

## 資料

### 医療情報の分析における背景知識の利用

豊田 修一<sup>1)</sup>

キーワード：医療情報、電子カルテ、コンセプト・ハイアラキ、背景知識、データウェアハウス

#### I. はじめに

日本では、ここ十数年で、企業の情報化・システム化が急速に普及してきた。企業では、従来、紙に書いて処理していたデータを電子化し業務の効率化を図るだけでなく、意思決定の迅速化も実現している。例えば、雨の日に最も売れた商品は何か、といった情報は一瞬にしてわかるようになってきている。このような情報は、インターネットなどを通じて、経営者から新入社員まで広く共有されている。データの電子化は、業務の効率化や経営の最適化に貢献している。

一方、医療における情報の電子化は遅れている。医療情報が守秘性の高い個人情報であることなどが影響しているとも言われている。しかしながら、最近では、医療の発展に貢献できるような情報化・システム化の重要性が認識され、医療機関ではオーダーリングシステムや電子カルテシステムの導入がすすみ、業務の効率化にも効果をあげてきている。しかしながら、情報化を意思決定にまで活用している段階には達していない。

また、医療分野でも、電子化された医療情報の蓄積量が膨大なものになってきており、その利用に注目が集まっている。そこで、本稿では、これらの膨大なデータの分析に有効とされる背景知識を構築し、その有効性を考察することとする。

#### II. データベースとデータウェアハウス

医療情報のデータ分析を目的とした背景知識の構築にあたり、データウェアハウス(Data Warehouse)技術に着目した。データウェアハウスは、蓄積されたデータを利用し、知識集約業務や経営管理に必要な意思決定のためのデータ分析に焦点を当てた処理を目的とす

るデータベースである。本章では、データウェアハウス技術の背景と特徴を明らかにするために、業務系データベースと情報系データベースの比較、医療機関のデータベース、データウェアハウスの特徴について述べる。

##### 1. 業務系データベースと情報系データベースの比較

組織で構築されるデータベースには、業務系データベースと情報系データベースの2種類がある。

企業や医療機関などの組織の日常業務からは、多量の複雑なデータが発生する。業務系データベースは、これらのデータを定義できる詳細レベルなデータベースである。業務系データベースでは、定型的な日常業務を効率よく行うために、必要な値を保持しており、日々データが更新される。このため、データ処理単位はトランザクション単位である。このような処理形態は、オンライン・トランザクション処理(Online Transaction Processing:OLTP)と呼ばれる。OLTPは、データを頻繁に参照・更新するため、素早いレスポンスが求められる。業務系データベースでは、データの同期制御(concurrent control mechanism)やデータの障害回復(recovery mechanism)などのデータベース管理システム(Database Management System:DBMS)の機能を十分に利用している。

一方、組織は、意思決定を行わなければならない。効率的な意思決定を支援するためには、長期的なデータ保存が必要となる。情報系データベースは、データの分析を支援する目的で収集・保存しており、経営者やアナリストの利用が多いデータベースである。このようなデータベースは、データウェアハウスと呼ばれ

1) 上武大学看護学部

る。データウェアハウスの設計には、トップダウン的・経営的視点が重要である。また、データウェアハウスを利用して、情報を高速にアクセスし、多様な視点から一貫性のある分析結果を出力するシステムは、オンライン分析処理(Online Analytical Processing : OLAP)と呼ばれる。

このように、業務系データベースと情報系データベースは、データの利用目的、データ構造、保存期間などが大きく異なる。表1は、それらの比較を示す。

表1 業務系データベースと情報系データベース

	業務系	情報系
利用目的	業務遂行	効率改善、意思決定
利用者	担当者	担当者、マネージャ
利用形態	トランザクション処理	分析処理
使用頻度	高い	低い
処理速度	高レスポンス	重要でない
データ更新	有	無
データ内容	現在の状況	過去の傾向・履歴
保存期間	一定期間保存後消去	長期間保存

## 2. 医療機関の業務系システム

医療機関における業務系システムの代表的なものに、オーダーリングシステムがある。オーダーリングシステムにおける臨床検査オーダーを考えてみる。診察室では、該当患者に対する検査オーダーが入力される。このデータは臨床検査部門や医事会計部門に転送される。臨床検査部門では、オーダーリングシステムと臨床検査システムが接続されていることが多い。

オーダーリングシステムでは、ある部門で患者データを更新中は、他の部門でその編集に矛盾するようなデータ更新をすることはできない。これがデータの同期制御機構である。また、検査の大まかな進捗状況は、診察室や医事会計部門でも確認することができる。部門間のデータ共有化が実現される。さらに、データの記録媒体であるハードディスク装置に障害が発生した場合には、バックアップデータを利用したデータ復旧も可能である。データの障害回復機能を提供している。

## 3. データウェアハウスの特徴

データウェアハウスの特徴として、サブジェクト指向(Subject-oriented)、統合(Integrated)、時間従属(Time-variant)、不変性(Nonvolatile)の4つのキーワードをあげ

ることができる。サブジェクト指向とは、特定の対象やその周辺に簡潔な視点を提供することである。意思決定に不要なデータは除去される。統合とは、複数のデータソースを統合して、視点を提供することである。属性名やデータのコード化構造などの一貫性を維持する。時間従属とは、時間経過的な見方で情報を提供することである。それぞれのデータは、時間要素を直接的・間接的に含んでいる。不変性とは、データに対する問い合わせのみで、更新されないということである。履歴データを蓄積し続けている。

データウェアハウスを使用したシステム例に、データマイニングシステムがある。このシステムは、特定の大容量データベースに埋もれている知識を見出すために使用される。しかしながら、データマイニングシステムを利用しての知識の発見には、人間の介入や助言が必要とされる。効率的な知識発見には、探索したい知識の種類、知識発見プロセスに有効な背景知識、発見された知識の視覚化方式などの基本的要素知識・技術が不可欠となっている。背景知識とは、知識発見の過程で、役にたつ可能性がある対象領域の情報のことである。背景知識の一種として、コンセプト・ハイアラキー(Concept hierarchy)がある。

## III. 医療情報における背景知識の利用

本章では、データ分析に有効な背景知識の一つであるコンセプト・ハイアラキーを医療情報の特定のディメンジョンに対して構成した結果について述べる。

### 1. コンセプト・ハイアラキー

コンセプト・ハイアラキーは、多様な視点からデータの集計・分析を行うために、データベースの一部分に対して、低レベル概念から高レベル概念に概念のマッピングの列を定義したものである。分析の視点には、ディメンジョンの軸を変化させる視点と、データの集約・分析の視点がある。ディメンジョンは、データを分析する視点と考えることができる。

選択されたデータベースの部分が同一であっても、ユーザの視点が異なれば、異なるコンセプト・ハイアラキーが構成されることもある。コンセプト・ハイアラキーの主な形式には、スキーマ型とセット分類型がある。スキーマ型構造は、データベースのスキーマにおける属性間の部分的(全体的)順序であり、属性間に存在する意味的順序を表現することが多い。セット分類型構造は、与えられた属性間に対して、定数のグ

ループ化や値の範囲の組織化で定義する。小さな関係を定義する時に使用されることが多い。

次に、コンセプト・ハイアラキにおいて、階層を上下することを考える。より高い概念の階層に移動する操作はロールアップ(Roll up)、より低い概念の階層に移動する操作はドリルダウン(Drill down)と呼ばれる。ロールアップやドリルダウンによって、ユーザは様々なレベルからデータをみることができる。

ロールアップは、意味的にはより高い概念で捉えることであり、データの取り扱いでは視点数を集約することである。一方、ドリルダウンは、意味的にはより低い概念で捉えることであり、データの取り扱いでは次元の追加や概念の詳細化を行うことである。

## 2. 医療情報に対するコンセプト・ハイアラキの構成

医療情報に対するコンセプト・ハイアラキの構成について、診療報酬データの分析と指示データの分析を例にとり、述べる。

診療報酬データを時間の視点と診療科の視点で分析する場合を考える。診療報酬データは、時系列的・診療科別に診療ごとの診療行為別(初診、再診、処方、検査、等)点数として用意する。分析の視点のデータ属性に対しては、定数のグループ化や値の範囲の組織化を定義する。つまり、セット分類型の構造を定義する。例えば、時間データを表現するための値には、“分”、“時”、“日”、“週”、“月”、“4半期”、“年”などがある。診療報酬データの経営的分析を目的にして、時間データを取り扱う場合は、“日” < “月” < “4半期” < “年”といった多層性が合理的である。つまり、図1のようなコンセプト・ハイアラキを構成することができる。

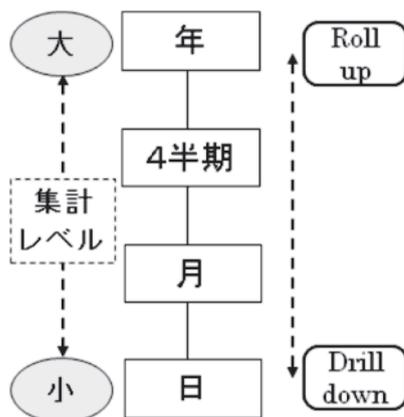


図1 診療報酬データを分析するための時間のコンセプト・ハイアラキ

次に、指示データの分析について考える。指示データの属性には、“患者”、“診療科”、“医師”、“年月日”、“処方単位”、“薬剤名”、“用法”、“検体”、“検査項目”などがある。属性間の関係を考えると、処方指示という視点では、“患者”、“診療科”、“医師”、“年月日”、“処方単位”、“薬剤名”、“用法”、などの属性に意味的関連性があると言える。同様に、検査指示という視点では、“患者”、“診療科”、“医師”、“年月日”、“検体”、“検査項目”、などの属性に意味的関連性があると言える。つまり、指示データに対しては、スキーマ型構造のコンセプト・ハイアラキを定義することができる。図2は、処方指示に関連する属性の意味的順序を表わすコンセプト・ハイアラキの定義である。同様に、検査指示のコンセプト・ハイアラキも定義した。

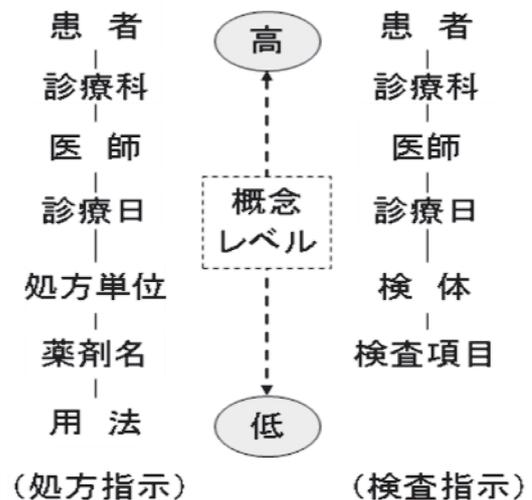


図2 指示データのコンセプト・ハイアラキ

## IV. 考察

医療機関の経営分析では、構成したセット分類型コンセプト・ハイアラキを利用して、情報の集約化(Roll up)や絞り込み(Drill down)などを段階的に実施することができる。つまり、医療機関の経営層が行う知的作業の一つである診療科別・診療行為別の時系列的分析を支援できると言える。さらに、派生的に季節変動要因などの把握にも貢献できる。また、指示データに対するスキーマ型コンセプト・ハイアラキは、指示データを意味的順序に従って再構成することや、指示データを多様な面から把握することに貢献できる。ここでは、処方指示に対するコンセプト・ハイアラキについて、詳しく考察する。

処方指示は、患者記録の中で内容的にも量的にも主

要な構成要素となっている。患者一人一人の処方指示の特性は、時間的には継続性を有する記録であり、内容的には繰り返し・類似傾向が強い記録である。図3は、処方指示を時系列的に表現したものである。紙カルテで運用していた時代や多くのオーダーリングシステムでは、このような表現が採用されている。しかしながら、時系列的表現においては、薬剤の使用状況を辿る場合や、検索する場合など、記録の再利用において効率が極めて低くなっている。さらに、時系列的表現においては、使用頻度の高い薬剤の記述の羅列の中に、使用頻度の低い薬剤の記述が埋もれ、多忙な医師が見逃すリスクがあった。同様に、微妙な変更内容(例えば、多くの薬剤を継続的に処方されている患者に対する1種類の薬剤の用法の変更)も見落とされるリスクもあった。また、データ量が増加するとデータ検索の負担も増大し、その利用効率は急速に低下する傾向にある。

09. 28	内科、 医師 X
	薬B 薬C
10. 12	内科、 医師 X
	薬A
	薬B 薬C
10. 29	内科、 医師 X
	薬B 薬C
11. 10	内科、 医師 X
	薬A
	薬B 薬C
11. 15	整形、 医師 Y
	薬E
11. 24	内科、 医師 X
	薬A
	薬B 薬C
	薬E
12. 08	内科、 医師 X
	薬B 薬C
12. 22	内科、 医師 X
	薬A
	薬B 薬C

図3 処方指示の時系列的表現

コンセプト・ハイラーキは、このような課題を解決する背景知識としても利用できる。図4に示すように、コンセプト・ハイラーキを利用して、処方指示を構成する属性を意味的に3レベルに分類する。最上位レベルには、属性“患者”が含まれる。上位レベルには、属性“診療科”、“医師”、“診療日”が含まれ、診療経過を表現していると言える。下位レベルには、属性“処方単位”、“薬剤名”、“用法”が含まれ、診療ごと

の処方内容を表現していると言える。

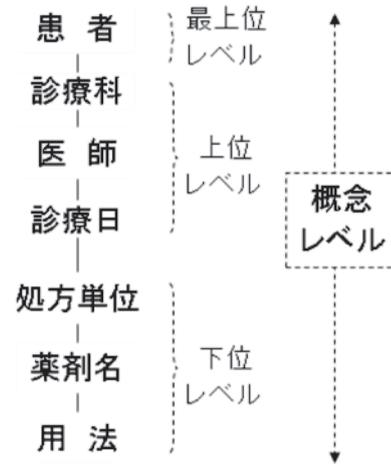


図4 処方指示を構成する属性の意味的な区分

本稿では、この上位レベル属性と下位レベル属性を利用した意味的概念対応表現を提案する。この表現は、図5に示すように、指示データを上位レベル属性と下位レベル属性に分離して表現し、上位レベル属性と下位レベル属性を関連付けるリンク情報を生成する。つまり、患者の診療経緯と処方指示の種類を分離して表現するものである。この表現は、上述のリスク軽減の貢献など、意思決定の支援を実現できる。

コンセプト・ハイラーキの新しい利用法として、指示データの意味的概念対応表現への適用について述べた。これは、領域の背景知識がデータの効率的な内容把握に貢献することを示している。さらに、同一内容の下位レベル属性のデータ表現を1回に集約することで、データ圧縮表現も同時に実現している。

## V. おわりに

最近のオーダーリングシステムの普及により、医療機関においてデータベースという言葉もよく聞かれるようになってきた。医療機関では、医事会計システムやオーダーリングシステムで処理したデータから、データウェアハウスを構築することができる。しかし、蓄積したデータを利用したデータ分析を積極的に行っている医療機関は多くないようだ。

意味的順序を表現するコンセプト・ハイラーキの利用目的は、データウェアハウス中に保存されたデータの背景知識を構築し、データ分析を支援することである。このような利用例として、経営分析におけるコンセプト・ハイラーキの構築がある。本稿では、コ

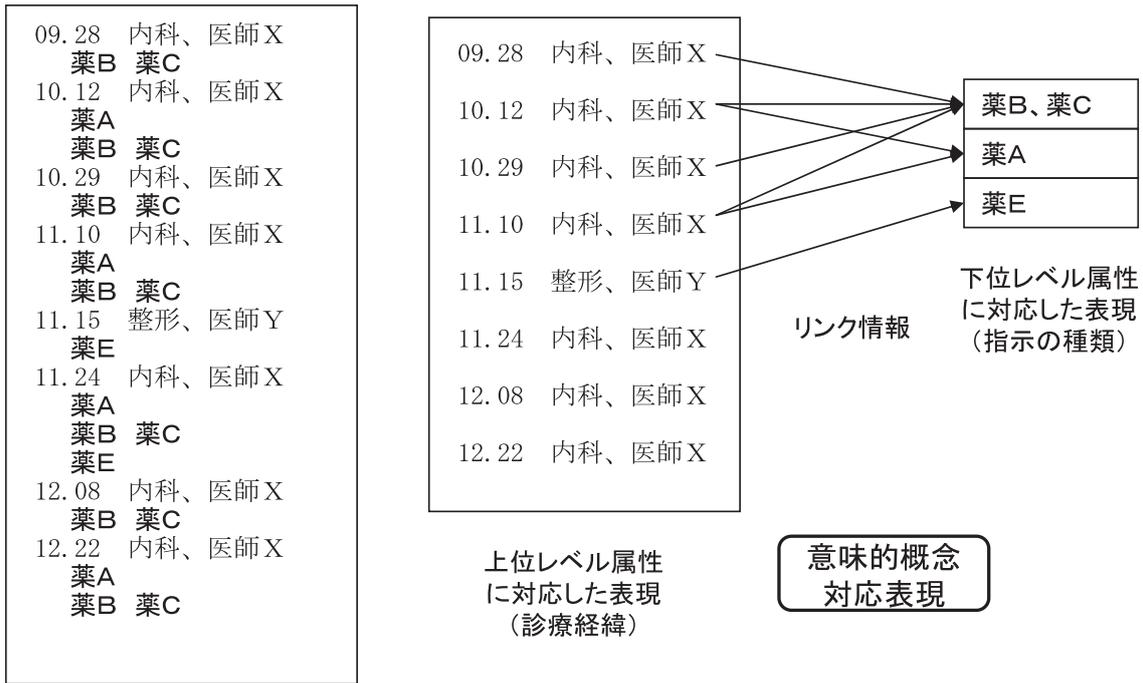


図5 時系列的表現と意味的概念対応表現

ンセプト・ハイアラーキの新しい利用法として、患者の処方指示の意味的上位レベル属性と下位レベル属性に分離して表現する方法を提案し、その有用性についても述べた。

今後、データウェアハウスを利用して、データ分析が多角的に行われ、医療資源の有効活用などにも貢献できるようになると考える。

なお、本研究には、上武大学特別研究費(平成21年度)を使用した。

参考文献

Jiawei Han, Micheline Kamber (2001), "Data Mining Primitives, Languages, and System Architectures," Data Mining : Concepts and Techniques, Morgan Kaufmann Publishers, 145- 178

Shuichi Toyoda, Noboru Niki, Hiromu Nishitani (2007), "SAKURA-Viewer : Intelligent Order History Viewer based on Two-Viewpoint Architecture," IEEE Transactions of Information Technology in Biomedicine, vol. 11, no. 2, 141- 152

大園博美 2002 : 最新データウェアハウスの基本と仕組み、秀和システム

豊田修一 2008 : 識者が語るITシステムの動向と選定のポイント、CLINIC BAMBOO, 日本医療企画, NO. 321, 66- 71