

概念モデリングにおける判断基準の有効性評価と支援ツール開発

大木 幹雄† 秋山 構平‡

An Evaluation of the Conceptual Modeling Criteria and its Support Tool Development

Mikio OHKI Kohei AKIYAMA

あらまし オブジェクト指向情報システムやデータベースのスキーマ設計において、概念モデルの設計が重要な役割を果たす。しかしながら、概念（またはクラス、実体型）の抽出方法については、発見的な方法の域を出ておらず、初心者が概念を抽出するときの具体的な判断基準は存在しないといつてよい。このような状況を背景に、本研究ではオブジェクト指向の基本的な特徴に着目して、基本データ（すなわち、他から導出されないデータ）にもとづいて概念を抽出する判断基準を考察し、3つの具体的な概念の抽出方法に関する判断基準を定めた。さらに、判断基準の有効性を評価するため、ERモデル設計演習授業の受講者を対象にして、一定時間内で実体関連（ER）モデル構築を行わせたときの正解率に関する比較実験を行い、その効果を確認した。また、判断基準を機械的に適用して概念を抽出する支援ツールを開発し、支援ツールから出力された概念モデルと人手（分析者）による概念モデルとの構造一致度の比較実験を行った。本論では、これらの2つの実験を通して確認した判断基準の有効性について紹介する。

キーワード オブジェクト指向分析設計、概念モデル、CASE ツール、モデリング、判断基準

1. はじめに

オブジェクト指向開発の特徴として、クラス構造を中心にスパイラル的に分析設計内容を具体化してゆける点あげられる。そのため、オブジェクト指向分析設計法では、クラスの抽出と設計クラス図の作成に最も重きを置いており、クラス抽出に関して、さまざまなアプローチが考案されている。例えば、オブジェクトが隠蔽する対象の捉え方によって、責任駆動型[1][2]、データ駆動[3]、プロセス駆動、イベント駆動等が提唱され、実用に供されている。しかし、いずれのアプローチにおいても、クラスの抽出は発見的に行うものとされている。発見には経験が必要とされることから、例えば責任駆動型アプローチでは、CRC

(Class Responsibility Collaboration) カードを用いて、分析者がオブジェクトの気持ちになり、外部に対して負うべき責任や他との協調動作を協議することで、発見を容易にする方法がとられている。

いずれの分析アプローチをとるにせよ、システムを構成するクラスは、3つの役割をもつクラス群、すなわち、データを格納する Model クラス、外部との通信・データ交換を担当する Interface クラス、操作イベントを監視し、Interface クラスと Model クラスとの仲介を行う Controller クラスに分けられることから、最終的には、これらのクラス決定をどのような優先度で行うかの順序の相違に帰着されることになる。

データベースを中心とした情報システムでは、従来から、格納データの構造に対応する Model クラスの抽

† 日本工業大学情報工学科
Nippon Institute of Technology
4-1 Gakuendai Miyashiro Saitama ,345-8501 Japan
‡ (株)グリーンウッド

出を最も優先すべきものとしている。ビジネスアプリケーションを対象にして、オブジェクト指向分析の有用性を解いた方法論の多くは、Model クラスの原型となる概念モデルの構築作業を最重要作業として位置付けている(例えば[4])。しかし、概念(本論ではUMLに準拠して、以後、Model クラスの原型、あるいはERモデルにおける実体型と同義として用いる)の抽出は、依然として発見的に行うものでしかない。ユースケース法でも、まずシステムの利用シナリオに焦点を当てて、それらを列挙し、シナリオ中の名詞句から概念の候補を発見するとしか扱っていない[4][5]。

筆者も情報システムの分析設計に関して、概念モデルの構築を最優先すべきと考えている。しかし、いざユースケース法を用いて設計経験の浅い初心者が概念モデルの抽出を行うとなると、正しいモデルを抽出できるか否か疑わしい。せいぜい初心者に対して、概念候補になりやすい概念カテゴリリストを提供する程度の支援しか行えず、抽出時における明確で具体的な判断基準が示されていないからである。クラスが負うべき責任に着目した責任駆動型アプローチにしても、発見に対する最終的な判断は、分析者の経験に依存するとして突き放している。現在のオブジェクト指向CASEツール[5]もこの点については無力で、依然として分析設計者の技能に頼らなければならない状況にある。

本論では、このような状況を踏まえて、初心者でも概念モデル構築を正しく効率的に行えるようにする“概念抽出に関する3つの判断基準”を提案し、その根拠と有効性を論じる。同時に、判断基準を正規化と比較したときの優位性についても論じる。

以下では、まず概念認識の基本的な考え方を述べた後、その考えに従って導出した概念モデル構築の判断基準を示す。次いで、判断基準の有効性を評価するために、情報系学部の学生を対象にして行った概念モデル構築の比較実験結果を述べる。さらに、判断基準の機械的な適用が熟練者と同等の概念モデルを導くかを評価するために開発した“概念モデル構築支援ツール”の動作をいくつかの具体的な事例を通して確認する。最後に、判断基準の機械的な適用の有効性と、今後の課題について述べる。

2. 概念認識の考え方

概念モデルの作成は、“概念”の構成要素である「概念名称」と「属性」に主に焦点を当てて行われる。このとき、概念を認識し抽出する基本的な姿勢として、大きく次の2つ考え方が存在する。

- (1) 抽出すべき概念は、はじめから存在するものとして、それらを見出すのが分析であるとする考え方である。システム化の対象範囲で用いられている用語や負うべき責任に着目して、ふさわしい概念(およびその名称)を見出す。概念間の関連や属性は、概念の抽出後に決定してゆく。しかし分析の初心者にとって、この考え方に沿った分析アプローチは誤解が生じやすい[11]。
- (2) 概念は、共通する属性をもった何らかの「個別の事例(インスタンス)」の集まりに対して便宜的に名称をつけた結果であり、概念は事例の総称として獲得するものであるとの考え方である。概念は属性の格納する入れ物に過ぎない。概念の名称は、共通する属性をもったインスタンスの集合体を表すものとして、後から便宜的に命名する。

いずれのアプローチも、振る舞いの側面や他との相互作用的な側面は、二義的に決定するものとして位置付けている。

(1)の方法は、専門用語を含むシステム化対象業務で用いられる概念候補を発見的に抽出するとの考え方であるのに対して、(2)の方法は、概念の属性となりうるデータ項目(以後、これを属性候補と呼ぶ)の集合をグループ化しながら概念候補を抽出する。

筆者らは、後者の立場に立ち、かつ『概念は常にその属性と同時に認識されるべきもの』とする考え方をとる。これは、同じ事物であっても、視点や状況が異なれば異なる概念として扱われることがあるからである。“必須科目”と“選択科目”といった一見異なる概念に見えても、突き詰めると最低取得単位に組み込むべき科目か否かの二値属性値に帰着できる場合などは、その好例である。多数のデータ項目を扱う情報システムでは、微妙な相違を言い表す専門用語が氾濫することが多い。これらを個別に扱うのではなく、属性値の相違によって説明できるものは、属性の追加によって概念を整理・体系化させる意味で、概念と属性は同時認識の考え方は、以後の議論の基盤をなしてい

る。

この立場をとると、逆に概念の認識が困難なとき、一定の判断基準(具体的には時間軸と属性構造に着目した判断基準)を持ち込むことにより、属性候補集合を分類し、そこから概念候補を抽出するような迂回路が作成できることになる。そこで以下では、概念候補が特定できないとき、データ分析から洗い出したデータ項目群を分類・整理して概念候補を絞り込む概念抽出の判断基準を考察してゆく。なお、以後、業務分析、書式分析等によって洗い出されたデータ項目の中で、他のデータ項目から導出されないデータ項目を“基本データ”と呼ぶことにする。

3. 概念を抽出する判断基準の導出

3.1 判断基準を導出する着眼点

基本データを分類し、それらを属性候補として含む概念候補を推測する判断基準を考察する上で、オブジェクト指向がもつ最も基本的な機構上の特徴であるインスタンスの生成、インスタンスの識別性、クラスの属性継承の3点に着目する。ただし、ここでいうクラスとは、概念を実装した Model クラスを意味するものとする。

(1) クラスからのインスタンス生成

クラスからインスタンスを生成する操作は、オブジェクト指向がもつ操作の中で基本的なものの一つである。この操作によってインスタンスが生成されると、インスタンスがもつすべての属性の初期値が“同時”に実現値として決定される。基本データ集合は判明しているが、存在させるべき概念が明らかになっていない状況では、この事実は逆に概念抽出の判断基準として用いることができる。すなわち、“基本データ集合の中で実現値が決定される時点が同じものがあれば、それらは概念を実装したクラスから生成されたインスタンスの属性である”としてよい。言い換えると、同じ初期値の決定時点(以後単に決定時点と呼ぶ)をもつ基本データは、同じ概念に属すべき属性候補となる。概念が明らかになっていない状況では、決定時点は、システム化対象の業務の中における一時点(例えば、注文といった外部からのイベント発生時点)といった大雑把なもので充分である。

このような判断基準に類似する基準をもつものと

して、データ中心の分析方法論である DATARUN がある[12][13]。DATARUN では、基本データを発生するきっかけとなるプロセス (PDG: Primary Data Generator) をまず認識する。次いで、PDG よって発生させられ記録させられる基本データを抽出し、それらをエンティティに分類・格納する手法を提唱している。しかしながら、PDG と抽出すべき概念の関係を陽に定義しておらず、またなぜ PDG に着目するかの根拠も明確にされていない。ここで述べた、インスタンス生成における初期値の同時決定性に着目した判断基準は、DATARUN で分析の出発点である PDG になぜ着目しなければならないかの根拠を与えている。

(2) インスタンスの識別性

生成された個々のインスタンスは、単一の実現値をもつ属性値(=単値属性)の組で一意的に識別される。もちろん、すべての属性値が同じであるインスタンスが存在することから、オブジェクト指向では、それらを識別する隠れたキー属性としてオブジェクト ID を便宜的に導入している。これによって、データベースの関係モデルでは重複して存在が許されない“すべての属性値が同じ”属性の組も、オブジェクト指向ではインスタンスとして存在できる。しかしながら、概念や概念がもつ属性候補が明らかでない状況では、オブジェクト ID は、単値属性の組によるインスタンスの一意識別性に代替するものではない。

この識別性によって、逆に単値の基本データと混在する形で、同時に複数の値をもつ基本データ(=多値データ)や選択的に値が決まる基本データ(=選択データ)をもつ概念は存在が許されないことになる[13]。これは、関係モデルにおける第1正規化と従属性をもとにした関係の分解の考え方に対応している。すなわち、関係の集まりを、一意識別可能な要素から構成される関係の集合に分解してゆく考え方に対応している。

以上から逆に“多値データあるいは選択データ部分は、(他の概念に)分離する”との判断基準を導かれることになる。

(3) クラスの属性継承

継承関係にあるクラスが存在したとき、下位のクラスが生成したインスタンスは、上位クラスで定義された属性を継承してすべて保有する。(1)で示した初

期値の決定時点にもとづき属性候補を分類する判断基準と整合させるには、上位クラスの属性集合も下位クラスの属性集合も同一時点で実現値が決定されなければならない。事実、下位クラスのインスタンスは、そのように生成されるので問題ない。しかし、決定時点だけにもとづく判断基準では、上位クラスと下位クラスを分離できない。上位クラスの属性集合は、下位クラスのもつ属性集合の共通部分集合を形成しており、上位クラスが複数の下位クラスをもつとき、上位クラスの属性は、下位にある各クラスのインスタンス生成に対応した複数の決定時点をもつことになる。

そこで、クラス継承関係を形成する判断基準を矛盾なく導くために、属性毎に固有の決定時点集合を対応させる。通常は、属性毎の決定時点集合の要素数は1であるが、上位クラスは下位クラスの数に応じて、1以上の要素数をもつ。ここで、決定時点集合から、それぞれ特定の決定時点を選択するパラメータを“状況”と呼ぶことにし、特定の決定時点を選択したパラメータ値を“状況値”、決定時点集合の要素数を“状況数”と呼ぶことにする。下位クラスをもたない限り、状況がもつ唯一の状況値によって決定時点が選択される。

インスタンス生成の基本動作が、“インスタンスの初期値を、一つの決定時点ですべて決定する”ことから、状況数が1以上である属性の集合は、同じクラスに混在させることはできない。すなわち、属性の中で状況数が異なるものは他のクラスに分離されなければならないことになる。逆に考えれば、“複数の状況数をもつ基本データは、継承関係から見て上位の概念の属性候補である”との判断基準を導くことができる。このとき、上位概念の属性候補名は、一致していなければならない。

概念が明らかでない状況では、状況値は、基本データの初期値決定がいくつかの状況に分けて行われるときの状況名といった大雑把なもので十分である。

3.2 導出された判断基準の種類

以上から、基本データをもとに概念候補を抽出する判断基準をまとめると次のようになる。

【判断基準1】初期値の決定時点の同時性

初期値が決定される時点が同一の基本データは、同じ概念に属する属性候補とする。

【判断基準2】実現値の構造的性

多値や選択的に実現値が決まる基本データは、単値基本データと混在する形で概念の属性候補となりえない。別の概念の属性候補として分離しなければならない。

【判断基準3】初期値の決定状況の単一性

初期値の決定時点が複数の状況に依存する基本データは、状況によって一意的に決まる初期値の決定時点をもった概念の上位概念に属する属性候補である。ただし、上位概念に置くととき属性候補名は、一致しなければならない。

判断基準2は、構造化分析設計においてすでに用いられているが、判断基準1と判断基準3は、筆者らが新たに追加したものである。これらの判断基準は、属性候補を実現値の決定時点、値構造、利用状況の3つの視点から分類するものになっている[14]。

3.3 適用事例による判断基準の有効性確認

前述の判断基準が実際に有効であるかを、いくつかの事例を用いて確かめてみる。なお、すでに述べたように、基本データ集合とそれらの実現値の決定時点、および決定される状況は、あらかじめデータ分析によって明らかにされているものとする。

(1) 基本データの実現値の決定時点が異なるとき

例えば、注文伝票をもとに次のような基本データ集合が洗い出されたと仮定して、先に定めた判断基準にしたがって分類してみる。

・基本データ集合₁: { 顧客番号, 日付, 単品注文番号, 商品番号, 顧客名, 住所, 電話番号, 商品名, 商品単価 }

これらの基本データは、注文を受付けたときにすべて伝票に記入される。しかし、実現値の決定時点は異なる。例えば、注文番号や日付は、伝票に記入されたときが実現値の決定時点であるが、顧客番号や顧客名、顧客の電話番号、住所の基本データは、取引を開始するにあたって、顧客情報を登録するときに実現値が決定されている。注文受付時に注文伝票に記入される値は、それらの値の参照にすぎない。

同様に、商品番号や商品名、商品単価の基本データは、取扱商品として商品登録するときに実現値は決定され、注文受付時にはそれらを参照しているに過ぎない。したがって、判断基準1にしたがって、基本デー

タ集合 ϕ_1 を分類すると、図 1 のようになる。図 1 は基本データの実現値の決定時点によって、3 つのグループ、すなわち 3 つの概念候補に分類できることを示している。実現値の参照は、一般に一对多の関係にあることから、図 1 は図 2 で示すような概念モデルに対応することになる。ここで、図 2 中の関連の正確な多重度は、概念候補の関連分析から別途決定している。

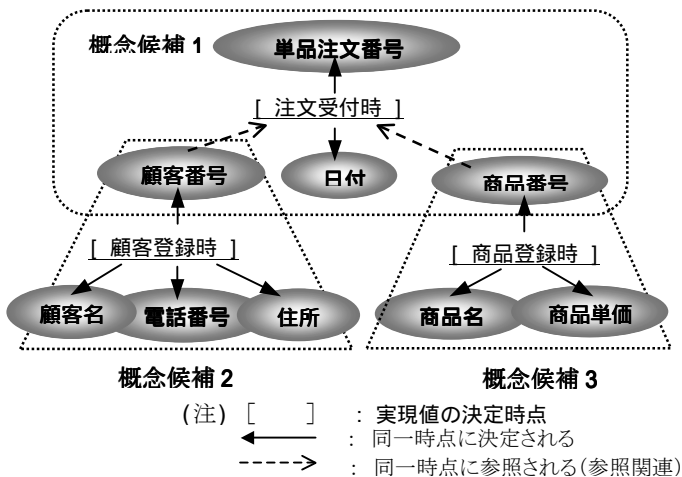


図 1 判断基準 1 を用いた基本データ集合 ϕ_1 の分類
Fig.1 Classification of Primary Data Set ϕ_1 using Criteria-1

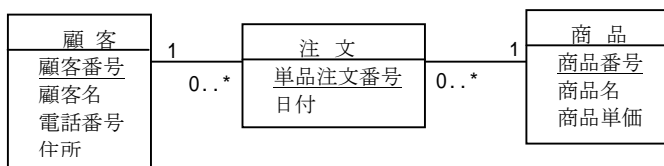


図 2 基本データ集合 ϕ_1 の分類に対応する概念モデル
Fig.2 Corresponding Conceptual Model to Fig.1

(2) 決定時点は同一で、実現値構造が異なるとき

同様に次のような基本データ集合を判断基準にしたがって分類してみる。

- 基本データ集合 ϕ_2 : { 日付, 注文番号, 商品名, 商品単価, 明細番号*, 明細注文数*, 商品番号*, }

ここで、*印の付いた基本データは、実現値の決定が同一時点で複数個行われるような多値構造をもつことを示している。基本データ集合 ϕ_1 と同様にして、判断基準 1 を用いて基本データ集合 ϕ_2 を分類すると図 3 で示すように分類される。

概念候補 1 の属性候補集合を見てみると、実現値の決定時点が同一で多値構造をもつ明細番号、明細注文数を含んでいる。そこで、判断基準 2 にしたがって、それらを

別の概念候補に格納すべき属性候補として強制的に分離する。強制分離後の分類を概念モデルに対応つけて示すと、図 4 のようになる。実現値の決定時点が同一ではあるが、実現値構造が異なるために強制的に分離され生成された概念候補同士の関連は、識別子依存関係に対応することに注意する必要がある。

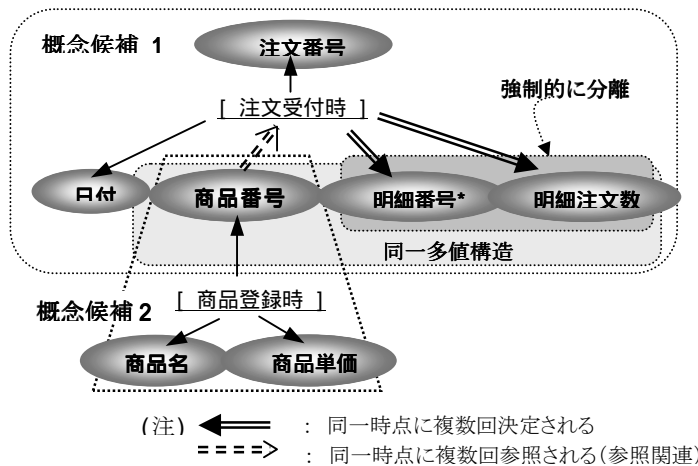


図 3 判断基準 1 を用いた基本データ集合 ϕ_2 の分類
Fig.3 Classification of Primary Data Set ϕ_2 using Criteria-1

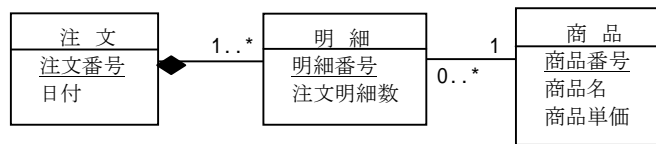


図 4 基本データ集合 ϕ_2 の分類に対応する概念モデル
Fig.4 Corresponding Conceptual Model to Fig.3

(3) 決定時点が異なり、かつ実現値構造が異なるとき

販売店において、特定の日に販売した商品と顧客の実績データを集計したときのような、それぞれが多値構造をもつような基本データ集合 ϕ_3 を判断基準にしたがい分類してみよう。

- 基本データ集合 ϕ_3 : { 日付, 顧客番号*, 顧客名*, 商品番号*, 価格* }

基本データ集合 ϕ_3 を ϕ_1 , ϕ_2 と同様にして分類してみる。実現値の決定時点は“販売商品実績の集計時”, “商品値決め時”, “顧客登録時”の 3 通りあるので、判断基準 1 にしたがえば、図 5 のような 3 つの概念候補に分類されるかのように見える。しかし、実現値の決定時点“販売商品実績の集計時”は、構造の異なる 2 つの多値属性を参照しているに過ぎない。したがって、図 6 で示す概念モデルと対応が付く。

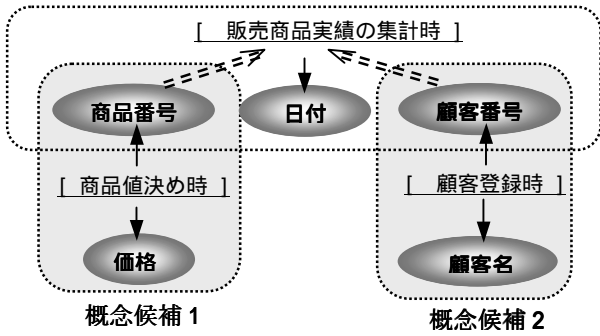


図5 判断基準1,2を用いた基本データ集合 ϕ_3 の分類
Fig.5 Classification of Primary Data Set ϕ_3 using Criteria-1, 2



図6 基本データ集合 ϕ_3 の分類に対応するクラス構造
Fig.6 Corresponding Conceptual Model to Fig.5

(4) 実現値の決定時点が状況に依存するとき

以上に示してきた事例では、実現値の決定時点が状況に依存しないことを前提に示してきた。しかしながら、状況に依存して、用いられる基本データ集合が異なる場合が考えられる。例えば、次のような事例が考えられる。

< 状況=顧客受注による発注 > 基本データ集合 ϕ_4 :
{ 発注番号, 日付, 発注業者番号, 発注数量, 納入予定日, 納入顧客番号, 運送業者番号 }

< 状況=在庫不足による発注 > 基本データ集合 ϕ_5 :
{ 発注番号, 日付, 発注業者番号, 発注数量, 在庫予定日, 倉庫番号 }

この基本データ集合 ϕ_4, ϕ_5 は、商品を発注するとき、“顧客受注による発注”と“在庫不足による発注”の2つの状況があることを示している。状況に対応して、それぞれ実現値の決定時点をもつ。このような場合には、判断基準1,2の適用に先立って判断基準3を適用して、複数の状況をもつ属性候補を継承関係から見て上位の概念候補群に属すべき属性候補として分離する。例えば、複数の状況をもつ基本データ部分集合 { 発注番号, 日付, 発注業者番号, 発注数量 } と、それ以外の基本データ部分集合 { 納入予定日, 納入顧客番号, 運送業者番号 }, { 在庫予定日, 倉庫番号 } に分離してから、それぞれの部分集合に判断基準1,2を適用する。このとき、複数の状況を

もつ部分集合とそれ以外の部分集合間には、継承関係で結ばれる。事例では状況は2つであったが、状況がn個に分けられるときは、状況数nに対して $(2^n - 1)$ 個の部分集合に分離される。例えば、3つの状況に対応する基本データの部分集合をA,B,Cとしたとき、図7で示すような継承関係をもつ部分集合に分離される。

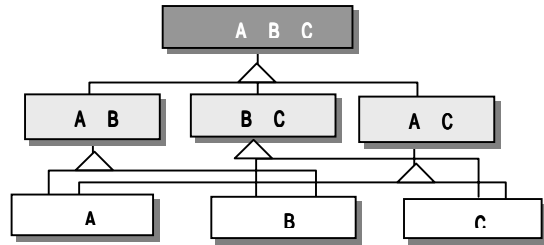


図7 状況数が3のときの基本データサブ集合間継承構造
Fig.7 Inheritance Structure of Primary Data Subset in the case of situation size 3

さて、基本データ集合 ϕ_4, ϕ_5 に判断基準3,1を適用して分類された構造は図8で示すようになる。これを概念モデルに対応つけると、図9で示すようになる。

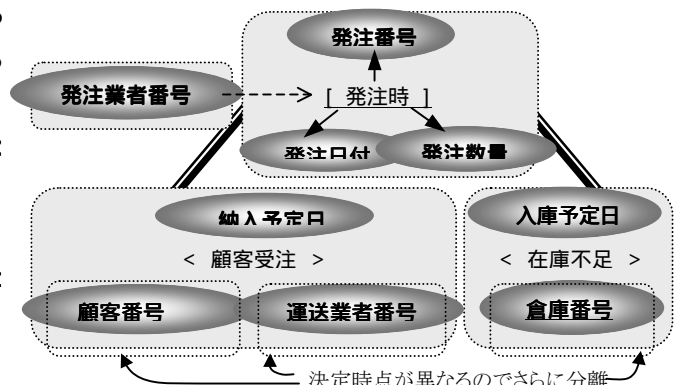


図8 判断基準3,1を用いた基本データ集合 ϕ_4, ϕ_5 の分類
Fig.8 Classification of Primary Data Set ϕ_4, ϕ_5 using Criteria-3, 1

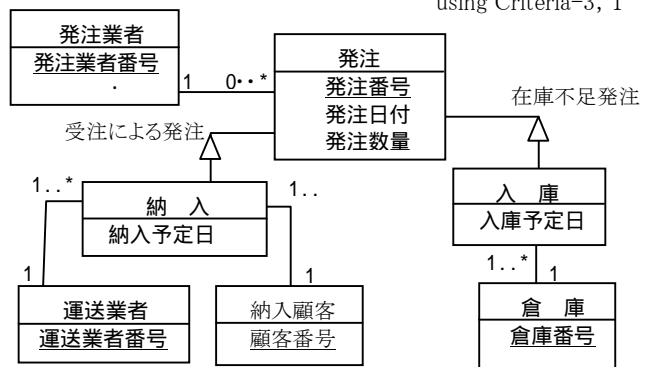


図9 基本データ集合 ϕ_4, ϕ_5 の分類に対応する概念モデル
Fig.9 Corresponding Conceptual Model to Fig.8

3.4 判断基準のまとめ

前項で示した事例を参考にして、判断基準 1,2 の適用順序と、分類後の構造、および対応する概念モデルとの関係をまとめると、図 10 のように図示することができる[16]。ただし、判断基準 3 は、前項(4)の事例で述べたように、状況数によって基本データ集合を分離するのに用いられる。

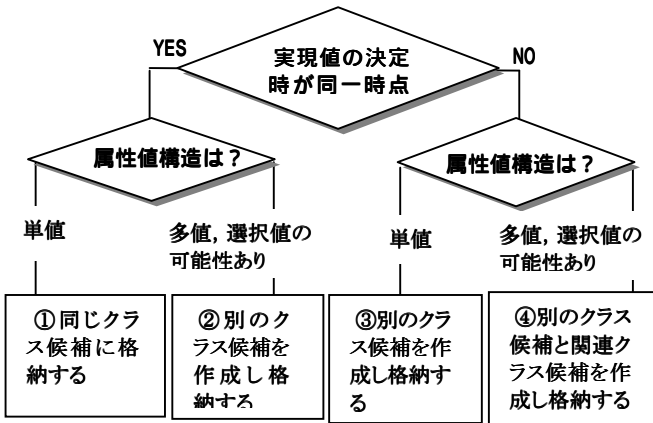


図 10 分類判断基準のまとめと関連の多重度との関係
Fig.10 The Summary of criteria and the relation of cardinality

3.5 正規化による ER モデル抽出との比較

図 1 から図 5 までに示した事例で作成した概念モデルは、いったん属性が抽出されれば、属性間の関数従属性、推移関数従属性、および自明な多値従属性の概念にもとづいた正規化の考え方でも、同様に導くことができる。しかしながら、正規化の考え方は、すべてのインスタンスに対する操作において不都合が生じないように、キー属性と他の属性間の意味的な従属関係もとに、属性間の関係を整理するためのものであり、分析の初心者がこのような意味的な従属関係を、あらかじめ、すべてのインスタンスを見通して発見することは困難といわざるをえない。

一方、正規化において基本的な役割を果たす関数従属性の定義 = “ある属性(通常、キー属性)の値が決まると対応する他の属性の値が一意的に決まる性質” を、意味的な因果関係として捉えるのではなく、値の決定時点の言及であると捉えてみる。すると、関数従属性とは、「ある属性の値が決まった時点(値の決定時点)で、他の属性の値が同時に、かつ一意的(すなわち、同じ実現値構造をもって)決まる」性質であると再解釈できる。これは、2.2 で示した判断基準 1,2

の根拠に対応することになる。同様に推移関数従属性は、値の決定時点の推移的な性質、自明な多値従属性は、値の決定時点が異なる属性を含まない性質と読みかえることができ、属性を分類する基準としての観点で、3.2 で示した判断基準 1,2 と対応付けが可能になる。

分析の初心者にとって、正規化の出発点となる基本データの意味的なつながりを発見することは、業務を熟知していないと判断できない面があるが、初期値の決定時点、および実現値の構造に着目した判断基準であれば、基本データの発生時点、発生仕方さえ判明すれば、正しい属性を含む概念を機械的に抽出することができる。これらの判断基準を適用して、機械的に概念モデルを抽出する支援ツールについて、5. で詳しく議論する。

4. 判断基準の有効性評価

4.1 評価実験の概要

前述の判断基準が、分析設計の初心者にとって正しい結果を導き出すガイドラインとして有効に機能しているかを評価するために、概念モデル構築のスキルを要求されるデータベース・スキーマ設計の演習授業において、比較実験を行った。具体的には、「判断基準を用いたグループ」と「用いないグループ」の正解率を比較することで有効性を評価した。

4.2 ER 分析演習での比較実験

筆者は、工業系大学においてデータベースの理論・応用講座を担当しており、演習授業において、具体的なアプリケーション業務の分析から実体関連図(ER図)を用いた概念モデル構築、データベース・スキーマ設計、MS ACCESS による実装までを行わせている。しかしながら演習授業の性質上、同一期間に判断基準の有効性を評価するために、「概念抽出の判断基準を示したグループ」と「示さないグループ」に分割することは困難である。そこで、年度を分けて行った2回の演習授業をもとに比較実験を行った。具体的には、判断基準を用いない年度の受講生が作成した ER モデルの正解率と、判断基準を示した年度の受講生の正解率について比較実験を行った。再履修者は、比較の母集団から除外し、両年度の受講者がもつ実体関連分析の意義、表記法の内容に関する熟知度のレベルは同水

準と仮定した。また、往々にして受講者が前年度の解答を非公式に入手するような学習効果が予想されるため、両年度で重複しないアプリケーションをサンプルとして採用した。ただし、用いるデータ項目名が異なるだけでスキーマ構造が類似するようなアプリケーション（例えば、在庫管理や販売管理等）は、分析段階では類似性の点で影響しないと判断し除外はしていない。

実験の概要は以下のとおり。

- (1) 母集団 1999年度 9 サンプル
2000年度 11 サンプル
- (2) 実験の実施方法
 - ・ 被験者は、両年度とも、3名以下のチームに分割し、チーム毎にアプリケーションの決定を行わせた。データ項目の収集・整理方法、ERモデルの記述方法に関する解説は、同時間を費やした。
 - ・ ERモデルの作成は、両年度ともCASEツールを用い3コマの演習時間を割り当てて行った。
 - ・ 判断基準の適用を試みた年度は、演習に先立って、60分の判断基準の根拠を解説する時間を設けた。

(3) サンプル・アプリケーション

野球選手成績管理、アルバイトシフト管理、レンタカーDB、図書館DB、カラオケDB、パソコン部品DB、通信プロバイダ検索、アパート検索、就職企業検索、中古車在庫管理、電気店営業管理、邦楽アーティスト検索、コミック雑誌検索、競馬レース成績・血統管理、観光地DB、カクテル・レシピDB、新作ゲームDBの20アプリケーションをサンプルとした。サンプル全体の特性は表1の通りである。

表1 サンプルアプリケーションの全体特性
table.1 Property of Sample Applications

	1999年度	2000年度
サンプル数	9	11
平均Entity数	4.2	5.7
平均Relation数	9.1	8.3
平均Attribute数	14.9	18.1

(4) 正解の判定方法

ERモデルの正しさの判定は、1999年度は筆者が完成度にしたがって採点した。2000年度はチーム毎のアプリケーション開発発表時において、受講者全員が相互に採点した完成度の得点平均を用いた。なお、祭典

における配点比率は、実体と属性等の正しさに対して70%、関連の正しさに対して30%とした。なお、審査の着眼点および配点は、以下のとおりである。

- 実体の種類と属性の正しさ(配点50点)
- 関連付けすべき実体の正しさ(配点20点)
- 関連多重度の設定の正しさ(配点30点)

このほかに、主キー属性の設定の正しさも含まれているが、判断基準の有効性評価とは直接関係が薄いため、比較実験の得点からは削除してある。

- (5) ERモデル作成時における誤り分析
各年度の誤りの種類と誤りサンプルに占めるそれらの発生比率は以下のようであった。

- 異なる種類の属性の混在
属性は実体固有の性質を列挙するとの認識はできているものの、認識し管理する実体の範囲があいまいなため、実体として関連つけるところを属性とする誤りである。
- 属性と実現値の混同
実体中に属性と実現値を混同して列挙する誤りで、実体と属性の意味が理解できていない誤りである。
- 実体属性と関連属性の混同
関連がもつ属性とすべきところを実体の属性とした誤りである。
- 関連付けすべき実体の誤り
関連の多重度設定の誤り

表2 誤りサンプルに占める誤りの種類の比率
table.2 Error factor and Ratio

誤りの種類	誤りサンプルに占める比率	
	1999年度	2000年度
	88%	40%
	25%	10%
	13%	10%
	13%	20%
	67%	60%

表2からも分かるように、1999年度に比較して、判断基準を陽に提示し、学生に対して機械的に基本データを分類し、実体を抽出するようにした2000年度の演習では、実体に異なる種類の属性を混在させる誤りは半減している。ただし、誤りが0%にいたらなかった理由として、初期値の決定時点に対する理解不足や決定時点を必要以上に詳細に分けすぎた点が上げられる。関連の多重度の設定に関しては、大きな改

善は見られなかった。これは関連の多重度の正確な決定が、ビジネス上のルールに依存するため、ルールの分析が不十分であったことに起因するものと考えられる。

(6) 判断基準の有効性の評価

判断基準の有効性を ER 図の完成度から比較するために、各サンプルの正解率平均値の有意性検定を行い以下の結果を得た。

有意性検定 (t 検定)			
1999 年度	平均	43.9	分散 743.4
2000 年度	平均	62.7	分散 591.8

危険率 5%で判断基準を用いた年度に関しては、ER モデルの正解率平均値の向上が確認された。よって、判断基準の有効性が確認された。

(7) モデル抽出における正規化の適用状況

今回の比較実験は、両年度とも ER モデルの表記方法、属性は実体の固有の性質であるといった属性の意味、正規化の考え方を学んだ学生を対象にしている。また、ER モデルを作成させるに当たって、管理すべきデータの列挙と基本データの整理を行っている。しかしながら、表 2 での 1999 年度の誤り①の発生比率からもわかるように、利用上の不都合をまでを分析し、正規化の考え方を適用できる者はほとんどいなかった。正確に述べれば、経験の浅い者は、そこまで分析できなかった。

3 種類の判断基準を用いる手法は、初心者にとっても ER モデルを機械的に作成できる点で、正規化による方法より優れた方法といえる。ただし、初期値の発生時点をどのように認識するかが新たな問題として生じた。

5 . 概念モデル抽出ツールの動作確認

5.1 実験の目的

基本データと判断基準に必要な初期値の決定時点、実現値構造が分析できていれば、判断基準は機械的に適用可能である。そこで、判断基準をルールとして取り込んだ概念モデル構築支援ツールを試作し[17]、熟練分析設計者と同品質の概念モデルを抽出するかの実験をいくつかの事例データをもとに行った。

正しい概念モデルが導出できるのであれば、分析の初心者のみならず、熟練者にとっても、基本データの収集や利用状況のヒアリング、ユーザ要求の分析等の分析者が本来行うべき作業に専念できる点で、利用価

値は高い。

5.2 ツールの主要機能

開発ツールの機能は以下の通りである。

- (1) 基本データおよび付加情報の入力機能
- (2) 判断基準 1,2 による基本データの分類機能
- (3) 判断基準 3 による継承関係の導出機能
- (4) 概念モデルの表示機能

5.3 事例による概念モデル抽出の比較実験

ここでは、4 つの事例に関して、支援ツールへのデータ入力形式と作成した概念モデル図、および分析設計者が作成した概念モデル図との比較を示してゆく。ただし、支援ツールは、属性を整理した概念候補を示すのみであり、分析設計者が作成した概念名とは必ずしも一致していない。また、概念名が明らかに含む属性名は、分析設計者が作成した概念モデルでは、省略しているときがある。

【 事例データ 1 】基本データ集合 4

状況が複数存在するようなとき、判断基準 3 にしたがって、概念の継承関係が出力されるかを見てみる。図 11 で示すような、前項 (4) の適用事例で用いた基本データ 4 を入力する。図 11 において、< > で囲まれたものは状況を示しており、[] で囲まれたものは、実現値の決定される時点を示している。図 11 は、状況が 2 つあり { } でまとめられた基本データ群の実現値が 決定時点 “ 発注時 ” に決定されることを示している。なお、{ } 内の基本データ名の後にさらに決定時点が付加されたものは、その決定時点が優先することを示している。

<顧客発注> // 状況名 { 発注番号 発注日付 発注業者番号[発注業者登録時] 発注数量 納入予定日 顧客番号 [顧客登録時] 運送業者番号[運送業者登録時] } [発注時] // 決定時点 1	<在庫不足> // 状況名 { 発注番号 発注日付 発注業者番号[発注業者登録時] 発注数量 在庫予定日 倉庫番号[倉庫登録時] } [発注時] // 決定時点 1
---	---

図 11 事例データ 1: 基本データ集合 4
Fig.11 Sample Data1: Primary Data Set 4

図 12 は、支援ツールが作成した概念モデルの原型である。図 9 で示した概念モデルと構造的に合致している。ただし、図 12 で示した属性候補のグループに対する名称付けは、属性候補全体を見渡しながら行わ

なければならない。また関連の多重度の決定も、基本的な種類は図 10 で示した決定手順から決定できるものの、詳細な妥当性については別途検討しなければならない。詳細な検討の必要性は、以下の事例すべてに対しても同様にいえる。

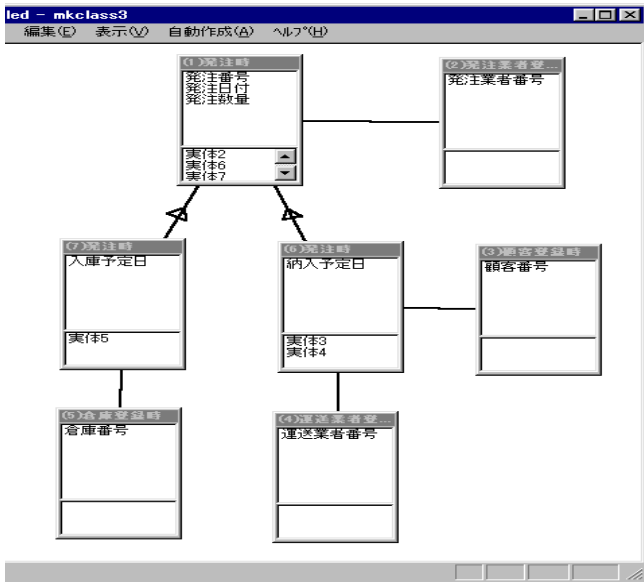


図 12 事例データ 1 から抽出した概念モデル図原型
Fig.12 Output Conceptual Model from Sample Data1

【事例データ 2】類似する基本データを含むときに多重継承を生じる可能性がある図 13 で示すような基本データ集合を入力する。その出力結果は図 14 で示す通りである。

<正社員採用> { 氏名 社員コード 月給 勤続年数 } [社員登録時]	<派遣社員> { 氏名 社員コード 契約期限 派遣元会社名 月給 } [社員登録時]	<バイト採用> { 氏名 社員コード 契約期限 時給 } [社員登録時]
---	--	---

図 13 事例データ 2：類似する基本データを含む
Fig.13 Sample Data2: Containing Similar Primary Data Set

一方、分析者によって、図 13 で示すような基本データをもとに概念モデルを作成したものが図 15 である。この事例においても、支援ツールからの出力と基本的な構造は一致する。

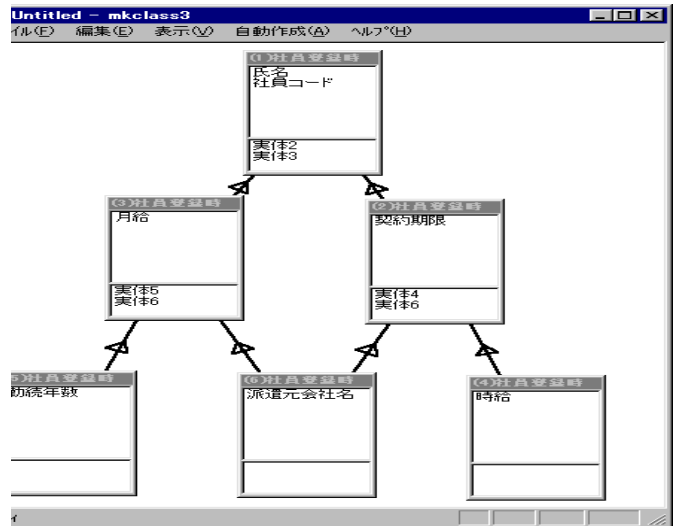


図 14 事例データ 2 から抽出した概念モデル図原型
Fig.14 Output Conceptual Model from Sample Data2

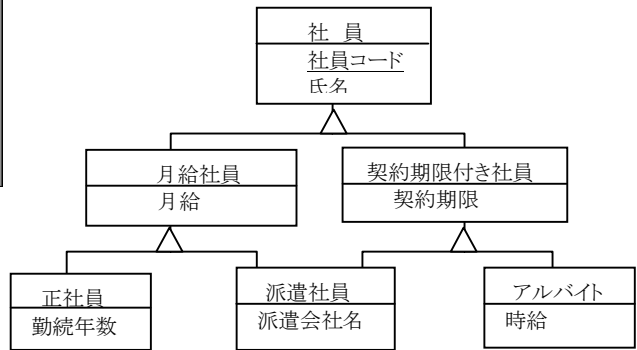


図 15 事例データ 2 もとに分析者が抽出した概念モデル
Fig.15 Output Conceptual Model from Sample Data2 by Analyst

【事例データ 3】決定時点の異なる類似基本データ
事例データ 2 と同様に、図 16 で示すような、3つの状況下で発注するときの基本データ集合を考える。この例では、事例(1)に「同時発注」の状況が追加されていると共に、実現値の決定時点が異なる基本データをそれぞれの状況が含んでいる。ただし、「同時発注」の状況が現実にかかることは稀であり、この事例データはあくまで検証が目的のものである。

支援ツールからの出力結果は、図 17 で示す通りである。一方、分析者によって抽出された概念モデルは、図 18 で示す通りである。概念の配置は異なるが、基本的な構造は一致している。

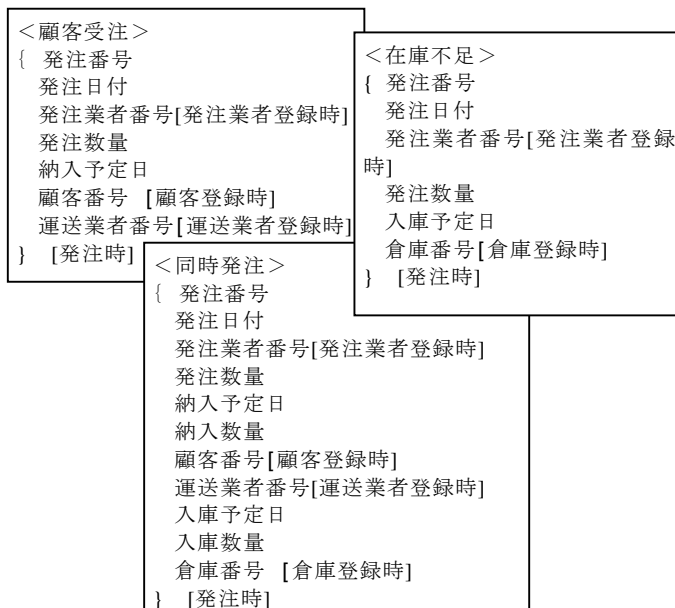


図 16 事例データ 3：決定時点の異なる類似基本データを含む
Fig.13 Sample Data2: Containing Similar Primary Data Set

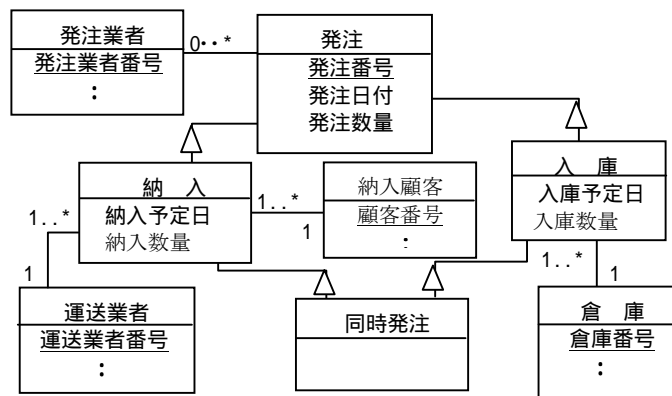


図 18 事例データ 3 もとに分析者が抽出した概念モデル
Fig.18 Output Conceptual Model from Sample Data3
by Analyst

【 事例データ 4 】 複数の状況をもつ決定時点の異なる基本データ

最後の事例として、レンタルビデオショップの事例を取り上げる。入力する基本データは、レンタルビデオ業務で用いられている伝票、帳票類を分析して、洗い出した図 19 で示すような基本データとする。支援ツールからの出力結果は、図 20 で示す通りであり、分析者によって抽出された概念モデルは、図 21 で示す通りである。

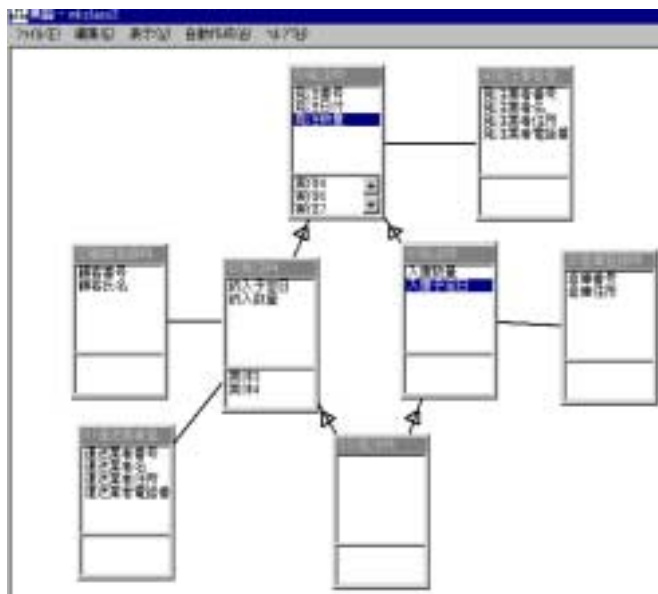


図 17 事例データ 3 から抽出した概念モデル図原型
Fig.17 Output Conceptual Model from Sample Data3

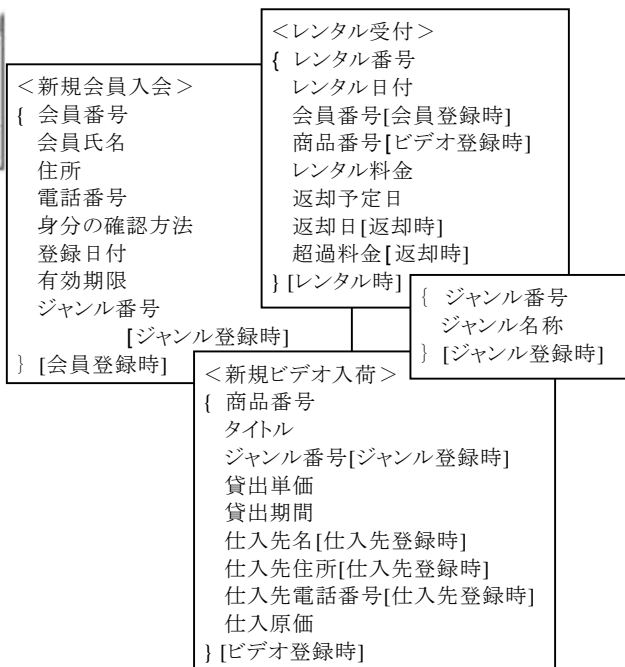


図 19 事例データ 4：複数の状況と決定時点をもつ基本データ
Fig.19 Sample Data4: Containing several Situations and Determination Time

この事例では、構造的なものが多少異なる。相違する点として、支援ツールでは、“貸出期限”と“貸出明細”が、1つにまとめられているのに対して、分析者が作成した概念モデルでは、2つの概念に分けている。これは、分析者が貸出明細と貸出期限を必要に応じて分けて管理したいとの意向が働いたためと考えられる。管理上の検討については、一般には概念モデル原型の作成後に行う作業であり、支援ツールの出力結果が誤りであったわけではない。

6. まとめと検討課題

以上から、概念の抽出に関して、3.2で示した判断基準は初心者分析作業にも、あるいは支援ツールを用いた機械的な適用に対しても有効に機能することが検証できた。しかしながら、さらに広範囲の概念モデル構築に役立つには、以下のような検討課題を残している。

(1) 概念間での自己関連

例えば、図22で示すような事例を考える。このとき、社員番号と氏名を属性候補として含む概念候補に、上司と部下の「役割」の相違があることを示す必要が生じたとき、役割が基本データとして明示されていない限り、現在の判断基準は役立たない。図22に示すように仮想的に基本データ“役割”を導入したとしても、“役割”を属性候補として含む概念候補が作成されるだけであり、上司・部下の関係を関連で表すことはできない。

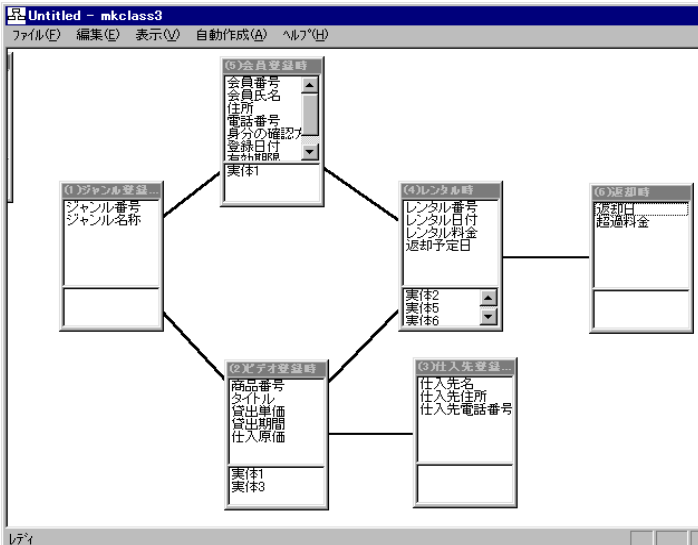


図20 事例データ4から抽出した概念モデル図原型
Fig.20 Output Conceptual Model from Sample Data4

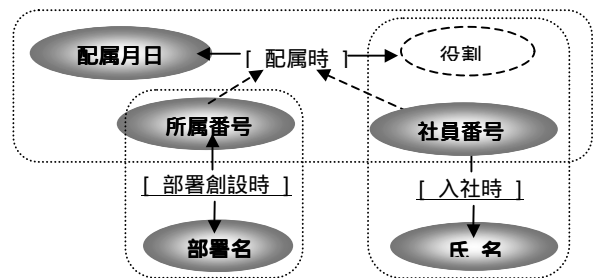


図22 役割の相違の扱い方
Fig.22 Basic Idea of how to treat the role

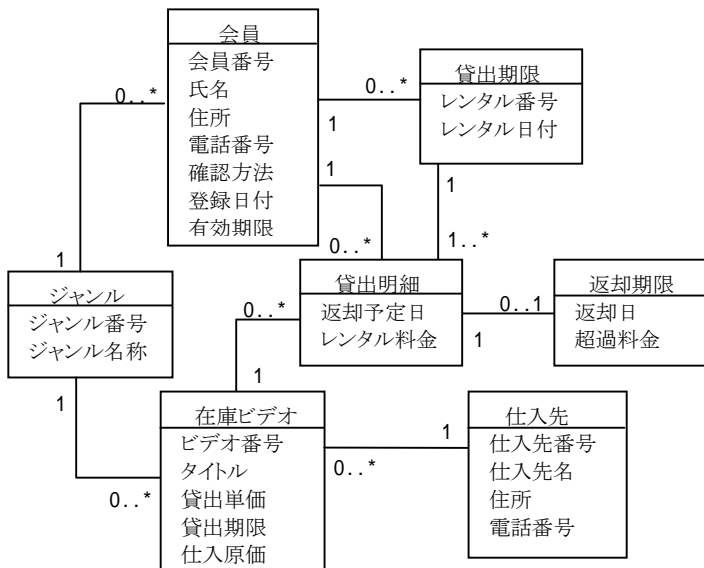


図21 事例データ4もとに分析者が抽出した概念モデル
Fig.21 Output Conceptual Model from Sample Data4
by Analyst

(2) 決定時点の認識方法

本論で述べた判断基準は、実現値の決定時点が大きな役割を果たしており、その認識が判断基準の適用結果を左右する。DATARUNにおいてもPDGの発見を重要視しており、ビジネスモデルからそれらを見出すとしている。しかし、概念モデルの構築は、ビジネスを対象としたシステム以外でも、重要な作業であることから、一般的な決定時点の認識方法を定める必要がある。

7. おわりに

入力された基本データを分類判断基準によって分類する機構については、筆者らが考案した協調オブジェクトの機構[18][19]に基づいている。今後、協調動作オブジ

エクトの研究と併せて、上述の課題に関する研究を進めて行きたい。なお、本研究の一部は情報サービス産業 HITOCC の補助金をもとに行われている。

参考文献

- [1]R.Wirfs-Brock, B.Wilkerson, " Object-Oriented Design: A Responsibility-Driven Approach," Proc of OOPSLA ' 89, ACM, pp. 71-75, 1989.
- [2]R.Wirfs-Brock, " Designing Objects and Their Interactions: A Brief Look at Responsibility-Driven Design, " Carroll, J. M. ed., Scenario-Based Design, John Wiley & Sons,1995.
- [3]S.Shaer,S.Mellor, " Object-Oriented Systems Analysis-Modeling the World in Data-, " Prentice-Hall, 1988.
- [4]Craig Larman, " Applying UML and Patterns:an introduction analysis and design " Prentice Hall,1998.
- [5]Rational Software Corporation, " UML Notation Guide, " <http://www.rational.com>
- [6]C.Sharble, S.Cohen, " The Object-Oriented Brewery: A Comparison of Two Object-Oriented Development Methods, " ACM SIGSOFT SE Notes, Vol. 18, No. 2, pp. 60-73, 1993.
- [7]K.Rubin, A.Goldberg, " Object Behavior Analysis, " CACM, Vol.35, No.9, pp. 48-62 ,1992.
- [8]B.Beck, W.Cunningham, " A Laboratory for Teaching Object-Oriented Thinking, " Proc. of OOPSLA'89, ACM, pp.1-6 ,1989.
- [9]W.Harrison, H. Ossher, " Subject-Oriented Programming (A Critique of Pure Objects), " Proc. of OOPSLA ' 93 ACM, pp. 411-428 ,1993.
- [10]M.Fowler, " Analysis Patterns: Reusable Object Models, " Addison-Wesley ,1997.
- [11]磯田定宏: " 実世界モデル化有害論-オブジェクトモデル化技法の解明, " 電子情報通信学会論文誌 D-I Vol.83, No.9, pp.946-959,2000.
- [12]D.パスコット, " C/S データベース設計入門 DATARUN, " 日経 BP, 1997.
- [13] 加藤貞行, " オブジェクト識別についての一考察とその効果, " 情報処理学会 研究報告, SE Vol . 98, No . 100, 1998.
- [14]大木幹雄, " データベース設計の基礎, " 日本理工出版会,1998.
- [15]大木幹雄, " 協調動作機構の実体関連分析プロセスへの適用, " 情報処理学会 00 ' 99 シンポジウム, pp.127-136, 1999.
- [16]大木幹雄, 稲田晃, " クラスモデリングにおける判断基準, " 電子情報通信学会 信学技報 Vol.99,No.312, pp. 9-16,1999.
- [17]大木幹雄,秋山構平, " 多重継承を含むクラス自動抽出の検証, " 電子情報通信学会 信学技報 Vol.100 No.90 pp.25-30,2000.
- [18] 大木幹雄, 坂本康治, " 動的な依存関係による分散オブジェクトの協調機構, " オブジェクト指向最前線 ' 98, 情報処理学会 00 ' 98 シンポジウム, pp.133-140, 1998.
- [19] 大木幹雄, " 動的な依存関係を利用した分散オブジェクトの協調機構に関する提案, " 電子情報通信学会 信学技報 KBSE Vol . 98, No . 238, 1998.