情 報 学 基 礎 46－1
（1 9 9 7． 7． 2 4）

表形式データの情報論的検討

古閑 政
九州東海大学工学部経営管理学科
E-mail:koga@bm-1.ktokai-u.ac.jp

（概要）

企業における経営資料を初めとして，各種データには表形式で現されるものが多い。しかし，それはデータのまとめ方として行われるのであり，大量のデータの場合等一般的に決して見やすいものではない。その結果，表の一部を抜粋したり，それをグラフ表示したりして，有益な情報を得やすいようにする工夫が行われている。これは経験に基づいて行われているだけで，今までその過程を理論的に検討した例は無いと思われる．

この報告では，表形式のデータが有する情報量を情報エントロピーによって表現することにより，表示変換を行ったときの情報量がどう変化するかを検討している。その具体例として，データ構成が３項目２要素の簡単な事例を取り上げ，詳細な検討を行った。この際，重要な示唆を与えるのは，システム情報量という指標であり，データ間の関連度の強さを示すものである．また，最近注目されるデータマイニングにおける有用性についてはは未検討であるが，その可能性が期待される．

Study of Information about Table-form Data

Masashi Koga

Kyushutokai University
School of Business Engineering, Faculty of Engineering
9 - 1 - 1, Toroku, Kumamoto City, Japan (post code 862)

(abstract)

There are a lot of data expressed in the table-form, for example corporate management data et al. The table-form is convenient for data arrangement, but it is not easy to understand the meaning of data, especially in case of large amount of data. Therefore, the summary or graph presentation are designed to find the important information. In this article, the difference of information is studied when the expression of the table-formed data is changed. About the case that the three-item data table is decomposed into the three kinds of two-item data table, the information change can be measured by using the new idea 'system information'. This concept is expected to be useful for data mining.

1．はじめに

最近の表計算ソフトの高機能化及びグループウエア・ソフトの普及は、企業等の職場において表形式データのパソコン処理を益々盛んにさせている。インターネットの普及は，この傾向に拍車をかけこそすれ，決してそれを不要にするものではない。したがって，表形式データの重要性は企業にあって一層高まるものと思われる。

もともとデータは，表形式にまとめられることによって、整然としたものとなり、見やすくなる。事務処理の第一歩が表形式データの作成であったのも当然と言えよう。その第２段階としてパソコンによる処理が行われるようになり，当初に述べたような風潮がみられるようになった訳である。

ところで筆者は，パソコンによるデータ処理の改善を目指して，いろんな検討[1]を行ってきた。その一環として表形式データにはどんな意味があり，それが有する情報をどのように抽出したらよいのかを研究課題に取り上げている[2]。そのための手法として情報エントロピーに基づく情報量の算出を行っており，この報告においても、その考え方が基礎となっている。ただ表形式データの場合，筆者が独自構想に基づいて提案しているシステム情報量なる概念を用いて，情報量を計測するのがよいと考えている。これは，とくに表形式データの分解にあたって役立つ指標である。

近年の組織体におけるイントラネットやグループウエアの活用のひとつに，表形式データの共有がみられる。このような傾向は，事務の生産性を向上させると期待されているが，本当にそうなのか評価してみる必要がある。勿論，否定するのが目的ではなく，単なる導入で満足せずに，生かすための工夫が必要ではなかろうかということである。

筆者は、コミュニケーションにおける情報伝達の問題[3]も含め、グループウエア時代ににおける有効な情報活用の方策を理論的に検討したいと思っている。

ここでは、先ず重要なデータ表現法のひとつである表形式がもつ情報量の問題を取り上げ，そのための手掛かりとしたい。

2．表形式データの情報量について

最初に，表形式データの情報量をどのように捉えるかについて説明したいと思う。具体例として、衣服販売のチェーン店を展開している企業があるとしよう。この企業の売上実績表は、店名（X）月次（Y）商品名（Z）の３項目から構成されている。チェーン店の数がｍで、商品の種類がｎとすると，この表の内容は以下の集合で表せる。

$$X = \{ X_1, X_2, X_3 \cdots\cdots X_m \} \qquad (1)$$

$$Y = \{ Y_1, Y_2, Y_3 \cdots\cdots Y_{12} \} \qquad (2)$$

$$Z = \{ Z_1, Z_2, Z_3 \cdots\cdots Z_n \} \qquad (3)$$

ここで， X_1, X_2, X_3 等はそれぞれ１号店，２号店，３号店等の売上高であり， Y_1, Y_2, Y_3 等は１月，２月，３月……１２月等の売上高であり，$Z_1 Z_2, Z_3$ 等は婦人服，子供服，下着等の売上高を意味している。したがって、これらの要素項目には売上高の数値が対応して存在している。次に各項目の売上高の和で要素項目値を除した値を小文字で表せば，以下のベクトル表示が得られる。

$$x = [\ x_1 \ x_2 \ x_3 \ \cdots\cdots x_m\ ] \qquad (4)$$

$$y = [\ y_1 \ y_2 \ y_3 \ \cdots\cdots y_{12}\ ] \qquad (5)$$

$$z = [\ z_1 \ z_2 \ z_3 \ \cdots\cdots z_n\ ] \qquad (6)$$

当然ながら次式が成り立っている。

$$x_1+x_2+x_3 \cdots\cdots +x_m = 1$$

$$y_1+y_2+y_3 \cdots\cdots +y_{12} = 1$$

$$z_1+z_2+z_3 \cdots\cdots +z_n = 1$$

この結果ｘ，ｙ，ｚについて情報エントロピーを定義できることとなり，以下の諸式が得られる。

$$H(X) = \sum_{i=1}^{m} x_i \log x_i \qquad (7)$$

$$H(Y) = \sum_{j=1}^{12} y_j \log y_j \qquad (8)$$

$$H(Z) = \sum_{k=1}^{n} z_k \log z_k \qquad (9)$$

そして，行ベクトルx，y，z間の関係を表わすシステム情報量が以下に掲げる式によって与えられる。

$$S(X;Y;Z) = H(X \cdot Y \cdot Z) - H(XYZ) \quad (10)$$

この式で、 $H(X \cdot Y \cdot Z)$ は要素事象の組み合わせ分布

$x_1y_1z_1, x_1y_1z_2, \cdots\cdots x_my_{11}z_{n-1}, x_my_{12}z_n$

（要素数は 12mn 個）

にたいする情報エントロピー値であり、第2項の $H(XYZ)$ は要素事象の同時分布について求めた情報エントロピー値である。すなわち要素事象 x_1 について考慮するyの要素事象を y_{11}, y_{12}, $\cdots\cdots$ y_{112} と置き，要素事象 x_2 について考慮するyの要素事象を y_{21}, $y_{22}\cdots\cdots$ y_{212}, $\cdots\cdots$, 要素事象 x_m について考慮するyの要素事象を y_{m1}, $y_{m2}\cdots\cdots$ y_{m12} と置き、複合要素事象 x_1y_{11} について考慮するzの要素事象を z_{111}, z_{112} $\cdots\cdots$ z_{11n}, $\cdots\cdots$, $\cdots\cdots$，複合要素事象 x_my_{m12} について考慮するzの要素事象を z_{m121}, z_{m122} $\cdots\cdots$ z_{m12n} 等々と置いたとき、xyzの同時分布は，次の項目

$x_1y_{11}z_{111}, x_1y_{11}z_{112}, \cdots\cdots$

$x_my_{m12}z_{m121}, \cdots\cdots x_my_{m12}z_{m12n}$

のように表せ，その要素数は同じく12mn個である。したがって，（１０）式は次のようにも表現できる．

$$S(X;Y;Z) = H(X \times Y \times Z) - H(X \cdot Y_X \cdot Z_{XY})$$

ここで右辺第１項の×は，集合X，Y，Zの直積集合を表し，Y_X はXの要素事象に対応するYの要素事象の集合，Z_{XY} は Y_X の要素事象に対応するZの要素事象の集合を表すようにすれば，それらの積により同時集合を表せる．そしてX，Y，Zは，互いに従属の関係にあるが，その順序は任意であるので，同時集合について以下の関係が成立している．

$$X \cdot Y_X \cdot Z_{XY} = X \cdot Z_X \cdot Y_{XZ} = Y \cdot Z_Y \cdot X_{YZ}$$
$$= Y \cdot X_Y \cdot Z_{YX} = Z \cdot X_Z \cdot Y_{ZX} = Z \cdot Y_Z \cdot X_{ZY}$$

ところで事象X，Y，Z間に何の相互関係も存在しなければ，

XYZの分布＝X・Y・Zの分布

となるので，システム情報量の値は零となる．

また，（１０）式は次のように書き直すことができる[4]。

$$S(X;Y;Z) = \{H(X)+H(Y)+H(Z)\} - H(XYZ) \quad (11)$$

この式の第１項（中括弧の中身）は，事象X，Y，Zが個々に存在しているときの情報エントロピーの総和であり，同時事象というものが考えられないときの最大値である。これに対し，第２項は同時分布における一様性の程度（換言すれば，偏りの無さ）を示している。そこで，（１１）式の右辺を言葉で表現してみると，次のようになる。

「個々の分布の様相の総和－同時分布を考慮するときに生じる分散の程度」

これがシステム情報量の与える意味である。

このことから，対象としている事象群についてシステム情報量の大小を問題とすることは，要素事象間の因果関係を明らかにすることに通ずると言えよう。

さらに，上記の検討では表面に出てこないが，実際にシステム情報量を計算するときにはベイジァン・モデルを下敷きにしているのであり，この事実からもシステム情報量による検討は，実は因果関係の検討と等価であるとみなせる。

したがって，複数の項目から構成されるグループデータについてシステム情報量を求めれば，項目間の因果関係（項目の生起順序を考慮する必要が無いときは相互関係）の目安が得られる。これについては，情報エントロピーに関するベン図を考慮することにより，理解が深まると思われる。議論を簡単にするため，２次元の場合を取り上げると，（１１）式は次のようになる。

$$S(X;Y) = \{H(X)+H(Y)\} - H(XY) \quad (12)$$

この式をベン図に表現したのが次頁の図１である。この図で，同時分布に関する情報エントロピーH（XY）に対応する領域は灰色の部分で示されている．この領域は，同時分布が一様である程（つまり，XとY間の相互関

係の度合いが弱い程）広くなる．反対に，XとY間の相互関係の度合いが強い程狭くなる．

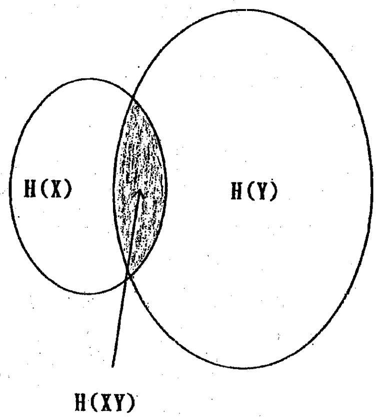

図１　２次元のシステム情報量

したがって，両円内の残りの領域で与えられるシステム情報量の大小と事象間の相互関係の強弱は，同じ傾向を有すると言える。

３．簡単な例の表示変換と情報量

表１に掲示するような，簡単な販売データについて考察する。

表１　或る企業の販売実績

店名	時期	商品名	
		婦人服	子供服
1号店 x_1	上期 y_{11}	z_{111}	z_{112}
	下期 y_{12}	z_{121}	z_{122}
2号店 x_2	上期 y_{21}	z_{211}	z_{212}
	下期 y_{22}	z_{221}	z_{222}

表２　時期と商品名の関係

時期	商品名	
	婦人服	子供服
上期 y_1	z_{11}	z_{12}
下期 y_2	z_{21}	z_{22}

表３　商品名と店名の関係

商品名	店名	
	1号店	2号店
婦人服 z_1	x_{11}	x_{12}
子供服 z_2	x_{21}	x_{22}

この表では２章と同じく，xは店毎の売上高，yは時期毎の売上高，zは商品毎の売上高を意味しているが，各事象の要素数は全て２としており，添字により区分けがしてある．そして、それぞれの合計に対する比をとって次の正規化を行い，

$$x_1 + x_2 = 1$$

$$y_{11} + y_{12} = y_{21} + y_{22} = 1$$

$$z_{111}+z_{112}=z_{121}+z_{122}=z_{211}+z_{212}=z_{221}+z_{222}=1$$

が成立するように数値変換したとする．

また，表１から「yとzとの２項関係」及び「zとxとの２項関係」を知りたい場合がある．（けだし，xとyとの関係は表１から一目瞭然である）．それらは，左側の表２と表３の形で得られる。ここで、表２，３に現れている記号の値は次の諸式で与えられる．

$$y_1 = x_1y_{11} + x_2y_{21},\quad y_2 = x_1y_{12} + x_2y_{22}$$

$$z_1 = x_1y_{11}z_{111} + x_2y_{21}z_{211} + x_1y_{12}z_{121} + x_2y_{22}z_{221}$$

$$z_2 = x_1y_{11}z_{112} + x_2y_{21}z_{212} + x_1y_{12}z_{122} + x_2y_{22}z_{222}$$

$$x_{11}=x_1(y_{11}z_{111}+y_{12}z_{121})/z_1$$

$$x_{12}=x_2(y_{21}z_{211}+y_{22}z_{221})/z_1$$

$$x_{21}=x_1(y_{11}z_{112}+y_{12}z_{122})/z_2$$

$$x_{22}=x_2(y_{21}z_{212}+y_{22}z_{222})/z_2$$

$$z_{11}=(x_1y_{11}z_{111}+x_2y_{21}z_{211})/y_1$$

$$z_{12}=(x_1y_{11}z_{112}+x_2y_{21}z_{212})/y_1$$

$$z_{21}=(x_1y_{12}z_{121}+x_2y_{22}z_{221})/y_2$$

$$z_{22}=(x_1y_{12}z_{122}+x_2y_{22}z_{222})/y_2$$

また，それぞれの場合におけるシステム情報量は次式で定義される．

$$S(x;y;z) = H(x \cdot y \cdot z) - H(xyz) \quad (11)$$

$$S(x;y) = H(x \cdot y) - H(xy) \quad (12)$$

$$S(y;z) = H(y \cdot z) - H(yz) \quad (13)$$

$$S(z;x) = H(z \cdot x) - H(zx) \quad (14)$$

ここには，初めの３次元の値と３種類の２次元の値が与えられているが，後者は前者を分解したものと考えられる．これを「システム情報量の分解」と称している．

また，情報エントロピーの計算に必要な，これらの式における組み合わせ分布は次に示す通りである。

3次元事象については,

$\{x\cdot y\cdot z\} = \{x_1y_1z_1, x_1y_1z_2, x_1y_2z_1, x_1y_2z_2, x_2y_1z_1, x_2y_1z_2, x_2y_2z_1, x_2y_2z_2\}$.

2次元事象については,

$\{x\cdot y\} = \{x_1y_1, x_1y_2, x_2y_1, x_2y_2\}$,

$\{y\cdot z\} = \{y_1z_1, y_1z_2, y_2z_1, y_2z_2\}$,

$\{z\cdot x\} = \{z_1x_1, z_1x_2, z_2x_1, z_2x_2\}$.

さらに,同時分布については以下の通りである.

$\{xyz\} = \{x_1y_{11}z_{111}, x_1y_{11}z_{112}, x_1y_{12}z_{121}, x_1y_{12}z_{122}, x_2y_{21}z_{211}, x_2y_{21}z_{212}, x_2y_{22}z_{221}, x_2y_{22}z_{222}\}$,

$\{xy\} = \{x_1y_{11}, x_1y_{12}, x_2y_{21}, x_2y_{22}\}$,

$\{yz\} = \{y_1z_{11}, y_1z_{12}, y_2z_{21}, y_2z_{22}\}$,

$\{zx\} = \{z_1x_{11}, z_1x_{12}, z_2x_{21}, z_2x_{22}\}$.

ところで,これらの事象間の関係は,図2に掲げるような幾何学的構造によっても表現される。図2の中央に列記したのが上記の同時分布の値である。ここではX,Y,Zの順序で記載しているが,この順序をどのように入れ換えても同時分布の値に変化は生じない.

なお表1の内容は,この図の左側に対応しており,右側は事象Zの要素値が得られる計算過程を示している。その結果については既に記述したとおりである。

この図2の構造のグラフを相補的分枝構造と称している。「相補的」の謂われは,左側で分散してゆく事象が右側では纏められてゆくからである.

そして,X－Y,Y－Z,Z－X間の相補的分枝構造図は,図3,4と図5(次頁)のように与えられる.これらの3種類の2次元構造図と初めの3次元構造図とは数学的に等価であり,前者から後者へ,またその逆でも簡単な演算により導出が可能である[5].しかしながら,情報量は必ずしも同じではない.

つまり,3次元データの場合のシステム情報量 $S(x;y;z)$ と,2次元データの場合のシステム情報量の総和

$$S(x;y)+S(y;z)+S(z;x)$$

との大小比較は次の判別式[6]

$$D = (x_{11}-x_{21})(y_{11}-y_{21})(z_{11}-z_{21}) \qquad (15)$$

の正負により左右される。

もし,Dが正であれば,2次元データのシステム情報量の総和が元の3次元データのそれよりも大となるので,3種類の2次元(2

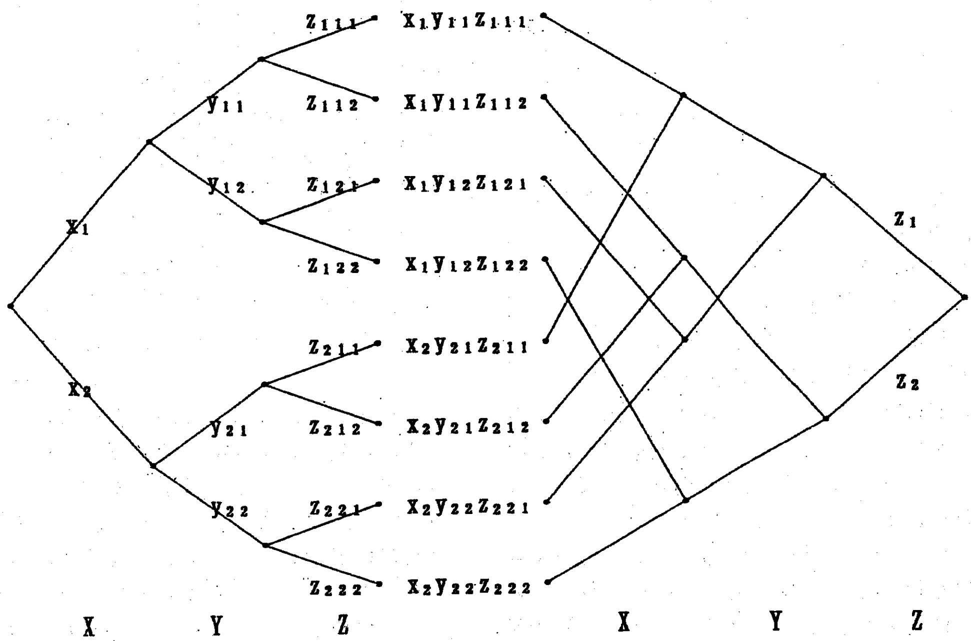

図2　3次元不確定状態系の相補的分枝構造

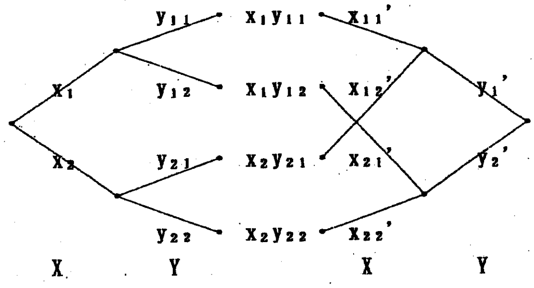

図３　２次元（X・Y）に分解した構造

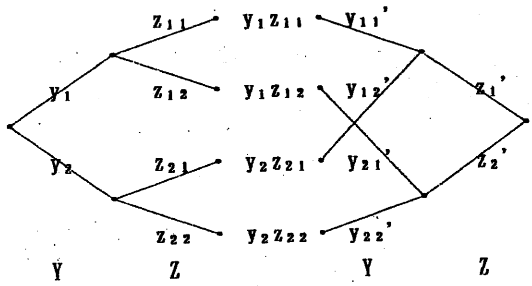

図４　２次元（Y・Z）に分解した構造

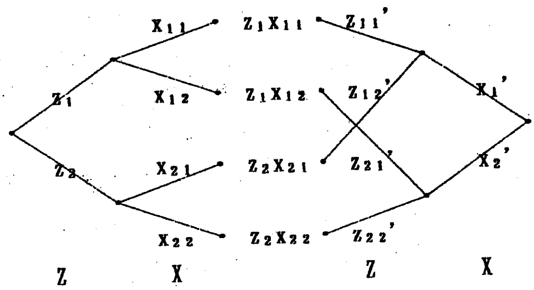

図５　２次元（Z・X）に分解した構造

項目間）データに変換することにより，それまで見えなかったデータ間の関係が明らかになると期待できる．

これに対し，Dが負であれば，２次元データのシステム情報量の総和が元の３次元データのそれよりも大となるので，３種類の２次元（２項目間）データに変換することにより，データ間の相互関係に関する情報が失われる恐れがあるということになる．

このように，データの構造によって表示変換が有利になったり，不利になったりするのは当然考えられることであり，その判断根拠をシステム情報量が提供するのである．

４．数値例による検討

前掲の企業について，具体的な販売高の構成が次の表４のように与えられたとする．最初の表からは，各時期における商品の販売実績や各商品がどの店でどのように売れているかは一見してわかるというわけにはいかないので，このような情報を知りたいときには表５のように編集しなおす必要がある．

表４　或る企業の販売構成比

店名	時期	商品名	
		婦人服	子供服
1号店	上期　0.25	0.167	0.833
0.333	下期　0.75	0.200	0.800
2号店	上期　0.60	0.125	0.875
0.667	下期　0.40	0.875	0.125

表５　分解して得られる構成表

店名	時期	
	上期	下期
1号店　0.333	0.25	0.75
2号店　0.667	0.60	0.40

時期	商品名	
	婦人服	子供服
上期　0.483	0.133	0.867
下期　0.517	0.548	0.452

商品名	店名	
	1号店	2号店
婦人服 0.347	0.184	0.816
子供服 0.653	0.413	0.587

確かに表５によれば、２項目間の関係はよくわかるようになる．しかしながら，表４の情報量と表５のように２項目関係に還元したものが提供する情報量は果して同一であるのかという問題が生じる．これを評価する指標が前記に提案したシステム情報量である．

そこで，３次元（表４の場合）及び２次元（表５の場合）に分解したシステムについてその値を求めると，

$S(x;y;z) = 0.413$,　$S(x;y) = 0.082$,
$S(y;z) = 0.078$,　$S(z;x) = 0.040$.
したがって，

$$S(x;y)+S(y;z)+S(z;x) = 0.200$$

となるので，

$$S(x;y;z) > S(x;y)+S(y;z)+S(z;x)$$

この比較の結果が示唆することは，元の表を3種類の2項目間関係表に分解したことで，情報量に欠損が生じたのではなかろうか，ということである．

事実，表4と表5とを比較したとき，2号店上期の婦人服の売上に問題があることが一見して明らかなのに対し，表5では隠れてしまっている（表5からは2号店の婦人服の売上高構成は約82％とみられるので、寧ろ問題ないように判断される）ことが指摘できるのではなかろうか．

次に、システム情報量の増減の問題を検討するために、表4の内容を表6（右側）に示すように一部変更してみる。しかし、その問題点はもとのまま残されている．

これを2項目間の関係表に変換した（表5に対応する）ものがその下に掲げるところの表7である。このときのシステム情報量を求めてみると，

$S(x;y;z) = 0.238$,　$S(x;y) = 0.082$,
$S(y;z) = 0.155$,　$S(z;x) = 0.022$,
等となる。したがって，

$$S(x;y)+S(y;z)+S(z;x) = 0.259$$

となるので，

$$S(x;y;z) < S(x;y)+S(y;z)+S(z;x)$$

であるが、両者の差は僅かである．

つまり，この場合は情報量の喪失は起こらないと判断される。事実、前例と比較して言えることは、表7の下段から婦人服の売上高に問題がありそうなことであり、上段からそれはとくに上期の売上に原因がある点が指摘されることであろう。

情報量がやや増えるという点は、次の3点に集約されていると思われる。

（1）店や時期に関係なく、婦人服の売上に少し問題があることが表7から読み取れるということ．

（2）それが2号店の販売に起因することが表7より一目瞭然であること．

（3）子供服は、全体として問題ないが、1号店の子供服の売上が少ないことが明示されていること。

表6　或る企業の販売構成比

店名	時期	婦人服	子供服
1号店 0.333	上期 0.25	0.167	0.833
	下期 0.75	0.571	0.429
2号店 0.667	上期 0.60	0.125	0.875
	下期 0.40	0.556	0.444

表7　分解して得られる構成表

店名	時期 上期	時期 下期
1号店 0.333	0.25	0.75
2号店 0.667	0.60	0.40

時期	商品名 婦人服	商品名 子供服
上期 0.483	0.133	0.867
下期 0.517	0.563	0.437

商品名	店名 1号店	店名 2号店
婦人服 0.355	0.442	0.558
子供服 0.645	0.273	0.727

5．考察

上の例ではデータ数が少ないため，表示変換の影響をシステム情報量により検討することのメリットは，それほど感じられないが，データ量が膨大になれば問題のある箇所の発見が困難になるので，ここで述べたような分析法が役立つ局面が出てくると予想される．そのときの方法として，次にあげるような手順が考えられる。

（1）データを幾つかのグループに分割して，それぞれのシステム情報量を求め，グ

ループ間の比較を行う．データの性質を考慮しながら，分割のやり方を変えて，同じ検討を行う．これにより，個々のデータ間の結びつきの強いグループが検出される．このときのシステム情報量をグループ情報量という．

（２）データの次元数を減らしたとき，システム情報量が増減する様子を調べ，データ間アソシエーションのレベルを知る．例えば，表４と表５とを比べたとき表４の情報量が大であることは，このデータのアソシエーション・レベルが３（次元）であることを意味している．

（３）膨大な量のマトリックス・データ（ベース）については，種々の縦横の分割によるグループ情報量と幾通りかのアソシエーション・レベルを調査することによって，重要な情報が含まれているデータ組を検出する．

６．おわりに

コンピュータによるデータ処理の有用性に疑問の余地はないが，パソコンの普及によりデータ処理の大衆化が進展している今日，単なる経験則だけによってそれを行っていればよいとは思われない．そこで，筆者は「データ処理の意義」についてささやかな研究を実施してきたのであり、今回は表形式データの処理にかかわる評価法を提案し，諸賢による批判を仰ぐことにしたものである．

すなわち，表形式のデータが３次元２・２・２要素事象から成る場合について，とくに表構成に関し分解前と後の関係を論じ，かつシステム情報量なる指標を提案することによって，その情報量がどう変化するかを検討してみた．この手法の骨子は次の２点にある．

（１）データ構成内容を一元的に示すシステム情報量が表形式データの表示情報量となり得ること．

（２）分解前後のシステム情報量の比較方式が理論的に正しいこと．

これらを事例に沿い，検討している．

つまり，簡単な事例ではあるが、チェーン店の売上高の表示法に関する具体的数値をあげて，表示変換の影響を述べた。とくに、情報量の増減がどのような形で生じているかを考察してみた。本論文に掲げた例では、その判断にかなり微妙なものがあるが、一応納得できる根拠を提示できたと考えている．

今後の課題として，要素数が多くなるときの一般的解析及び具体例の検討を進める必要があろう。さらに、巨大な表データの中に隠された重要な知識を発掘する問題との共通性に鑑みて，Ｄａｔａ　Ｍｉｎｉｎｇ 技法[7]との関連を検討するという立場にもたってみる必要があろう。

－参考文献－

１）古閑：データ処理の意味の検討―情報エントロピーによる評価法の提案―パソコンリテラシー，Vol.22，No.2，pp.38-44(1997/2)

２）古閑他：表形式データの表示変換に関する情報論的検討，平成８年度電気関係学会九州連合大会講演論文集，p.675（1996/10）

３）古閑：情報論による組織内コミュニケーションの研究（コミュニケーション・パターンの基礎的検討），第１１回熊本県産学官技術交流会資料集，NO.11408（1997/1）

４）古閑：不確定状態系の基礎理論，情報処理学会九州支部研究会報告　pp.212-221（1997/3）

５）古閑：状態変換システムの諸特性の検討（その２），情報処理学会研究会報告 Vol.96，No.91，pp.17-24（1996/9）

６）古閑：状態変換システムのシステム情報量に関する検討，情報処理学会第５３回全国大会講演論文集，pp.1-155~1-156（1996/9）

７）M-S. Chen, J. Han, P. S. Yu: Data Mining: An Overview from a Database Perspective, IEEE Trans. Knowledge and Data Eng. pp.866-883（Dec.1996）