

データベース品質劣化要因分析モデルの提案

6P-4

中村敏夫 伊土誠一 石垣昭一郎
NTT情報通信網研究所

1. はじめに

企業内の重要な情報の多くが、データベース(DB)化される状況になってきている。しかし、DBの品質²に関する重要性が十分認識されていないにもかかわらず、データ品質についての検討は少ない。[1]

本稿では、データ誤り(不良)の混入するプロセスを特定し、そのプロセスを改善すれば、データ誤りが発生しなくなるようなDB品質向上対策を検討するための、DB品質劣化の要因分析モデルを提案した。

2. DB信頼度曲線

DBの品質に関しては、ソフトウェア、或いは、ハードウェアの場合とは異なり、

① DBは通常、運用中に頻繁に更新され、DB量も増える。それに伴い、DBの不良も図1のカーブIのように増加し、無視できない、

また、

② DBの個々のデータは、データ毎に色々なプロセスを経て生成/更新される。その各プロセスでは人やプログラムなどが係わるため、各局面で

DBの不良が起きる可能性がある、

という特徴を有する。

DBの品質を向上させるには、DBが生成/更新される全てのプロセスを明らかにし、それぞれの誤りの起こったプロセスを特定し、防止策を検討することが必要である。そのため、DB品質劣化要因分析用のモデルを作ることが重要となる。

図1は、DBの信頼度成長曲線である。

DBを運用中に何等の品質改善を行わない場合、DBの品質はカーブIの曲線をたどる。運用中に何等かの品質改善活動や運用中に発見される不良を修正すると、カーブIIの曲線を推移する。実際のDBシステムでは、この両方のことが起こり、DBの信頼度成長曲線は、カーブIIIのようなになる。

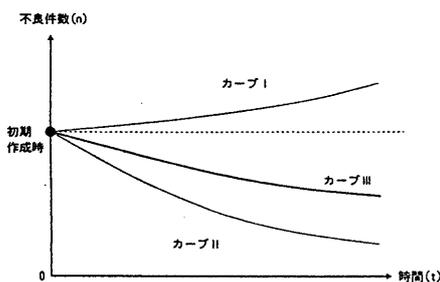


図1 DB信頼度成長曲線

* 本稿での「DB品質」とは、狭義の意味での「DB信頼性」を指す。

Model to analyze the factors deteriorating database quality

Toshio NAKAMURA, Seiichi IDO, Shoichiro ISHIGAKI
Network Information Systems Labs., NTT

3. DB品質評価基本参照モデル

DBは、現実世界の情報(ここでは、これを実体と呼ぶ)を写像したものである。

実体とは、例えば、設備に関するDBでは設備(交換機、線路、電話機、等)、顧客に関するDBではお客様注文(オーダ)であったりする。

DB品質評価を行う上での基本参照モデルを図2に示す。(評価対象の)DBの成り立ちとしては、

- ① 実体を直接写像したもの、
- ② 写像した別DB(CD/FDのデータも含む)から流通したもの、

の2種類に大別できる。なお、DBへの更新という観点からは、運用中における業務用プログラムからの更新(追加/変更)もある。

また、写像の手段としては、以下の2つがある。

- ① 自動入力
- ② 人手を介する入力

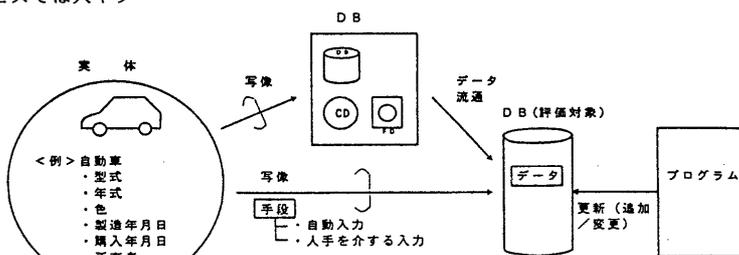


図2 DB品質評価基本参照モデル

4. DB品質劣化要因分析モデル

実体の変化に対して、DBは適正なタイムラグの許容範囲内で追従していく(デッドライン制御)必要がある。

この追従のメカニズムには、

- ① 実体の変化を検出する検出メカニズム
- ② 実体の変化をDBに反映する反映メカニズム
- ③ 何等かのデータ誤りに伴う修正メカニズム
- ④ 業務用プログラムによる運用中での更新メカニズム

これらのメカニズム中の全ての処理(プロセス)において、不良が混入する可能性がある。

DB品質劣化要因の分析モデルを、図3に示す。品質劣化要因としては、以下のものが想定できる。

- ①-1 ... ソフトウェアによる実体変化の検出不良
- ②-1 ... ハードウェアによる実体変化の検出不良
- ③-1 ... 人間による実体変化の検出不良(聞き取り誤り/記入誤り/確認誤りまたは該DBに関する担当者への連絡漏れ)
- ④ ... 別DBからの不良データの利用
- ①-2-a ... ソフトウェア(データ流通プログラム)の不良
- ①-2-b ... ソフトウェア(更新プログラム)の不良
- ③-2-x ... 実体変化に伴い、別のデータ項目を追加するが、その際の人間による不良
- ③-2-y ... 人間による投入時誤り(コーディング誤り含)

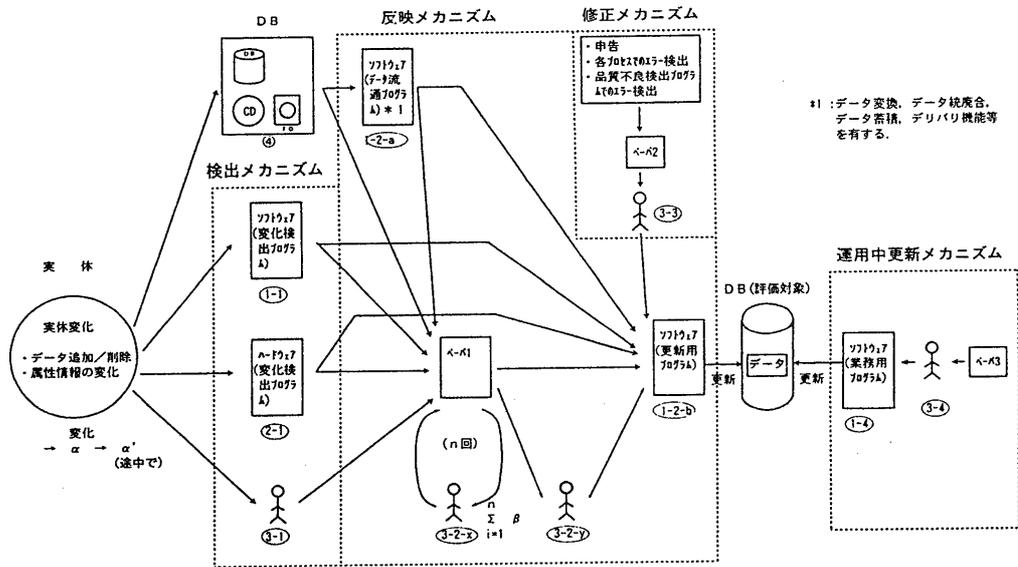


図3 DB品質劣化要因分析モデル

表1 DB品質劣化要因の分析結果

誤り種別	1-1 検出不良	2-1 検出不良	3-1 記入誤り、連絡漏れ	不正利用	1-2-a データ流通プログラム不良	1-2-b 更新プログラム不良	3-2-x 項目追加時の不良	3-2-y 投入時誤り/漏れ	3-3 修正漏れ/誤り	1-4 業務用プログラム不良	3-4 1-4の投入時誤り等
DB品質劣化要因	(77)	(n-1)	(人間)		(77)	(77)	(人間)	(人間)	(人間)	(77)	(人間)
1. DB-A a地域			100								
2. DB-A b地域			83				13	4			
3. DB-B a地域			15				59	25			
4. DB-B b地域							36	70			

③-3 … データ誤りを検出しているが、そのエラー修正の漏れ/誤り
 ①-4 … ソフトウェア(業務用プログラム)の不良
 ③-4 … 運用中の業務用プログラムに対する、人間による投入時誤り等
 ここで、③-1、③-2-x、③-2-y、③-3、③-4は、ヒューマン・エラー[2]である。
 本モデルに基づき、データ誤り(不良)の要因を分析した結果を表1に示す。なお、評価対象DBは、DB-AとDB-Bの2つである。
 特徴的なことを、以下に示す。
 (1) 不良は、全てヒューマン・エラーであり、その内訳は、
 ③-2 … 記入誤り/連絡漏れ
 ③-2-y … DBへの投入時誤り/漏れ
 ③-3 … エラー検出後の修正漏れ/誤り
 であった。
 (2) 人間による実体変化の検出不良の内訳では、DB-A(a地域)では35%、DB-A(b地域)では、65%が、他担当者への連絡漏れであり、その内容は、自担当のDBの更新は行ったが、関連して更新する必要のある他DBの担当者に連絡していなかった場合であった。
 (3) DB-Bにおけるエラーの修正漏れ/修正誤りについては、大部分がデータ流通プログラム(①-2-a)で、データ交換処理(コード体系が異なることに起因)等のエラーが出力されているが、その修正漏れであった。これは更新処理でのエラー出力に比べて、おろそかにされていた面があった。

5. まとめ

本稿では、データ誤り(不良)の混入プロセスを考慮し

た、DB品質劣化の要因分析モデルを提案した。本モデルにより、DB不良の混入プロセスを特定することができ、今後の不良防止策の検討に役立つものと期待できる。

今後の課題としては、以下のものがある。

- (1) 他DBシステムにおける検証及びヒューマン・エラーの“真の原因”の分析及び不良混入を防ぐための対策
- (2) 不良の検出方法

参考文献

[1] Huh Y U et al. : Data quality, Information and software technology, Vol. 32, No. 8, pp. 559-565 (1990).
 [2] 菅野文友 : ヒューマン・エラーのメカニズム, 日科技連(1980).