

ポスター発表

[P22-1B～P22-4B] ポスター

2018年6月22日(金) 13:00 ～ 14:00 第5会場 (3階・中会議室301B)

[P22-2B] Semantic Data Modelの導入による診療データ分析プロセスの改善

島川 龍載（広島赤十字・原爆病院）

Semantic Data Model の導入による 診療データ分析プロセスの改善

島川 龍哉^{*1*2}, 小園 菜美^{*1}, 鈴木 英夫^{*2}, 紀ノ定 保臣^{*2*3}

^{*1} 広島赤十字・原爆病院, ^{*2} 一般社団法人 SDM コンソーシアム,

^{*3} 岐阜大学医学部附属病院 医療情報部

On Improvement of Clinical Data Analytical Process by Using Semantic Data Model

Tatsunori Shimakawa^{*1}, Nami Kozono^{*1}, Hideo Suzuki^{*2}, Yasutomi Kinoshita^{*3}

^{*1} Hiroshima Red Cross Hospital & Atomic-bomb Survivors Hospital

^{*2} General Incorporated Association SDM Consortium

^{*3} Department of Medical Informatics, Gifu University Hospital

抄録: 【背景】医療機関における診療データの利活用促進を妨げる問題点を解決するため、医療用 SDM の有用性に関する調査が進められてきた。一方、医療用 SDM の導入が診療データの分析業務に好影響を与えている要因が何かを十分に検証できていないという課題もあった。【目的】医療用 SDM の導入が診療データの分析業務に好影響を与えている要因を明らかにする。【方法】データ分析の標準化手法となる CRISP-DM に実際の診療データ分析プロセスを当てはめ、医療用 SDM の導入前後における診療データの分析プロセスを比較する。顕著な差があったプロセスに対して、作業工数の平均値を比較する。【結果】「データの理解」、「データの準備」のプロセスに関する工数削減が、効率化に影響を与えていた。【考察】医療用 SDM の構造特性が、スキルレベルに関わらず発生する前処理の工数削減に寄与しており、データ分析プロセスの効率化にも、少なからず寄与していると言える。

キーワード: DWH、SDM、CRISP-DM、データ分析、分析プロセス

1. 背景

病院情報システムの普及に伴い、診療データの利活用が期待されているものの、統一されない DWH のデータ構造に起因するデータ利活用の促進を妨げる課題[1]が指摘されてきた。

これらの課題を解決するため、一般社団法人 SDM コンソーシアムでは、2014 年より、普遍的な定義のデータモデル化を目指し、医療情報に関する DWH (Data Warehouse) の設計書として、項目間の関係を含めた構造を有する医療用 SDM (Semantic Data Model) の開発[2]を行ってきた。

一方、医療用 SDM の実務への適用は端緒に就いたところであり、現状では、医療機関での導入効果は十分に検証されてこなかった。

これらの状況において、広島赤十字・原爆病院では、医療用 SDM の有用性に関する調査を開始し、医療用 SDM の実務での適合状況が概ね高い傾向であることや、導入後のデータ抽出や分析の依頼件数が増加傾向にあることから、医療用

SDM の導入が業務効率に好影響を与えていることを報告[3]しているものの、その要因分析が十分に検証できているとは言えない状況であった。

2. 目的

医療用 SDM の導入が業務効率に影響を与えている要因は何かを明らかにするため、以下の 2 つの調査を実施する。

1) データ分析プロセスの比較

SDM 導入前後でのデータ分析プロセスを比較し、どのプロセスに関して差があったかを定性的に検討し、その要因に関する分析を行う。

2) データ分析工数の比較

プロセス間の比較において差が生じた処理に関して、実務における工数の変化をステップ数として定量的に検討する。これにより、これまで SDM 導入のメリットとして提唱されていたデータの抽出、加工など、分析過程の変動要因への効率化に対する検証を行う。

3. 方法

1) データ分析プロセスの比較

データ分析の標準化手法となる CRISP-DM (CRoss-Industry Standard Process for Data Mining) に実際のデータ分析のプロジェクト工程を当てはめる。

CRISP-DM は、①ビジネスの理解、②データの理解、③データの準備、④データモデリング、⑤評価、⑥展開のプロセスを明確にすることで、最適な結果に結びつけることができる。試行錯誤を繰り返し、上手くいかない場合はどこからやり直す必要があるか、見極めるためのプロセスモデル[4]として、実務で使われていることから、プロセスを比較するためのツールとして適している。

この CRISP-DM に対して、医療用 SDM の導入前と導入後のプロセスを当てはめて比較し、医療用 SDM が持つ構造の特性を考察する。

2) データ分析工数の比較

医療用 SDM の導入前後で差があったプロセスに着目し、そのステップ数の変化を定量的に比較し、具体的な事例をもとに要因を分析する。

ステップ数は、IBMModeler を利用して、プロセスを可視化して、変化の差を比較する。

4. 結果

1) データ分析プロセスの比較

医療用 SDM の導入後は、普遍的な定義を持ったデータモデルとして集約されるため、「データの理解」における「収集」「記述」「調査」「検証」に関する新たな作業は不要となる。また、「データの準備」における「選択」「整理」「構築」「結合」「設定」に関しても、SDM の共通ヘッダーに含まれる普遍的な ER(Entity Relation)により、作業がパターン化され、部品の再利用が可能となるため、これまでのデータ加工における無駄な作業を削減することができる。なお、「データモデリング」や「評価」に関するプロセスにも差は生じているが、データ分析者のスキル向上や分析ツールの違い、さらには、習熟度に関する要因も含んでいると思われるため、医療用 SDM だけの効果ではないと言える。

2) データ分析工数の比較

分析プロセス全体において、前処理に該当する「データの理解」、「データの準備」の工数削減

は、クレンジング処理した内容が、テーブルに保存されていることが大きな要因であることがわかった。実際に、医療用 SDM 導入により、全体の40%程度の工数削減が結果として得られている。医療用 SDM の前処理の工程は、前処理時間の毎回の処理を抑えることや、マスタ等の更新における結果出力の差異が発生しないように、複数テーブルの情報が、前処理やアウトプットに適したテーブル構成となっている。また、電子カルテベンダーが提供する DWH は、テーブル構成、項目名、データの定義が、ベンダー毎に異なるが、医療用 SDM は、テーブル構造とテーブル間リレーション、項目名、定義が統一されている。

5. 考察

データ分析の前処理は、データ分析者のスキルレベルに関わらず発生する工程であるため、「データの理解」「データの準備」の工数削減が業務効率に大きく影響する。したがって、医療用 SDM の構造の特性が、データ分析プロセスの効率化に少なからず寄与していると言える。

6. 結語

医療用 SDM の実務での導入評価を行うことで、DWH の問題解決につながり、データの利活用の促進が期待できるため、今後の研究の継続が重要であると考ええる。

参考文献

- [1] 紀ノ定保臣:SDM コンソーシアムの役割とその社会的な背景, 中部 IT 融合セミナー 2014, [http://sdm-c.org/], 2014
- [2] 紀ノ定保臣:SDM コンソーシアムの目的と意義,第 36 回医療情報学連合大会 36th JCMJ(Nov.2016):158-159,2016
- [3] 島川 龍載,鈴木 英夫,津久間秀彦他:
DWH を指向する医療用 Semantic Data Model の BI に関する有用性の検討, 第 37 回医療情報学連合大会. 554-557,2017
- [4] Wirth, Rüdiger, and Jochen Hipp:
CRISP-DM: Towards a standard process model for data mining . Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 2000