートへの回答一覧

○ アンケートの方法：インターネットによる WEB アンケート

○ アンケート期間：2004 年 12 月から 2005 年 2 月

○ 依頼と周知法：

(1) シグナル検出のシンポジウム参加者への個別メール。

(2) 日本計量生物学会のメーリングリストによる依頼

○ アンケート項目：

1. これまで本科学研究費の主催によるシンポジウムや研究会に参加したか

2. これまでの安全性情報の集積と活用の問題点

3. これからの安全性情報の集積に関する要望

4. シグナル検出のシステムについての要望

5. シグナル検出システムで障害となると思われる点（克服法も）

○ 回答者：22 名（主として製薬メーカーの安全性担当者）

○ アンケートに寄せられた回答一覧

1. これまでシンポジウムや研究会に参加したか

開催したシンポジウムの参加者にメールを送ってアンケートへの回答を依

頼したこともあり、回答者 22 名中 21 名がシンポジウム参加者であった。

2. これまでの安全性情報の集積と活用の問題点

・Data の品質、活用のための知識不足、時間とマンパワー不足

・収集担当者が収集目的を認識していない。なんでもかんでも調べようとしている（現実にそぐわない）。

・弊社では、副作用情報をシステムにて管理しているものの、十分なデータベース化ができておらず、シグナル検出を含む、副作用データの活用については模索段階です。

・製品の特定が困難なため、うまく活用できなかった。

・自発報告に関しては、評価に十分な情報の収集が難しい。詳細な情報の提供に関して医療機関の協力が得にくい環境にある。

・集積：医療関係者（医師）からの自発報告が情報の主体になるが、発現した重篤な疾患がどの位の割合で報告されているかが不明。

・活用：情報の最終的な活用目的がエンドユーザーである患者であるとする、国（厚労省）－製薬業界団体の枠に縛られる（逆に言えば、もたれ合い）ため、限られた情報しか活用されていない。

・弊社では市販後部門における（グラフ・集計表作成以外のことができる）解析担当者がいない

・企業の状況と言う観点でお答えします。1) 情報の収集内容、電子化の方

法（項目設計、コーディングを含む）が不均一。原因としては、規制要件の変化、業務・体制上の制限、IT技術の進歩によるシステム改変等色々です。

2) 入力データのQCが不十分。企業のデータベースが、実は当局への報告や進捗管理のための道具に過ぎず、解析用データの様な均質なデータ蓄積を目指していないという側面があります。

3) 活用するには、相当なデータ処理スキルが必要。上記の問題に加え、解析用データにくらべ、データベース構造が格段に複雑である点も障害となります。さらに、今後は個人情報保護法への対応は如何に有るべきかも問題です。過去に患者本人の同意も無く収集した情報は活用していいのでしょうか？

- ・自己の問題として、データ集積の努力に見合った解析ができていない。

- ・企業から重篤、未知で軽微でない副作用・感染症の情報が集約され、副作用情報として公開されています。これはこれで意義あることですが、どうしても対応が遅れてしまいます。

- ・個別症例についての情報はリッチである反面、集積として見た場合、バイアスが大きく、特にレポートバイアスは看過できない。各社、情報収集手順、MRの教育により、努力を行っている場合もあるが、医師サイドのレポートバイアスを排除することはできない。また、一企業で特定分野（疾患、薬剤）について多くの情報を収集できることは稀であり、パブリックなシステムが必要である。

- ・市販後の知識は少ないですが、集めた安全性情報からシステムティックにシグナルを見つけ出すことがなされていないかと思えます。

- ・ネガティブイメージ：企業の受身的対応と当局の消極的対応（機械的・膠着的・表層的なアプローチと運用）広範かつ多面的な分析がなされなかった

- ・企業が収集している情報と、行政当局が収集している情報に差があるように感じられる。また、同一成分の情報が各企業に分散し、有効な活用が出来ない。医療機関からの安全性情報についても、医師個人の興味や学会の流れなどにより収集情報に偏りが感じられることが多い。

- ・これまでは当局対応を主体として情報の収集が行われてきており、結果として膨大な情報が集積されているという状況にある。今後は、有益な情報を迅速かつ適切に抽出できるよう、データ入力ルールも含め、意思の統一が必要と思われる。

- ・規制当局は、新薬承認後すべての薬剤について、市販後調査の6年分の生データを収集しているが、データベースとして、研究用にも活用しているとは、まったく思えない。

- ・薬による有害事象（と思しき場合も含む）が発生した場合、AEシートによって詳細を報告する形を取っているが、医師側の認知不徹底による報告怠慢、また医薬品メーカーからのフィードバックの遅滞等が大きな問題と感じる

- （特に前者。手間がかかるので、医師はあまり報告したがる傾向にある

のでは?)

・安全性情報を目的をもって集積されていないので、いざ解析をしようと思おうと限定された用途にしか使用できない。また、薬効分類及び薬剤名を過去のデータを含めて統一することが必要であると思う。

・産官学の連携が弱いのではないのでしょうか? シグナル検出では、米国 Phrma と FDA のワーキンググループのような活動が望まれます。

3. これからの安全性情報の集積に関する要望

・ICH3 極を含めて、統一性、互換性を確保していただきたい

・目的を限定する

・シグナル検出を行うために、データベースをどのように整備していくべきかを知りたい。

・重複した情報を除外するようにお願いしたい。

・医療機関からの報告をもっと推進してほしい。

・医療機関 (または、医師個人) の厚労省報告の (ペナルティを課した) 義務化によって、少なくとも重篤な症状の発症は、政府が国家として把握する姿勢を示して欲しい。

・治験なみの DM が必要

・企業が個々にデータを保持するのは、そろそろ時代遅れではないかと思いません。企業から運用資金を調達して、少なくとも日本国内の情報については一元管理する組織、データベースの運用は出来ないか? と思います。また、法

規制による報告制度による情報収集のみでは、重篤例に偏る点が情報活用という点では価値を下げると考えます。

・解析目的によっては、ビタミン、輸液等の主要でない併用薬のような詳細情報は集積不要とする方法論についても、方向性を判り易く示して頂ければありがたいです。

・新薬については、治験情報では不十分であり、市販後に重篤な副作用が見つかる場合があります。したがって、新薬については、因果関係とは無関係に有害事象としてバイアスを含まない情報を集約し、シグナル検出に活用した方がよいと思います。

・ICH の報告対象規定 (重篤、因果関係) にとらわれず、バイアスを少なくかつ広く (因果関係を問わず) 収集する制度が必要である。また、コーディングルールの統一 (MedDRA の功罪) も必要である。

・柔軟でタイムリーな情報収集体制と up-to-date な解析法の適用、各企業で収集したデータの有効な活用方法の考案

・情報が氾濫しても活用できない。集積した情報から何が重要であるかを見極められるか否かが必要。

・データ入力ルールの徹底が求められる中で、特に、情報源の適切な分類は必要不可欠である。例: 自社品が、他社における調査・試験で対照薬として使用された場合など

・企業から提出されている、安全性定期報告や再審査申請等の市販後調査報告を行政がデータベース化して、公開する事も必要と思われる。

・これこそが、科研費を用いた研究テーマだと思います。研究用としても使えない何かの事情があるのでしょうか。私はあるとまったく思いません。

・情報のオープンさと誰でも利用可能なこと。情報公開のスピード、公平性への配慮。消費者（ユーザ）保護の原則にて活動いただきたいです。

・1) 迅速なフィードバック 2) より簡便でアクセスしやすい報告システムの構築

・安全性情報で使用する項目を特定し、集積する。データの品質水準を定義し、ある程度均一の水準を確保する。用語及びコードの統一

・医療機関からの直接報告の質を高めることかと思えます。

4. シグナル検出のシステムについての要望

- ・互換性、平易性
- ・分母のないシグナルの検出よりも、分母のある集団使用成績調査などのシグナル検出の方がもっと大事
- ・統計学を専門に学んだ者でなくても使えるシステムになるのでしょうか。自社単独の副作用情報の集積量で検出が可能か。不可能ならば、どのような手立てがあるのか。
- ・何らかの形で被疑薬とされた製品を特定できるものにして欲しい。情報保護の観点から困難かとは思いますが、一般名と製品コードを示す等検討して頂きたい。
- ・現在米国 Lincoln 社と契約して FDA-AERS データベースに登録されて

いる薬剤のシグナルスコア

(EBGM/PRR) を算出できるシステムを利用しているが、日本でもこのようなシステムを構築して欲しい。

・現在の数学的処理の手法は、多国籍企業によって販売される大型医薬品について適用ができるものの、情報が少ない医薬品についてシグナル検出することはできない。何か、よい手法はないでしょうか。

・シグナルディテクションが各社で可能となるように、データベースの公開（使用料支払い可）

・設問の意図が良く分かりませんが、PMDA が検討しているシステムについての要望、というふうに解釈してお答えします。GPS のような、計算プログラム作成自体が困難な手法は採用して欲しくないと思います。例えば、将来「GPS でシグナルが出たからコメントを提出せよ」と PMDA に要求されたとしても、まともに対応できる企業は日本にどれほど有るのでしょうか。

・ぜひ、同システムについての公表・教育の場をお願いいたします。

・データマイニングの具体案を知りませんのでコメントできません。

・グローバルなシステム（データ、手法）であること、検出した後に、次のステップとして、恐らく個別の症例を見直し事となるが、その時、対象となった自社のものでないデータについてもレポートを取り出せるようなシステムを望む。

・副作用は保守的に検出できるシステムの方が良いと思います。

・自発報告だけがシグナル検出の対象になるのかどうか？自発報告に限らず、入手可能なすべての安全性情報をシグナル検出の対象にすべき（アプローチは検討の余地あり）

・情報提供が必要不可欠であるので、医療関係者の協力が必須。

・感度および特異度の算出するための指標が必要である。（MedDRA用語を品目問わず添付文書用語と対応させることは難しく、また、各事象の既知・未知を活用する場合でも、品目担当者間での評価のバラツキが散見されるため、感度・特異度を適切に求めるのには不適と考えられる。）

・企業が収集するデータには、限界があるので、公の機関が主導を取り、広く情報を収集する事が必要だと思う。また、このシステムで検出される（と思われる）シグナルには、旬があると思うので、タイムリーな情報公開が必須と思う。

・これまで、何回かプレゼンを聞いて、なるほど有益であると認識を新たにしました。しかしながら日本国内のデータへの小さな適用事例も示されないのは残念でなりません。

・情報公開のスピード、誰でも普通に、かつ無料でアクセス可能なことを希望します。

・アクセスしやすいシステムの構築に尽きるのでは？

・シグナル検出の手法が各団体で異なるので手法の統一、あるいは適用範囲を明確化し、これまでは予測できなかった（予測するのに時間がかかった）

関連性を検出できるような体制が必要であると思う。

・シグナル検出後の対応をどうするのかにつきます。日本は疫学データベースが存在しないため、ケースコントロールやコホート研究が低調です。現状ですと、シグナル検出後、詳細検討で注意喚起の可否を検討になりますが、臨床疫学的な検討もできる環境作りが望まれます。

5. シグナル検出システムで障害となると思われる点（克服法も）

・シグナルと認定するための方法論（ガイドラインの発行が必要と思われる）
安全性の部門に統計学、データマネジメント等のシグナル検出に必要な人材がいないか、少ない(規制をかけてでも人材の補充や育成を急ぐ必要がある)
・有害事象としてはPTコードでは細かすぎるし、SOCでは大まかすぎる。と
いって、HLTやHLGTが妥当な分類とも言えない。併用効果の探索で薬剤コードを使用する場合、MedDRAよりは
ましたがシグナル検出のための分類が
きっちりとなっていない

・古い症例と新しい症例ではデータの質が異なっている。MedDRAコーディングのばらつき。一社で保持する症例データの量の少なさ。使用者側の統計の専門知識の低さ。

・副作用用語の統一（MedDRA）と用語の読替の部分で、企業間に差が出るのではないかと思われる。

・シグナル検出を有効に活用するための大規模なデータベースがまだ集積さ

れていない。企業や医療関係者からの個別症例報告、企業の定期報告も含めたデータベースを作成してほしい。

- ・情報量が少ないドメスティック医薬品をどう扱うか。

- ・医学的判断の関与をシステム化できないか

- ・障害は思い付きますが、克服法を提案できる段階ではありません。よって、問題点のみお答えします。1) システム設計、データベース等への実装上の問題：分析対象集団の設定方法、薬剤、有害事象のグルーピング方法、シグナルが検出た有害事象について、新規性の有無を機械的に判定する仕組みと、その正確さを維持する仕組み シグナルが検出た有害事象が、未知である場合 既知の有害事象に対する、新たなシグナルである場合、ユーザインターフェイス設計、システム化後のバリデーション。2) その他の問題：「発生傾向の著しい変化」を如何に定義し、検出するか。

- ・重大な有害事象症状に関しては、その絶対数が注目され、データベース中の多寡を云々しても意味が無いのではないか。

- ・重篤な副作用に繋がる危険信号を十分に分析し、それらを医学的に関連付ける必要があると思います。

- ・医師（既知、未知、疾患領域、重篤の解釈）、報告企業等によるレポーティングバイアスが最大の障害であり、次に販売期間、頻度の大きな特定の事象の存在が障害と考えられる。制度面の対応として、医療機関側への報告制度

（当局から医療機関への指導、監査、服薬指導料のペナルティ等）の改善等。システム面の対応として、臨床検査を伴う事象、疾患分野で頻度の高い事象など、事象の検出、報告過程による分類、薬剤毎に、治験時に最も高頻度であった事象のレポーティングを標準指標としてデータベースの信頼性を評価するような工夫。

- ・安全性部門に統計の専門家を配置できるだけのマンパワーが社内にあるか心配です。

- ・過度な期待と機械的・膠着的な運用 ⇒常に性能評価を行いシステムの地道に性能改善に努めていくべき（予算化必要）

- ・今後、「有害事象」という観点で収集された情報と、「消費者情報」が収集されることになると、基本データの信頼性が揺らぐ可能性がある。

- ・因果関係判定の精度。客観的に判定できるか否かがポイントと考えます。

- ・検出されたシグナルに対して、これまでの集積評価とは異なった視点で、かつ人的評価の部分を最大限省いた形で適切に評価できるか？

- ・データベースの構築。医療現場からの情報提供の義務化。あまり過激では困りますが、意識改革。

- ・日本国内のデータでなければ、価値がありません。規制当局に出しているデータは一覧表程度のデータで、シグナル検出の研究用には使えません。一覧表程度のデータでも、更なる研究のためのデータの再集積のためにも役立ちます。

・各種営利団体間での利害が問題にならないかと思えます。対策として、公平な機関（ユーザ重視の立場に立った）を設立して運用という方法を取ってはどうかと思えます。

・現場からの報告がなければ、検出は難しいが、現場の認識は薄い。理由の一つとして、医師卒前教育課程において、医薬品認可に関する教育はほとんどなされていない、もしくは市販前に偏っているという点が考えられる。今後は現場にて密接に、また真っ先に関わる事になる市販後調査の話題を何らかの科目で取り上げる事が必要なのではないか。

・市販後データは治験ほど管理されて集積されていないので不明確なデータもあると考えられるので、最低限の集積するデータを特定し、意図をもって集積していくことが必要であると思う。

また、過去集積されたデータをどのように活用していくかは重要な問題であると思う。

・いつどのタイミングで実施し、誰が結果を検討するのかという問題です。結果を企業へ送付するだけですと、企業にとって大きな負担となるでしょう。上記活用を考えている場合は、判定基準をFDA並に高くするべきと考えます。

