

第6章

国の分離・統合 —— 時間データモデルの適用 ——

坂 本 英 陽

はじめに

統計データベースには種々のメタデータ（データに関するデータ）が必要である。アジア経済研究所の世界貿易統計データベース（AIDXT）は、データ保有状況、品目分類、国・関税地域、データ記録上の制約等の情報がメタデータである。データベースが利用者にとって容易に使いデータ処理を誤らないために、これらのメタデータが使いやすい形に構造化されてデータベースに蓄えられていることが望まれる。そのためにはメタデータを表現するデータモデルが必要となるが、メタデータの構造は多種多様かつ複雑であり、関係データベースに代表される伝統的データベース管理システム（DBMS）では処理できないことが多い。これは伝統的DBMSが事務処理を初めとする一般業務を主たる対象にして、単純なデータ構造（関係データベースではリレーションと呼ばれる単純な表構造）を前提に開発されてきたことによる。このために伝統的データベースシステムの限界が主張されており、エンジニアリング分野ではオブジェクト指向データベース管理システム（OODBMS）、統計分野では統計データベース管理システム（SDBMS）が研究されており、既に実用化されているものもある。しかし真のヒューマンニーズに合致したシステムにはまだまだ程遠く、依然として研究・開発途上の技術である。

本章は、AIDXTのメタデータの一つである国・関税地域の分離・統合に関する知識の表現と利用について考察する。貿易統計は1国のある期間における複数の取引相手国に対するある商品の流れを記録するものであり、国に関する情報は統計利用上重要である。世界には多くの国があり、また東ドイツと西ドイツの統合、シンガポールのマレーシアからの分離、ビルマのミャンマーへの国名変更等時間と共に変化しており、正しい国に関する知識を得るためには努力を要する。また実際世界での国の統合・分離の時期とそれがデータに反映され記録される時期とのタイムラグの存在は、データ処理において間違

いを起こす原因になるので注意を要する。

国の分離・統合は時間に関係しており、この知識を表現するためにはデータモデルに時間を組み込むことが必要である。このような時間に関するモデルを時間データモデルという。時間データモデルを扱えるDBMSは少ない[Segev 88]。現存するシステムではデータベースは単に現在のデータを蓄えておけばいいという考えから設計されている。しかし、時間に関連するデータを扱う応用分野は非常に多く、これまでも多くの研究開発がなされてきた。これらの研究は、（1）関係データベースのような伝統的データモデルを時間を扱えるように拡張したもの[Ariav 84, Clifford 87, Snodgrass 87]と、（2）既存のデータモデルと独立に、時間データを概念化し、特徴を調べ挙げることにより、時制データモデルをデザインするためのフレームワークを作成するSegevとSnodgrassの2つのタイプに類別される[Segev 87, Snodgrass 85]。SegevとShoshaniは多くの時間データモデルの研究が伝統的データモデルの拡張によりなされているのに対して、既存のデータモデルとは独立に時間データの持っている意味（セマンティック）を調べ、その特徴を明らかにしたうえで時間に関する種々の演算を規定している[Segev87]。彼らはこのアプローチを時間データを扱うデータモデルのフレームワークを開発する目的で採用しており、（2）の研究成果を特定のデータモデルに移植可能であるという立場をとっている。彼らはこの成果を関係データモデルに応用している[Segev88]。

本章ではさらにSegevの時間データモデルを概念拡張し、AIDXTの国・関税地域の時間に伴う変化を扱える時間データモデル概念と基本演算を提案する[Segev 87]。第1節から第3節までで、国・関税地域に関するAIDXTの問題点を明らかにし、国・関税地域メタデータの重要性を述べる。第4節では始めにSegevの時間データモデルの基本概念を概観し、その後この概念を国・関税地域問題を扱えるように拡張する。最後に拡張データモデルによる基本演算を紹介する。

第1節 貿易統計における国および関税地域

貿易統計は国および関税地域間の商品の流れの記録である（以下、国および関税地域を単に国と呼ぶ）。そのため国に関する知識は貿易統計を利用する上で必要不可欠なものと言える。貿易統計で用いられる報告国や相手国の「国」という概念は政治的に独立した国あるいは保護国とは必ずしも同じではない（注1）。貿易統計では植民地等でも独立の関税地域と見なせる地域を国として取り扱い、独立国や保護国も含めて統計地域という。しかも統計地域の範囲は各国独自の判断にまかされている。貨物の国別分類である貿易相手国は、貿易という経済的な行為を反映させることを目的として地理的に分類しているため、政治上の国の領域、または国境とは必ずしも一致しない。そのため、政治的に承認されていない国であっても1地域としてとらえている場合がある。また、香港のように英国の属領であっても別個の1地域として分類している貿易相手国もある。このような分類方法がとられているのは貨物の地理的な移動状況をより詳細に把握するためにほかならない。しかも、貿易の相手国となる国別の選択基準は貿易の形態が多様であることから様々な説があり、条約、国連勧告等においてもいくつかの方法は示しているものの統計的な定義は出されていない。このため、各国においてもその国情に応じた選定の方法を採用しているのが実状である。

UNおよびOECDの各統計局が貿易相手国としての国の選定を各国の判断にまかせているという状況のもとでは、このような各国における異なる統計地域の選定方法は当然AIDXTのデータ源であるUNやOECD貿易統計および台湾貿易統計にも影響を及ぼしてくる。UN貿易統計の国の範疇は「United Nations Standard Country Code」、OECDのそれは「Foreign Trade by Commodity Series C」の中の「OECD Geographical Nomenclature」で明らかにされている。例外事項が微妙に異なることを除けば多くの場合には国の範疇がUNとOECD貿易統計でほぼ同じなので問題は起こらないが台湾、旧植民地、アジア太平洋諸国・地域等では包括範囲が必ずしも同一とは限らないし、特殊地域の扱いかたも異なる（注2）。特に、台湾についてはOECD貿易統計では相手国として明確にコード化され取引額も計上されているのに対して、UN貿易統計では1970年以降はそれが計上されていない。UN貿易統計「Commodity Trade Statistics」の前書きでは、台湾に計上されるはずの取引額は「その他のアジア

ア」に計上されていることが明記されている。

国際比較にさいしてはUNおよびOECDの国コードを統一した概念で対応できるような別の体系が必要になる。アジア経済研究所の作成した「アジア研究統一国コード」はそうしたひとつの試みである。「GRT₃₂(IDE)およびアジア研究統一国コード表」の表2-1に「アジア研究統一国コード表」の一覧表が示されている。この表はアジア研究統一国コードを軸としてUNとOECD貿易統計の国コードを互いに関連づけたものである。この表からわかるように小さな国などはUNではコード化されているにもかかわらず、アジア研究統一コードはotherの箇所に対応しているものもある。

アジア研究統一コードの作成にあたってはアジア経済研究所が東南アジア各国の出版する貿易統計書から得られた取り引き額をもとに東南アジア貿易統計マトリクスを作成するため、「統一国名表」（統計参考資料1963-6-21 no.8、アジア経済研究所、1963年）において統計地域の体系をまとめてコード化したことがある。この段階の国コードはUNの編集する貿易統計の存在を確認できなかったため、アジア経済研究所独自の体系によるコード化であった。また、統計地域をコード化したという経験は日本貿易統計のみに限られてはいたが、相手国の時系列のつなぎを検討した「日本貿易統計時系列に関するノート」（統計参考資料1968-11 no.34、アジア経済研究所、1968年）、「日本貿易統計相手国名変遷表」（統計参考資料1967-8 no.24、アジア経済研究所、1967年）、「アジア研究編成『日本貿易統計時系列磁気テープ』利用の手引き」（統計参考資料1970-9 no.39、アジア経済研究所、1970年）で生かされ、相手国の統合および分割の状態が明らかにされている。また、アジア研究統一国コードの応用による世界貿易統計の評価は「世界貿易マトリクスおよび中国の貿易統計推計作業ノート（電子計算機によるデータ処理手順）」（統計参考資料73-5 no.82、アジア経済研究所、1973年）以降いろいろな場面で試みられている。

「GRT₃₂(IDE)およびアジア研究統一国コード表」の表2-1の「アジア研究統一国コード表」ではUNとOECDにおける国の比較は可能であるが、国の時間的な変化に対しては対応できない。国の範囲は歴史的に新生、分割および統合等の変遷が見られるため、貿易統計を時系列で利用しようとすれば国の範囲の検討は避けて通れない。特に、貿易統計を相手国別の時系列としてみる場合、時系列の全期間を通して最細分の単一相手国別のデータが得られることが望ましい。しかしながら、政治的な統合、独立、

経済的な重要性の変化、関税地域の設定の変更のために必ずしも単一国のデータが持続的に得られるとは限らない。得られるのは新生、分割および統合された国のデータである。UNの場合は国や関税地域の変更はそのつどUNの出版物である「International Trade Statistics Year Book」に記載されている国の一覧を比較して確かめる。OECDについては磁気テープを購入したときに国の変更のテーブルが付いてくるのでそのテーブルで確かめることができる。それらの情報や磁気テープから得られた情報を取りまとめて一覧表を作り、統合や分割の変遷を明らかにすると便利である。国コードの変遷を検討するにあたっては国コードの範疇の基準化や統一化が必要である。統一化の基準としては、(1) 国名あるいは国コードが年次間で異なっても実質的に同一地域を意味する場合には同一国とする、(2) 同一国名あるいは国コードであっても実質的にその包括範囲が異なる場合には別個の範疇の国とする、(3) 国の統合あるいは分割が行われた場合には互いに関連する国をその包括範囲に入れる、ことが考えられる。このようにして作成された国の変遷を明示的に示したのが「GRT₃₂(IDE)およびアジア研統一国コード表」の表2-2の「アジア研統一国コード推移表」である。前述したように国の選定を各国の判断にまかせているので、正確な情報を得ようとするには国ごとにこのような表が必要になる。

第2節 国分類表の問題点

本節では、国の統合・分離等に関しAIDXTデータベースの現状と問題点を明らかにする(注3)。AIDXTデータの国の範疇は「GRT₃₂(IDE)およびアジア研統一国コード表」の表2-1「アジア研統一国コード表」により国分類として仕分けされている。表2-1で表された国分類は国の分離・統合・名称変更に伴う履歴情報(国カテゴリーの時間的変化)を保有しておらず、現在の状態を表わしており、更新前の状態は破棄されてしまっている。このためデータベース利用の際次のような問題が生じる。

(1) 国の分離・統合の場合、分離・統合前と後の国カテゴリーが明確に区別されないで混在している。そのため、国を特定することが困難な場合がある。例として、ドイツ、ベトナム、マレーシア、シンガポールを取り上げる。ドイツについては1990年10月、ドイツ連邦共和国(西ドイツ)とドイツ民主共和国(東ドイツ)の2国は統一し、国名をドイツ連邦共和国(統一ドイツ)とした。「GRT₃₂(IDE)

およびアジア研統一国コード表」の表2-1の「アジア研統一国コード表」によればこの2国は、

223680 GERMANY FR

224660 GERM. S.Z.

と表されている。ここで、左側の6桁コードはアジア研統一国コード、右側は国名を表す。この国分類表ではGERMANY FRが統合前か後かの区別がつかない。そこでこの表2-2「アジア研統一国コード推移表」を利用すれば1990年までのGERMANY FRは統合前、1991年以降は統合後ということがわかる。ただし、すべての報告国が同一年度で相手国ドイツを切り替えているかどうかは明らかでない。ベトナムについてはVIETNAM N(北ベトナム)とVIET NAM(ベトナム)の2カテゴリーが分類表に含まれている。

128060 VIET NAM

127240 VIETNAM N

これから現在のベトナムがどちらのカテゴリーに属するのか判断できないが、「GRT₃₂(IDE)およびアジア研統一国コード表」の表2-2によればViet Namは1980年以前と以降で区別されていることがわかる。すべての報告国が同一年度で相手国ベトナムを切り替えているかどうかは明らかでない。またマレーシアの場合も1963年に西マレーシア、サバ、サラワクおよびシンガポールをあわせて独立連邦国マレーシアとなり、その後、1965年にシンガポールが分離独立した。表2-1による国分類は、

127650 MALAYSIA

127660 W MALAYSIA

127680 SINGAPORE

127740 SARAWAKU

127780 SABAH

であるが、127650で表されたマレーシアがシンガポールを含むのか含まないのか不明である。また、西マレーシア、サバ、サラワクとマレーシア、シンガポールとの関係も明示されていない。

(2) 国名の変化が記録されていない。例えばミャンマーとスリランカの場合国分類は次のようになっている。

137560 SRI LANKA

137540 MYANMAR

スリランカとミャンマーがそれぞれセイロンとビルマと呼ばれていた事実が明示されていない。表2-2からCeylonは1973年からSri Lankaに、またBurmaは1990年以降Myanmarに国名が変更されたことがわかる。貿易統計データのすべての報告国が同一年度で相手国名の変更を切り替えているかどうかは明らかではない。

このように表2-1および表2-2で表された国分類は利用者にとって分かりにくく、誤った処理をする可能性がある。またAIDXTにおいて利用者のデータチェックにかかる負担をシステムが代行するためにも、国の分離・統合に関する知識をメタデータとしてデータベースに具備する必要がある。

第3節 国の分離・統合にともなうデータ処理

現実世界で実際に起きた国の分離、統合および国名の変更以外に、データ処理の上で重要なものの一つに、データに現われる国カテゴリーが現実世界を反映していないこと、すなわち現実の国カテゴリーの変化がデータベースのデータに記録として現われるまでにタイムラグがあり、しかも報告国によりラグの大きさが異なることである。このデータ記録上のタイムラグに関する知識（メタデータ）の重要性を、ドイツとマレーシア・シンガポールを事例として以下に示す。

報告国ドイツの貿易統計は、1962年から1989年までドイツ連邦共和国（西ドイツ）、1990年が統一までの西ドイツと10月以降の統一ドイツの計、1991年以降が統一後のドイツ連邦共和国である（注4）。報告国東ドイツのデータはAIDXTには含まれていない。相手国に関しては、1989年まではドイツ連邦共和国（西ドイツ）とドイツ民主共和国（東ドイツ）2国が使われている。1991年以降は統一ドイツ一国で記載されている。1990年は次の2つのタイプが挙げられる。

- （1）1989年までと同様にドイツ連邦共和国とドイツ民主共和国を使用している。
- （2）ドイツ連邦共和国のみでドイツ民主共和国を使用していない。

（1）の場合、ドイツ連邦共和国は統一までの西ドイツと10月以降の統一ドイツの計であり、ドイツ民主共和国は統一までの東ドイツを表わす（図1イ）。AIDXTデータベースでは、大部分の報告国がこの方式をとっている。（2）の場合、1990年のデータを統一ドイツ一国にまとめたもの（図1ロ）で、統一までの東西ドイツと10月以降の統一ドイツの計を表わす。OECD編集のデータではオーストラリアとオーストラリアが、国連編集のデータでは香港、インドなど少数の国がこの方式を採用している。

データを利用するときこの知識は重要である。例えば、1990年の日本とオーストラリアのドイツ連邦共和国への輸出を扱っている場合、日本の統計では

東ドイツは含まれていないのに対し、オーストラリアの統計では東ドイツも含まれている。この知識なしで相手国「ドイツ連邦共和国」を指定したときは誤った結果を得ることになる。

1990年の両タイプのデータを同時に使う場合、分析目的により2つの方法が考えられる。

（1）1990年は、1991年以降と同様に統一ドイツとして扱う。この場合、東ドイツが明記されているデータは、ドイツ連邦共和国（統一ドイツ）とドイツ民主共和国（東ドイツ）を合計してドイツ連邦共和国（統一ドイツ）として扱う。

（2）1990年は、ドイツ連邦共和国と東ドイツに分けて処理する。この場合、ドイツ連邦共和国（統一ドイツ）一国で記載されているデータについては、東ドイツを含まないドイツ連邦共和国とドイツ民主共和国に分割する必要がある。東ドイツを含まないドイツ連邦共和国は、報告国「ドイツ連邦共和国（統一ドイツ）」を使い逆推計（注5）により代替値を求める。東ドイツは、東ドイツを含んでいるドイツ連邦共和国からこの逆推計で求めた西ドイツを差し引いて求める。

図1 ドイツの国表示（統一前後）

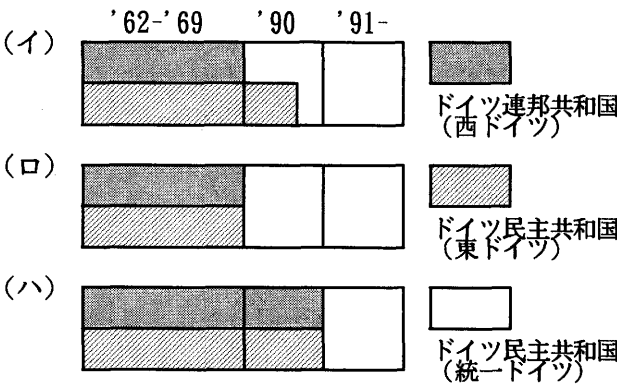


図2 マレーシアとシンガポールの貿易統計データ

1962	...	1967	1968	...	1992
西マレーシア			マレーシア		
サバ					
サラワク					
シンガポール					

次はマレーシアとシンガポールの場合である。図2はUN編集の報告国のカバリッジを示している。シンガポールは1965年に独立したにもかかわらず1962年からシンガポールとして利用できる。マレーシアについては、1962年から1967までは西マレーシア、サバ、サラワクの3地域に分かれており、1968

年からマレーシアとして利用できる。

OECD編集のデータと国連編集のデータとでは相手国の扱いが異なる。台湾貿易統計は国連編集の相手国の扱いが同じである。OECD編集のデータは、1962年から1965年までをマレーシア1国でまとめ、1966年以降はマレーシアとシンガポール別々に表記している（図3）。ただしフランスとドイツは例外で、1967年まではマレーシア1国で、1968年以降はマレーシアとシンガポールを別々に表記している。

図3 OECDデータの相手国マレーシア・シンガポール表示

1962	...	1965	1966	...	1992
マレーシア			マレーシア		
			シンガポール		

図4は、国連編集のデータと台湾データの相手国の取り扱いを示している。シンガポールは1962年から単独で表記されている。西マレーシア、サバ、サラワクの3地域は1962年から使われており、その後はマレーシアとして表記されている。相手国としてマレーシアを使用しはじめた年次は、フィリピン、香港、シンガポールはそれぞれ1973年、1974年、1976年から、モロッコ、ブラジル、イランはそれぞれ1962年、1964年、1965年からと言うように国によりばらつきがある。オーストラリア、ニュージーランド、タイ、インドは注意を要する例外である。オーストラリア、ニュージーランド（図5）は1962-1963年はシンガポールと西マレーシア、サバ、サラワクに分かれている。1964-1965年はマレーシア1国で、それ以降はマレーシア及びシンガポールの2国で表示されている。図6はタイとインドの相手国表示である。

図4 国連データと台湾データの相手国マレーシア・シンガポール表示

1962	1992
西マレーシア		マレーシア	
サバ			
サラワク			
シンガポール			

図5 オーストラリアとニュージーランドの相手国マレーシア・シンガポール表示

1962, 1963	1964, 1965	1966 …
西マレーシア	マレーシア	マレーシア
サバ		
サラワク		
シンガポール		シンガポール

図6 タイとインドの相手国マレーシア・シンガポール表示

(インド)			
1962-1973	1974	1975	1976…
西マレーシア	マレーシア	西マレーシア	マレーシア
サバ		サバ	
サラワク		サラワク	
シンガポール		シンガポール	シンガポール

(タイ)			
1962-1965	1966	1967-1969	1970…
西マレーシア	マレーシア	西マレーシア	マレーシア
サバ		サバ	
サラワク		サラワク	
シンガポール		シンガポール	シンガポール

データを利用するとき、この知識は重要である。例えば、1965年以降のフィリピン、シンガポールのマレーシアへの輸出入を調べているとき、単に相手国として「マレーシア」を指定すると、フィリピンの場合は1972年まで、シンガポールの場合は1975年までデータが検索されないことになってしまう（相手国として西マレーシア、サバ、サラワクを使用しているため）。また日本とフランスの1977年のマレーシアへの輸出入を扱うために相手国をマレーシアと指定すると、日本はマレーシアにシンガポールを含んでいないが、フランスはマレーシアにシンガポールの貿易額を含めているので、誤った解釈をしてしまう可能性がある。

貿易マトリクス作成上の処理は、つぎの方法が考えられる。

(1) 西マレーシア、サバ、サラワク3地域を合計して、マレーシアとする。報告国のたしあげの場合は域内取引（西マレーシアからサバ、サラワクへの輸出等）を調整する必要がある。

(2) 相手国について、シンガポールを含んでいるマレーシアをシンガポールとマレーシア（シンガポールは含まない）に分ける。この場合、1967年まで

は報告国「西マレーシア」、「サバ」、「サラワク」の3地域をたしあげたもの、または報告国「シンガポール」を使って逆推計する。1968年以降は、「報告国マレーシア」または「シンガポール」を使って逆推計する。

これらのメタデータは現行AIDXTデータベースには具備されていない。利用者がそれぞれデータを調べ、分析目的に合わせて必要とする知識を得ている。誤ったデータ処理を防ぎ、利用者の負担を軽減するために、このメタデータをデータベース化する必要がある。

第4節 時間データモデル (TS、TSC) の基本概念と拡張

前節で見たように国の分離・統合に伴うデータ処理は時間に関係しているために複雑である。データベース利用者の負担を軽減するためにデータ表現方法とそのデータに対する操作(演算)を規定したデータモデルを新たに開発する必要がある。このため本章では国の分離・統合に関する処理ができるように、SegevとShoshaniの時間データモデルの基本概念であるTS(タイムシーケンス)とTSC(タイムシーケンスコレクション)を拡張する[Segev 87]。その後次節で、前節で述べた国の分離・統合にともなうデータ処理を可能にする基本演算であるselectとcompareを例示する。

4-1 TSとTSC

Segevのデータモデルの基本概念はTS(タイムシーケンス)である。種々のオブジェクトは時間と共に変化する属性値を持っている。例えば、企業データベースでは、従業員オブジェクトは時間と共に値が変化する給与属性を持っている。また、貿易データベースでは、報告国オブジェクトは輸出額、輸入額など、時間と共に値が変化する属性を持っている。このオブジェクトの属性値の時間に関する順序列をTSと呼ぶ。

貿易統計データベースの例では、S国のt年の総輸出額を $x_s(t)$ と表したとき、総輸出額属性値の時間順序列は、

$$TS = \{ x_s(t_1), x_s(t_2), \dots, x_s(t_n) \}$$

である。ここで、 t_1, \dots, t_n は時間を表す。

クラスは同じ属性を持つオブジェクトの集まりである。同じクラスに属するTSの集まりをTSC(タイムシーケンスコレクション)と呼ぶ。先の貿易統計データベースの例では S_i 国のTSを TS_i と表すと

き、

$$TSC = \{ TS_1, TS_2, \dots, TS_m \}$$

となるTSの集合がTSCである。図7はTSCを2次元で表示したものである。時間順序列 TS_1, TS_2, \dots はそれぞれ国オブジェクト S_1, S_2, \dots のTSである。

図7 TSCの2次元表示

	t_1	t_2	...	t_n
TS_1	$x_{s1}(t_1)$	$x_{s1}(t_2)$...	$x_{s1}(t_n)$
TS_2	\vdots	\vdots	...	\vdots
\vdots	\vdots	\vdots	...	\vdots
TS_i	$x_{si}(t_1)$	$x_{si}(t_2)$...	$x_{si}(t_n)$
\vdots	\vdots	\vdots	...	\vdots
\vdots	\vdots	\vdots	...	\vdots
TS_m	$x_{sm}(t_1)$	$x_{sm}(t_2)$...	$x_{sm}(t_n)$

前述の貿易統計データベースの例では TS_1, TS_2, \dots, TS_n はそれぞれ日本、フランス、アメリカ等の国オブジェクトに対応するTSである。図8は各国オブジェクトを図示したものであり、有向線分→の上に表示されているのはオブジェクトの国名属性である。

TSCには、タイムグラニュアリティ、ライフスパン、タイプ、補間ルールという特性があり、時間データモデルの構成要素として定義される。以下これら諸概念について述べる。

(1) タイムグラニュアリティ

この特性はタイムポイントの細さを指示する。1,2,3,...という順序表示と年、月、日、...というカレンダー表示がある。

(2) ライフスパン

TSCのライフスパンは開始ポイントと終了ポイントで指示される。これは順序表示でもカレンダー表示でも可能である。

(3) タイプ

SegevはTSのタイムポイントとイベントポイントを区別している[Segev 88]。タイムポイントは、潜在的に取りうる可能性のある全ての時点である。一方、イベントポイントは、実際にTSCが値を持っている時点である。TSCにはステップワイズ、不連続、連続の3種類のタイプがある。タイプによってイベントポイントでないタイムポイントのデータ値を補間する方法が異なる。

以下は、日本の1972年から1993年までの輸出額を表すTSCである。

Object Class: 日本

Temporal Attribute Type: 輸出額

Time Granularity: 年

Life Span: Start=1972;End=1993

Type: ステップワイズ定数

このTSC定義は、輸出額が時間と共に変化する、1972年から1993年まで、各年で値が変化することを示している。また、ステップワイズ定数タイプを伝っているため、データが欠損している年次の輸出額を補間法で推定しないことを示している。SegevはTSCに対して時間を処理できる次の基本演算子を定義している。

select
aggregate
accumulate
restrict
composition

全ての基本演算子是一只またはそれ以上のTSCから一只のTSCを生み出す。そのため繰り返し基本演算子を適用することにより複雑な演算を実行できる。このうち、selectは1993年2月16日までというようなある条件を満足するTSだけを選択するような選択的抽象演算である。aggregateは地域別計の作成のような要約計算をおこなう。

4-2 TS概念の拡張

国の分離・統合と名称の変更は、次のように考える。

- (1) 国名の変更は、国クラスのあるオブジェクトの国名属性がある時点で変わった。
- (2) 国の統合は、国クラスの複数のオブジェクト

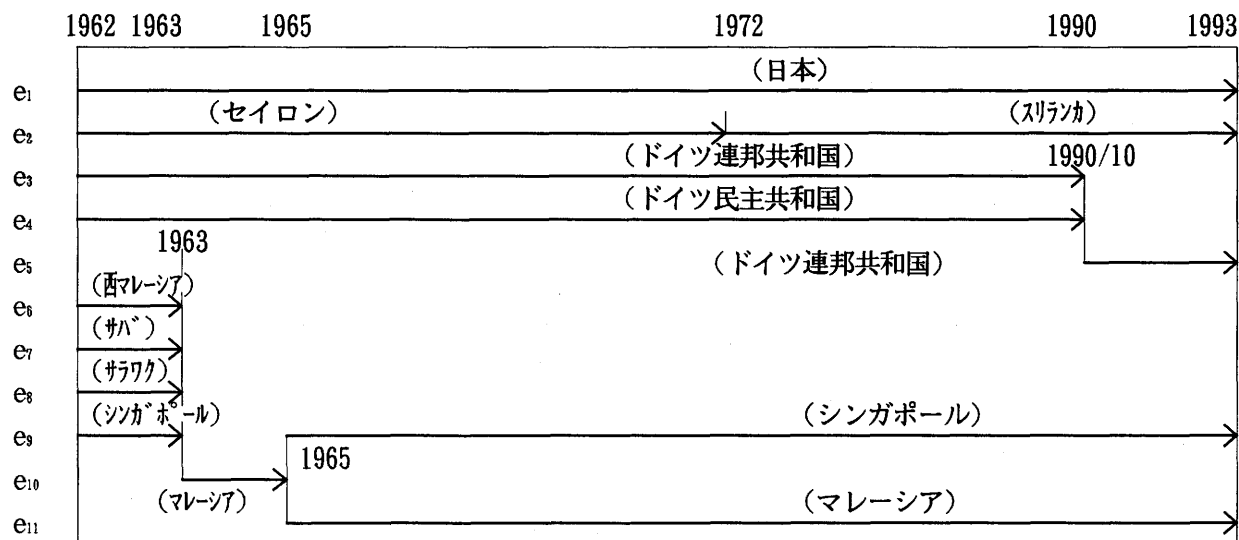
がある時点でまとまり一只の新しいオブジェクトが生成された。

(3) 国の分離は国クラスの一の国オブジェクトがある時点で複数の国オブジェクトに分割された。

(4) 国の一がある国またはその一部と合併された場合は分離と統合の組み合わせと考える。

前節で紹介したSegev 87のTSは同一クラスに属するオブジェクトの属性値の時間的変化を記述するモデルであるが、オブジェクト自体は不変であることを仮定している。従って国名の変更は、同一オブジェクトの国名属性値の変化とみなされるためTSモデルで表示することが可能であるが、国の分離・統合はクラス内の複数オブジェクト間の関係を扱う必要があり、TSモデルで記述することができない。図8は国分類(国クラス)の時系列変化をTSCの枠組のなかで表したものである。細線はオブジェクト間の関係を示している。オブジェクト e_1 の国名は日本であり、このオブジェクトは他のオブジェクトと関係を持たない。オブジェクト e_2 は国名がセイロンからスリランカに1972年に変わった。日本と同じく他のオブジェクトと関係を持たない。オブジェクト e_3, e_4, e_5 に関しては、オブジェクト e_3 と e_4 が統合して新たにオブジェクト e_5 が生成されたことを表す。このとき、(e_3, e_4, e_5)を関連オブジェクトグループと呼ぶ。オブジェクトグループ($e_6, e_7, e_8, e_9, e_{10}, e_{11}$)に関しては、オブジェクト e_6, e_7, e_8, e_9 が統合されて e_{10} となり、その後 e_{10} が分離して e_9 と e_{11} になったことを示す(注6)。

図8 国分類TSCx



このように国分類のオブジェクト間の関係は、有向グラフ（時間の流れにそったオブジェクト間のネットワーク）として表すことができる。TS、TSCという時間概念ではこのオブジェクト間の時間的關係を表現できない。以下に国の分離・統合を扱えるようにTSとTSCを概念拡張して関連オブジェクトTS_xと関連オブジェクトTSC_xを新たに定義する。

n個のオブジェクト e_1, e_2, \dots, e_n とそれに対応するTS₁, ..., TS_nが存在するとき、このオブジェクトの中で時間と共にまとまったり分かれたりする互いに関連するm個 ($n \geq m$) のオブジェクトの集合 $\{e'_1, e'_2, \dots, e'_m\}$ を関連オブジェクトグループと呼ぶ。

関連オブジェクトグループの要素 e'_1, e'_2, \dots, e'_m に対応するTSをTS'₁, TS'₂, ..., TS'_mとすると、このm個のTSの集合を拡張TSといい、

$$TS_x = \{TS'_1, TS'_2, \dots, TS'_m\}$$

で表す。定義からわかるとおり、拡張TSはTSCの部分集合である。さらに、TSCにk個の拡張TS、すなわち、TS_{x1}, TS_{x2}, ..., TS_{xk}が存在するとき、このk個の拡張TSの集合をTSC_xとする。

$$TSC_x = \{TS_{x1}, TS_{x2}, \dots, TS_{xk}\}$$

TSの集合がTSCであったのと同じように、TS_xの集合がTSC_xである。

図8の有向グラフを例に取れば、11個のオブジェクトとそれに対応するTSが存在する。また、関連するオブジェクトの塊は4個あるので、これに拡張TSを対応させれば、4個の拡張TSとなる。したがって、拡張TSであるTS_xは、

$$TS_{x1} = \{e_1\}$$

$$TS_{x2} = \{e_2\}$$

$$TS_{x3} = \{e_3, e_4, e_5\}$$

$$TS_{x4} = \{e_6, e_7, e_8, e_9, e_{10}, e_{11}\}$$

となる。このときの拡張TSCは、

$$TSC_x = \{TS_{x1}, TS_{x2}, \dots, TS_{x4}\}$$

である。オブジェクト e_1 は日本の国属性値を持ち、このTS₁は有効グラフの中では、

(日本)
→

と表される。また、オブジェクト e_2 のTS₂は1962年から1971年まではセイロン、1972年以降はスリランカの国属性値を持つ。これを図示すると、

(セイロン) (スリランカ)
→
1972

となる。

第5節 国に対する基本演算

この節では国の統合・分離を扱うための基本演算子について述べる。拡張モデルの基本演算子是一只またはそれ以上のTSC_xから一只または複数の新たなTSC_xを生み出す。そのため繰り返し基本演算子を適用することにより複雑な演算を実行できる。TSC_xはTSCの概念拡張であり、TSC_xの特殊な形がTSCであると見做すことができる。従ってTSC_xのオブジェクトが互いに関連していない（TSCと見做すことができる）ときは、TSC上で定義されているselect、aggregate、accumulate等の演算が可能である。互いに関連するオブジェクトを含むTSC_xでは、有向グラフを処理する新たな演算子が必要となる。ここでは国の分離・統合に関して必要となる基本的な演算子select、select*とcompare、compare*について述べる。

(1) select

国分類表は実際に世界で起きている国の分離・統合を時間的に表したTSC_x（図8）である。

(a)国分類TSC_xからオブジェクト e_3, e_4 について1970年から1992年まで検索する。

```
select WK1 国名
from 国分類
where s in (e3,e4) and t in (1970 to 1992)
```

国分類TSC_xからオブジェクトSが e_3, e_4 であり、かつ時間が1970年から1992年までの国名履歴をWK1と呼ばれるTSC_xにアウトプットする。結果は次のようになる。

WK1 TSC_x

	1970	1990/10	1992
e_3	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> (ドイツ連邦共和国) → </div> <div style="display: flex; justify-content: space-between; align-items: center;"> (ドイツ民主共和国) → </div> </div>		
e_4			

この結果より、1972年から1990年10月まではドイツ連邦共和国とドイツ民主共和国が現実世界に存在し、1990年10月以降変化が起こったことが見て取れる。

(a')国分類TSC_xからオブジェクト e_3, e_4 について1970年から1992年まで e_3, e_4 の関連オブジェクトグループも含めて検索する。関連オブジェクトを含めて検索する場合にはselect*を使用する。

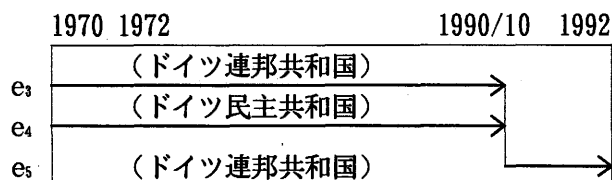
```
select* WK2 国名
```


from 国分類

where s in (e3,e4) and t in (1970 to 1992)

結果は次のようになる。

WK2 TSCx



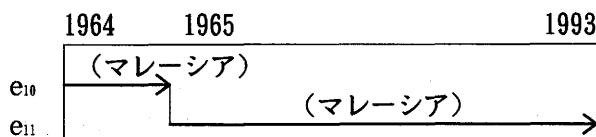
この結果から(a)の事実のほかに1990年10月からオブジェクトe3、e4が統合されて国名属性がドイツ連邦共和国であるe5が新生した。

(b)国名マレーシアを持つオブジェクトを国分類TSCxから1964年から1993年まで求める。

```
select WK3 国名
from 国分類
where 国名=マレーシア and t in (1964 to 1993)
```

結果は次のようになる。

WK3 TSCx



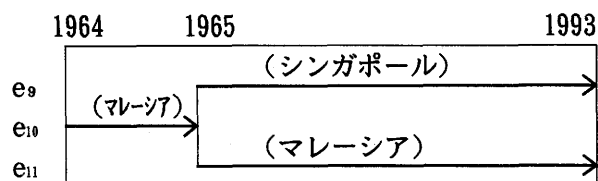
この結果より、マレーシアは1965年を境に異なったオブジェクトであることが見て取れる。

(b')国名マレーシアを持つオブジェクトを国分類TSCxから関連オブジェクトグループを含めて1964年から1993年まで求める。関連オブジェクトを含めて検索するときはselect*を使用する。

```
select* WK4 国名
from 国分類
where 国名=マレーシア and t in (1964 to 1993)
```

結果は次のようになる。

WK4 TSCx



この結果より、1965年以降オブジェクトe10はオブジェクトe9とe11に分かれ、国名がそれぞれシンガポールとマレーシアであることが見て取れる。

(2) compare

第1節で述べたように各国のデータで使われている相手国分類は、国の分離・統合に対して時間的なズレを持っている。ここでは日本とフランスのデータの記録に使われている相手国分類をそれぞれ日本データ相手国TSCx (図9)、フランスデータ相手国TSCx (図10) で表すことにする。

図9 日本データ相手国TSCx

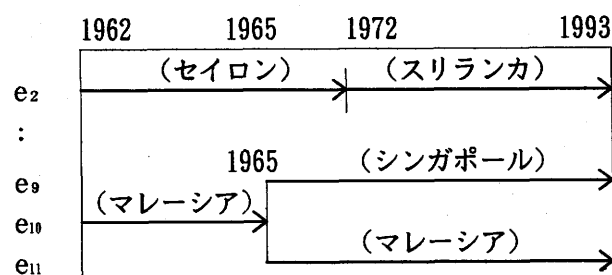
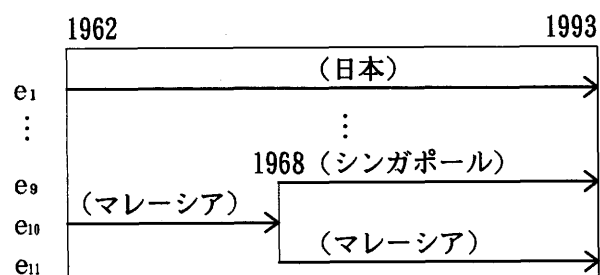


図10 フランスデータ相手国TSCx



(a)日本データ相手国TSCxとフランスデータ相手国TSCxにおいて、相手国マレーシアが同じオブジェクトである期間と異なる期間を求める。

```
compare (WK5,WK6) 国名
from 日本データ相手国分類、フランスデータ相手国分類
where 国名=マレーシア and t in (1962 to 1993)
```

日本データ相手国分類TSCxとフランスデータ相手

国分類TSCxを国名属性がマレーシアで時間が1962年から1993年までを比較して、同じ分類カテゴリーを使っている期間に対してはWK5に、異なる分類カテゴリーを使用している期間に対してはWK6にアウトプットする。結果は次のようになる。

WK5 TSCx

	1962	...	1965		1968	...	1993
e ₁₀	(マレーシア) 1965 →						
e ₁₁	1968 (マレーシア) →						

WK6 TSCx

	1962		1965		1968		1993
e ₁₀	(マレーシア) 1965 → 1968						
e ₁₁	→ (マレーシア)						

この結果より、報告国日本とフランスで相手国に関して、(1) 1962年から1964年までと1968年以降は同じオブジェクトとして国名マレーシアを表している、(2) 1965年から1968年までは国名マレーシアを別々のオブジェクトとして表している、ことが見て取れる。

(a') 日本データ相手国TSCxとフランスデータTSCxにおいてマレーシアおよびその関連オブジェクトグループが同じオブジェクトである期間と異なる期間を求める。関連オブジェクトを出力TSCxに含める場合にはcompare*を使用する。

```
compare* (WK7,WK8) 国名
from 日本データ相手国分類、フランスデータ相手国分類
where 国名=マレーシア and t in (1962 to 1993)
```

結果は次のようになる。

WK7 TSCx

	1962		1968	...	1993
e ₁₀	(マレーシア) 1965 →				
e ₁₁	1968 (シンガポール) →				
	(マレーシア) →				

WK8 TSCx

	1962		1993
e ₉	1965 (シンガポール) → 1968		
e ₁₀	(マレーシア) →		
e ₁₁	→ (マレーシア)		

この結果より上記で得られた結果に加えて、(1) シンガポールが1968年から1993年まで日本とフランスで同一オブジェクトとして表されている、(2) 日本とフランスではシンガポールが異なるオブジェクトとして表されている(フランスでは1965年から1968年まで相手国シンガポールは使われていない)。

本節では選択的データ抽出とデータ比較の基本2演算が例示されている。この演算機能だけでは第3節で述べたデータ処理を全ておこなうことはできない。そのため、本章で提示されているデータモデルを今後精緻化した上で、新たな基本演算を考案することが必須である。

おわりに

AIDXTのデータベースはUN、OECD貿易統計および台湾貿易統計から構成されている。各統計作成期間の国の範疇が必ずしも一致しないことからそれらを統一して取り扱えるように考えられた国の分類体系が「アジア統一国コード」である。国は時間と共に分離・統合、名称の変更、政治形態の変化等が繰り返し行なわれており、分離・統合した国が報告国・相手国として貿易統計に記載される時点は、現実に行っている分離・統合の時期と異なっている場合がある。貿易統計を利用して正しくデータ処理を行なうためには、国の範疇の調整と国の分離・統合に関する知識が必要である。

UNとOECDの異なるコード体系を統一したのが「GRT₃₂(IDE)およびアジア統一国コード表」の表2-1の「アジア統一国コード表」である。この表は1次元の表記なので時間の変化に対応する分離・統合の時点を表記できない。時間の変化に対する国の変遷を表したのがつぎの表2-2の「アジア統一国コード推移表」である。国の選定は各国の判断にまかせられているので、UNまたはOECDの国体系の中でも各国ごとに国の分離・統合の計上時点が必ずしも同一ではなく、国ごとの情報は表2-2では得ることができない。国ごとの正確な情報を得ようとするには国の数に相当する表2-2が要求され膨大な一覧表が必要になる。

国の分離・統合に関するデータ処理は貿易データ

ベースを扱う際必要となるが、国の分離・統合は時間と関係しているため複雑な処理手順をデータベース利用者に要求する。利用者の負担を軽減するために新たな時間データモデルを開発する必要がある。

SegevとShoshaniの時間データモデルは、時間データを扱うデータモデルのフレームワークを作ることとを目的に開発された論理モデルであるが、このままのモデルでは国の統合・分割の時間的変化の問題を処理できない。そのため彼らの論理時間データモデルを概念拡張し、国の分離・統合の時間的変化を扱う基本演算子を提示したのが拡張データモデルである。この拡張データモデルはまだ研究の緒に終わったばかりであり、現在AIDXTで課題となっている国の分離・統合に関する処理の一部を可能にするに留まる。そのため、AIDXTの国の分離・統合メタデータベースを構築するためには、更にこのデータモデルが精緻化され、新たな演算機能が案出される必要がある。

「GRT₃₂(IDE)およびアジア統一国コード表」の表2-2「アジア統一国コード推移表」はTSCxに形が似ているが、前者は国コードの推移を具体的に表示したものである。それに対して、後者は「意味データモデル」における主体(entity)の概念をもとにした形式的データモデルである。

(注1) 山岸一夫著による「わが国の貿易統計について—外国貿易等に関する統計基本通達とその内容—」(『貿易実務ダイジェスト』1991年5月号)を参照のこと。

(注2) 国コードには特殊地域(Special Category)があり、UNとOECD貿易統計の相手国コードとして使用されているコードとその名称は下表に示される。

UN貿易統計	OECD貿易統計
90900 free zone	8110 ships supplies
90896 area nes.	8150 miscellaneous nes
90898 unspecified	8210 secret
90904 for ships	9998 difference
90908 spec cat.	

(注3) 統計データベースが企業データベースと比べてより多くのメタデータを必要とする理由の一つに、企業データは個々の組織から作られるデータであり、例外事項も含めデータに関する情報が広く組織内で把握されているのに反し、統計データベース

では一般的にデータは他機関で作られたものであり、データに対する情報が不足している。AIDXTデータベースではデータ源は磁気テープの形で入手したOECDと国連編集の世界貿易統計と台湾政府作成の貿易統計である。台湾のデータは国連から得られないため、アジア研で台湾政府から入手したデータを再編集してAIDXTデータベースに加えている。OECD加盟国のデータはOECDから、その他の国のデータは国連から入手している。オーストラリアとニュージーランドはOECD加盟国であるが1987年までは国連編集のデータである。データを入手する際磁気テープと共にデータスペックが送られてくるが、データ利用に必要な全ての情報が含まれているわけではない。国に関するメタデータにおいても例外ではない。

(注4) 例えば1990年のドイツの年データの場合、データの内容として次の3とおりのケース(前出図1)が考えられる。(イ)統一までの西ドイツと10月以降の統一ドイツの計、(ロ)統一までの西ドイツと東ドイツ、および10月以降の統一ドイツの計、(ハ)統一までの西ドイツと10月以降の統一ドイツの旧西ドイツにあたる部分の計。このうちどれが正しいかは入手したデータスペックから判断することができなかった。またOECDと国連刊行の出版物からも情報を得ることができなかった。このようにメタデータの問題は、メタデータを収集することから始める必要がある。本章はメタデータそのものを述べるのではなく、メタデータの重要性について述べ、表現方法をモデル化することであるのでこれ以上この問題に係わらない。そのためドイツの1990年の事例ではドイツデータを(イ)統一までの西ドイツと10月以降の統一ドイツの計であるとした。以下の相手国がドイツの事例およびマレーシア、シンガポールの事例はデータを時系列的に調べて判断しているが、推測を伴う作業であるため多少の判断ミスがあると考えられる。これも本章の目的に甘んじて許しを乞うことにする。

(注5) A国のB国への輸出(輸入)をB国のA国への輸入(輸出)で代替することを逆推計と言う。

(注6) (イ)オブジェクト e_2 と e_5 の国名が同じドイツ連邦共和国であることに注意を払う必要がある。国名が同じでも異なったオブジェクトであると考えたほうが、国の分離・統合の時間的変化を扱うのに都合がいいためである。(ロ)オブジェクト e_9 が e_{10} に統合されたあと e_{10} が分離し、再度 e_9 が現われている。異なるオブジェクトとして表現可能であるが、ここでは国の地理的定義が同じ場合は同一オブジェクトと考えた。(ハ) e_{11} は e_{10} から分かれたオ

プロジェクトであるが、国名は両方ともマレーシアである。地理的定義が異なるために異なるオブジェクトと考えたほうが都合がいいためである。

【参考文献】

[1] 山岸一夫「わが国の貿易統計について—外国貿易等に関する統計基本通達とその内容—」（『貿易実務ダイジェスト』1991年5月号）。

[2] Clifford, J. L, Croker, A. , The Historical Relational Data Model(HRDM) and Algebra Based on Life Spans, Proceedings of the Third International Conference on Data Engineering, February 1987, pp .528-537.

[3] McCarthy, J. L., Metadata Management for Large Statistical Databases, Proceedings of the Eighth International Conference on Very Large Data Bases,

September 1982, pp.234-243.

[4] Segev, A., Shoshani, A., Logical Modeling of Temporal Semantics, Proceedings of the ACM SIGMOD International Conference on Management of Data, May 1987, pp.454-466.

[5] Segev, A., Shoshani, A., The Representation of a Temporal Data Model in the Relational Environment, Proceedings of the 4th International Conference on Statistical and Scientific Database management, 1988, pp.39-61.

[6] Snodgrass, R., Ahn, I., A Taxonomy of Time in Databases, Proceedings of the ACM SIGMOD International Conference on Management of Data, May 1985, pp.236-246.

[7] Snodgrass, R., The Temporal Query Language TQuel, ACM Transactions on Database Systems, 12,2, June 1987, pp.247-298.