

関手的データモデルにおける情報単調性

Information Monotonicity in Functorial Data Model

水野 隆文^{1*}

Takafumi Mizuno¹

¹ 名城大学

¹ Meijo University

Abstract: I analyze information monotonicity of database by using functorial data model. I specify the monotonicity, and describe that a kind of inconsistency occurs among order relations over results of queries to the database which has the monotonicity.

1 はじめに

近年、ブロックチェーン技術を実用する環境が整ったことにより、ビジネスの分野において、データベースを作成する際の前提に構造的変化が起きつつある。最も大きな変化は、民間銀行が独自通貨を発行できるように法改正がなされたことである [1]。いくつかの銀行は、2017 年秋より暗号通貨の発行が予定されている。これら暗号通貨は、通貨の機能の大部分を代替し、通常のモノや商品とは異なる扱いがなされることになり、市場の状況の推移によっては、既存の通貨に準ずる扱いが必要になる。このような暗号通貨は、自治体や企業も発行でき、数年間は数百種類が混在すると予想される。

すでに普及している仮想通貨では、ブロックチェーン技術を用い、取引記録の帳簿の正本のデータをノード上で共有しつつ、改ざんを防いでいる。自社のデータや限られた企業とのみ共有すべきデータは、従来ならば特定のサーバを用いて管理するものであった。仮想通貨に利用されている技術を応用し、これらデータのうちのいくつかをノード上で管理・共有するように、データベースの設計や運用が変更されると考えられる。

自社のデータベースの再設計は不要であっても、他社との取引を見据え、インターフェイスを検証する必要がある。データベースの再構成においては、次の点を前提にすることは妥当と考えられる。

- データベースの具体的なデータモデルやデータ操作の低水準な手続きは、高頻度に変更される。
- 他社と共有する部分を議論する際に、他社とのコミュニケーションが要求される。
- データは改ざんされずに単調に増加する。

これら前提の下、本研究では、関手的データモデルを用い、データベースの設計や分析を行うことを目指す。

関手的データモデルは Spivak[2] により提案された。関手的データモデルにおいては、スキーマは、有向グラフ上のパスとして表現される。個別のデータベースに依存するようなビジネスルールはパスの合同関係として表現される。データはスキーマの圏から **Set** への関手、データ操作は関手間の自然変換として表現される。そして、データベースへのデータ操作は、データの状態に変化をもたらすものとし、**Set** から **Set** への関手として表現される。

本研究で対象とするデータベースでは、データは単調に追加されていく。データを削除した場合も「そのデータが存在した」という情報は決して削除されないものとする。本研究では、この性質を情報単調性とよぶ。

本発表では、この、有向グラフそのものをデータベースとして表現する抽象度の高いモデルを用いて、情報単調性を考察する。

2 関手的データモデル

まず、つぎのようにスキーマを定義する。 $G = (V, A)$ を有向グラフとする。 V は頂点の集合、 A は矢の集合である。矢 $a \in A$ は、頂点 $src(p)$ から出て頂点 $tgt(p)$ に入る。このグラフのパスの集合を $Path_G$ とする。有向グラフ G と、パスの集合 $Path_G$ におけるある合同関係 \simeq の組 $C := (G, \simeq)$ がスキーマである。この合同関係 \simeq はつぎの条件を満たす。

1. $p \simeq q, p, q \in Path_G$ ならば、 $src(p) = src(q)$.
2. $p \simeq q, p, q \in Path_G$ ならば、 $tgt(p) = tgt(q)$.
3. $p, p', q, q' \in Path_G, a = src(p) = src(p'), b = tgt(p) = tgt(p'), b = src(q) = src(q'), c =$

*連絡先：名城大学 都市情報学部 都市情報学科
〒461-8534 愛知県名古屋市中区矢田南 4-102-9
E-mail: tmizuno@meijo-u.ac.jp

$tgt(q) = tgt(q')$, そして $++$ を2つのパスを連結する演算子とする. このとき, $p \simeq p'$ かつ $q \simeq q'$ ならば, $(p ++ q) \simeq (p' ++ q')$.

つぎに, データベースのデータ (インスタンス) を定義する. データベースは随時更新されており, そのときそのときの状態を, スキーマ C から \mathbf{Set} への関手 $I: C \rightarrow \mathbf{Set}$ として表現する. データベースの状態の変化は関手間の自然変換として表現する.

スキーマ C 上のデータを $C - \mathbf{Set}$ と表記する. データの抽出 (データベースへの問合せ) は, スキーマを変化させて, その変化したスキーマに対応するデータを取り出すと考えられる. スキーマの変化を $F: C \rightarrow D$ とする. このとき, D に対応するデータの抽出を表現する関手 Π_F は, データを無視する関手 Δ_F との随伴として表現できる.

$$\Delta_F: D - \mathbf{Set} \quad \rightleftarrows \quad C - \mathbf{Set} : \Pi_F$$

Π_F は SQL の SELECT 文に対応する.

3 情報単調性

ここで, どのようなデータモデルなら, データベースが情報単調性をもつか考える. ある時点 t のデータを I_t , その後データベースの状態が変化した時点 t' でのデータを $I_{t'}$ とする. 情報単調性を前提とするなら, t' の時点で削除されているデータも, 「 t の時点でそのデータが存在していた」という情報が残っており, 元の同じデータを取り出せるはずなので, つぎの mono 射 i を構成できる.

$$t < t' \quad \Rightarrow \quad i: I_t(C) \rightarrow I_{t'}(C). \quad (1)$$

このような条件は, データの有効時間のような情報やタグをデータに付与することにより実現できる. Goguen [3] はこのようなシステムを, 層を用いて構成した. \mathcal{T} をある開集合の集合とする. データに有効時間を付与したデータは, \mathcal{T} から $C - \mathbf{Set}$ の反変関手 $\mathcal{O}: \mathcal{T}^{op} \rightarrow C - \mathbf{Set}$ のうち, 層の条件を満たすものとして定義される. これが情報単調性を担保するデータを表現できるデータモデルのひとつである.

ところで, データベースの利用では, データを取り出すのみではなく, そのデータをもとに何らかの判断を行う. データを取り出す問合せは関手であった. ここで, a, b, c を問合せとする. ある時点での問合せ結果を

$$a(C) = X_1 \sqcup Z_1, \quad b(C) = X_2 \sqcup Y_2, \quad c(C) = X_3$$

とする. $X_1 \sqcup Z_1$ は X_1 と Z_1 の coproduct である. データ更新が行われた後に同じ問合せを行い c の結果が つぎのように変化したとする.

$$a(C) = X_1 \sqcup Z_1, \quad b(C) = X_2 \sqcup Y_2, \quad c(C) = X_3 \sqcup Z_0$$

もし射として $X_1 \rightarrow X_2, Y_1 \rightarrow Y_2, Z_0 \rightarrow Z_1$ のみが存在する場合, $a \rightarrow b \rightarrow c$ であった関係が, データ更新後には $a \rightarrow b \rightarrow c \rightarrow a$ となり, 循環が生じる.

例えば, 関係モデルで, つぎのような結果が得られたとする. x, y, z を有効時間やタグとする. 結果を評価する人は, 同じタグ同士の数値を比較し, 各結果の大小関係を提示する.

a	b	c
x 1	x 2	y 30
z 0.1	y 20	

比較の結果, $a < b < c$ である. そしてデータベースの状態が変化し, c の結果にデータが つぎのように追加されたとする.

a	b	c
x 1	x 2	y 30
z 0.1	y 20	z 0.0

大小関係は $a < b < c < a$ となる.

4 むすび

本研究では, 関手的データモデルにより情報単調性の分析を試みた. まず, 情報単調性を明示し (式 (1)), それを満たすデータ構造としてデータの有効時間をデータに付与する単純なモデルを用いた. そして, データベースが情報単調性をもつとしても, 問合せの結果同士の関係に循環 (矛盾) が生じうることを示した. このような循環は, 問合せの結果を coproduct で表したために生じている. 問合せの結果を数値化し, 数値を比較し何らかの判断をする際には, 結果が coproduct で表現されるモデルでは, 慎重に数値化する必要がある. 逆に, 分析のために十分に細かい問合せを用意しても, 問合せ結果の比較に矛盾や循環が生じる場合には, coproduct の構造が隠れていると推測される.

参考文献

- [1] 日本経済新聞, 「貨幣」認定, 金融庁が法改正へ, 2016年2月25日朝刊, p.5 (2016)
- [2] Spivak, D.: *Category Theory for the Sciences*, MIT Press.
- [3] Goguen, J.: Sheaf semantics for concurrent interacting objects, *Mathematical Structures in Computer Science*, Vol.2(2), pp.159–191 (1992)