

医療分野での意思決定を支えるデータサイエンス —VMS の活用を中心に—

紀ノ定 保臣

岐阜大学大学院

医学系研究科 医療情報学分野

連合創薬医療情報研究科 医療情報学専攻

1. はじめに

診療業務および病院運営を進める上で、①意思決定支援、②診療支援、③マネジメント(経営・管理)支援などの機能が重要な役割を果たす。一方、これら支援機能を充実するためには、医療情報システムなどに既に蓄積された膨大なデータの活用や各種ガイドラインの順守、理論的な解析等から導かれる知識等を効率的・効果的に利用できる環境の整備が不可欠である。診療データ等の二次利用の必要性が叫ばれる今日、このような環境を構築するためには技術的側面から DWH(Data Warehouse)の機能と役割について整理し、DWH を積極的に活用するための方法論や手法、方向性について議論することは大変有意義なことである。

本稿では、病院内での TQM(Total Quality Management)活動や問題解決、知識発見の観点から DWH の機能と役割、具備すべきツールや意思決定を支えるデータサイエンスについてまとめる。

2. 診療データを含む DWH の特殊性について

診療データは患者毎に経時的に発生するデータであり、基本的には木構造である。また、患者毎にデータ項目毎の発生頻度が異なるため、一般的な DWH 的データ処理を行なおうとする場合には欠損値の扱いが重要な問題となる。したがって、DWH の設計においては保存するデータ項目の選択もさることながら、どのように時系列データを管理・保存・処理を効率的に行なうかが最重要事項となる。さらに、処理の中には格納されているデータを検索する場合、表示する場合、個別データを処理する場合、大規模データセットとして処理する場合等、でその困難さ

あるいは効率的な対応は異なる。このような診療データを含む DWH の特殊性を考慮した形で、その構築と運用を行なうことが重要である。

3. 一般的な問題解決活動の手順

一般的な問題解決活動の手順を図 1 に示す。同図より明らかなように、問題解決には事実をデータとして収集し、データの処理と結果判断、対策行動の順に手順が示されている。同様に、EBM(Evidence Based Medicine)的な考察をする場合にも、問題の定式化、情報収集、批判的吟味、診療への適用という4つのステップを踏む。このような問題解決手順と DWH の機能を対比するとき、DWH が果たすべき役割は事実データ(Fact Data)を蓄積し、データ処理機能を提供し、意思決定に活用可能な情報として提供することである。換言するならば、DWH はこのような役割を果たす機能を持ったシステムとして構築・運用されるべきであり、診療業務で用いられる一般的な DB(Data Base)とは明確に区別されるべきものである。

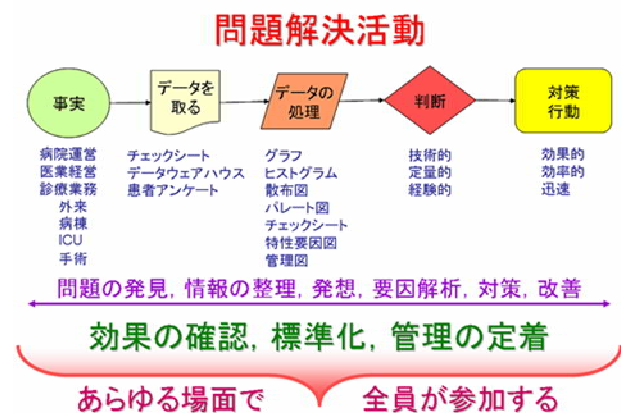


図 1. 問題解決活動から見た手順とそこで利用される手法や視点等

4. MyEHR: DWHとしての統合データバンク

MyEHR は、複数機関からの特定健診データ、健診データ、診療データ(病名、検査値、処方データ、など)を HL7-CDA 等の標準化された XML 形式の文書で電子的に受け取り、個人毎に突合、生涯に亘る健康情報として長期的に保存・管理・大規模母集団健康データの分析、個人毎の時系列データ分析、保健指導とその効果等の管理を効率的に行なうことを目的に開発したシステムである。図2に MyEHR の基本的な構成要素とその運用概念図を示す。



図2. MyEHR の構成・運用概念図

MyEHR は標準的なデータ交換規約をサポートし、医療機関や健診センター、健康保険組合などの医療保険者から、また病診連携ソフトなどを介して HL7-CDA などの XML 形式の文書を受信する機能を有する。また、受信された XML 文書を保存・管理するデータベース(DB)システムと格納されている健康情報等を Web を利用して参照・活用できる Web 型健康情報参照機能を装備した。Web 型健康情報参照メニューには、健康情報等を格納している個人・患者が自身の

データを参照・活用する場合と、医師あるいは MyEHR 管理部署が格納されている全データを参照・活用する場合とを想定して各種機能を用意した。さらに、健康情報の参照や利活用の利便性を高めるため、健診・診療情報等データベースの他に、市販医薬品の添付文書情報(JAPIC から提供)や公表されている医薬品副作用情報等データベース、健診・診療データの解釈を支援する各種知識情報 DB を装備した。詳細は以下である。

① 市販医薬品の添付文書情報(JAPIC から提供)をデータベース化し、MyEHR 上での多面的な利用を可能にした。

② 医薬品副作用情報は米国 FDA が有する Adverse Event Reporting System (AERS) をデータベース化、参照可能にした。

③ MedDRA/J 10.0 (ICH 国際医薬用語集日本語版 Ver.10.0)を実装し、有害事象や病名の辞書として使用できる環境を構築した。

④ AHFS (American Society of Health-System Pharmacists) Pharmacologic-Therapeutic Classification 2007 (アメリカヘルスシステム薬剤師会編集医薬品集の薬効分類)を実装し、薬効分類と医薬品(一般名)との関係データとして利用する。

⑤ Drug Bank を実装した。4,900 の薬剤について一般名、商品名、薬物体内動態、構造式、相互作用情報、作用点情報などがデータベース化されている。

5. MyEHR の動作環境

本研究で開発した MyEHR の構成は以下である。

- 1) OS: Windows Server 2003 R2
- 2) XMLDB: 自主開発
- 3) RDBMS: MySQL 5.0.16
- 4) Web サーバ: Apache_2.0.52
- 5) 開発環境: Visual C++ 2005 / C++6.0
- 6) ハードウェア: HP Proliant ML150
- 7) MyEHR: 自主開発
- 8) 解析には VMS (数理システム) を利用

6. データサイエンスの活用例

一般利用者が MyEHR を利用するとき、上記のような健康情報や専門的なデータを単に参照させる機能だけでは適切に理解・判断できるとは言い難い。特に、自身の健康データから生活習慣の改善と疾病予防へと行動を連鎖させるためには、自身の健康データの位置付けとその変化の方向性を容易に理解できる支援機能が不可欠である。

健康データの効果的な提示方法を検討するため、MyEHR に格納されていた肝疾患のデータについてデータ抽出(29名分)、分析、データの可視化を行なった。29名の内訳は健常者13名、肝炎患者8名、肝硬変患者8名である。血液検査結果からAST, ALT, LDH, ChE, アルブミン, ビリルビン, γ -GTP, ALP, 血小板の9項目を選択した。これらデータに対し要因分析法を応用し、主な3要因を抽出した。さらに、寄与率の大きい2要因(因子3と因子2)を用いて因子得点散布図を作成した。その結果を図3に示す。散布図上の「1」は肝硬変患者群を、「2」は肝炎患者群を、「3」は健常者群を示す。同図より明らかな如く、因子3-因子2からなる二次元平面上の散布図において、健常者群、肝炎患者群、肝硬変患者群と重症度が増すにつれてその分布は右上の方向に移動した。このようなデータの提示方法は、9項目の検査値を2次元平面上の2因子の関係にマッピングしたものであり、かつ重症度に応じてデータが分布する位置関係が容易に理解できる利点があった。

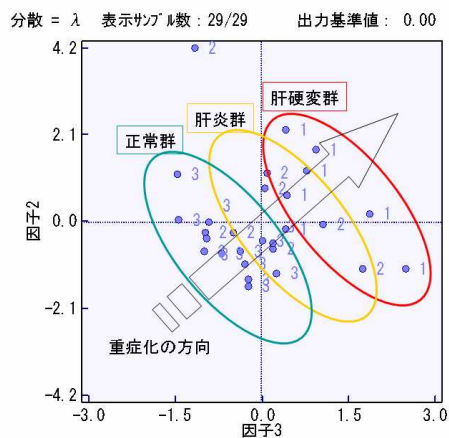


図3. 健康データの提示方法例

7. 問題解決活動と管理・改善

DWH に蓄積されたデータを用いて、データサイエンス手法を駆使しながら各種の問題を解決・意思決定する場面は多くある。この時、どのような姿勢で問題解決に取り組むか、あるいはデータを通してどのように管理・改善を行なうかという手順を明確にすることは重要である。図4に問題の構造化を示す。

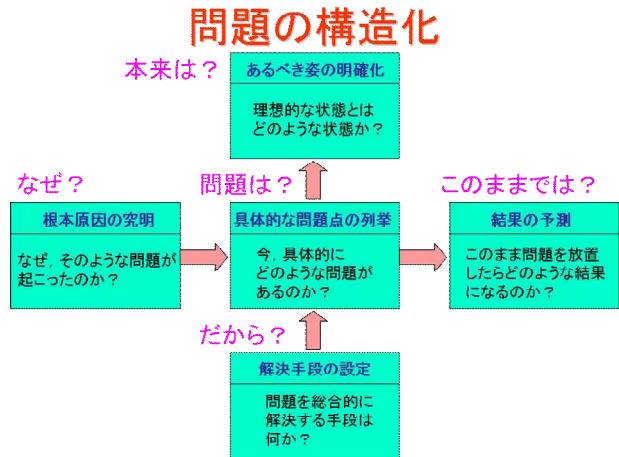


図4. 問題の構造化

問題(本来あるべき姿と現状とのギャップ)がある場合、何故そのような問題が起こったのかを明らかにすること(根本原因の究明)が求められる。また、このまま問題を放置すればどのような結果になるのか(結果の予測)を知ることも重要である。一方、問題を解決するためには本来のあるべき姿を明確にすると共に、現状とのギャップを埋める手段の設定が必要になる。前述のDWH やデータサイエンス手法は、このような場面に有効に機能することが求められる。

同様に、管理・改善する場面でも同様の機能が求められる。管理とは現状の維持であり、管理水準が設定されなければならない。

- ① 管理の手順は以下である。
 1. 管理水準を決める
 2. 作業標準を決める
 3. 教育・訓練を行なう
 4. 作業を行なう
 5. 作業の結果をチェックする
 6. 異常に対する処置をとる
 7. 処置の結果を確認する

一方、改善とは現状の打破である。

② 改善と管理の手順は以下である。

1. 問題点を見つける
2. テーマと目標を決める
3. 協力体制を作る
4. 現状を把握し、解析する
5. 改善案を作る
6. 改善を実施する
7. 効果を確認する
8. 歯止めする

問題解決活動と管理・改善について述べた内容と手順を、MyEHRに応用するならば、以下のことが言える。

- データを表示する場合には、現在の健康状態を示し、同時に本来あるべき望ましい状態との関係が容易に理解できることが必要である。
- 現在の健康状態は維持・管理すべき状態か、あるいは改善すべき状態かが容易に理解できることが必要である。
- もし問題がある場合、望ましい状態に移行させるためにはどのような手法・手段があるのかを示すことが必要である。
- 改善の効果が確認できることが必要である。

8. データサイエンスの発展と活用への期待

医療分野のIT化が進展し、従前にも増して大量のデータが発生、蓄積されつつある。また、このような大量のデータを活用しようとする機運が高まりつつある。この時、どのようなデータ分析環境が求められるのかを考察する。

① 大量のデータが処理できる環境でなければならない。

② データの分析には、その手順が具体的に可視化され、同時に用いられる処理手法と処理前後のデータが容易に確認できることが不可欠である。

③ 問題が発生した場合、その問題に対する原因究明に利用できるデータの分析手法が豊富に用意されていることが望ましい。

④ 問題解決手段を複数個設定する場合、

その効果判定や感度分析が容易にできる環境でなければならない。

⑤ データに対して、外れ値や欠損値に対する対応が柔軟でなければならない。

⑥ データを可視化する豊富な機能が用意されていないなければならない。

⑦ データ分析途中での中間データを管理する機能に優れていないなければならない。

⑧ 時系列データの処理が容易でなければならない。

⑨ 構造化されたデータの処理が容易でなければならない。

⑩ ネットワーク図の作成と分析が容易でなければならない。

⑪ 優れた GUI と処理の高速化が不可欠である。

⑫ 上記機能などが統合機能として実現されていないなければならない。

9. まとめ

医療分野での意思決定を支えるデータサイエンスについて概要をまとめた。また、MyEHRを紹介し、生涯に亘る健康情報を保存・管理、分析できるシステムについて、その運用例を紹介した。さらに、今後利用が増大すると思われるデータサイエンス手法や各種ツールに求められる機能について私見を述べた。

データサイエンスは大量データからの知識の発見に話題が集まる傾向にある。しかし、上述のように、医療分野における問題解決や管理・改善という場面においても同様の手法が有効に機能する。また、問題の構造化で述べたような論理的な思考を手助けしてくれる。

医療分野での意思決定を支えるデータサイエンスのツールとして、VMSは十分な機能を備えたソフトウェアであった。筆者等の研究に対し、平素の数理システムの協力に深謝する。

参考文献

[1] 紀ノ定保臣. 米国先進医療IT事情と日本の進むべき方向性についての考察. 新医療 2006;33(11):150-153.