情　報　学　基　礎　31－1
（1993．　9．　14）

秤量問題等における情報エントロピー及び分枝構造の意義

古閑　政
九州東海大学　工学部

初めに通報によって事象の生起確率が変化する時、どのように情報量を求めるのが妥当であるかを検討し、相互情報量を基礎とする考え方の利点を説明した。

そして、情報量が秤量問題に於けるヒューリスティクスとなる点について、秤量の各段階に於ける出現確率値を基準に、分枝構造を組み立てることによって、秤量計画が得られることを示した。さらに、事前確率分布が予想される事象にかかわる問題解決の手法に関しては、分枝構造を検討することが有効なアプローチとなるだろう、と主張した。

また、分枝構造を相補的にすることにより、相互情報量が簡単に求まり且つ情報エントロピーの理解にも役立つことを述べた。

SOME STUDY ABOUT INFORMATION-ENTROPY IN THE BAD COIN FINDING PROBLEM , ETC

Masashi Koga

Kyushutokai University
Department of Technology
9 -1 -1 , Toroku , Kumamoto City , 862 Japan

The method to measure information quantity is discussed when a message makes the occurrence probability of events change. In reference to this, it is emphasized that the concept of mutual information should be basic. While the information helps for heuristics about the bad coin finding problem, the plan of balancing is got by framing branching from the viewpoint of appearence probability at each step. If the priori probability distribution of the concerned problem is known, the study of branching gives good hints about the problem.

The complementary branching structure shows how to compute mutual information and is useful to understand information entropy in various cases.

1．はじめに

１９４８年シャノンが情報理論を提唱して以来、それは通信理論の基礎として、高い評価を獲得してきたが、計算機科学の中で情報理論は通信路符号化理論と等価視され、所謂情報処理と結びつくことは殆ど無かった。僅かに、情報量の単位であるビットが、ワード或いはデータの単位として或る種の共通概念をもって使われたに過ぎなかった。即ち情報理論は、通信系では生かされたが、情報処理系では生かされなかったと言えよう。

他方、１９６５年ザデーがファジィ集合の概念を提案し、その後ファジィ理論とファジィ制御が目覚ましい発達を遂げた。最近では、ファジィ情報処理[1]と言う分野も開けつつある。こうした動向のひとつに、あいまいさを表現する測度としてファジィ情報エントロピーが取り上げられている[2]。その応用分野は今後の課題であるが、おそらくファジィ推論の過程におけるヒューリスティクスの一種か、或いは評価値として役立つものと思われる。更に、ファジィデータを処理するシステムの特性値に相応しい面がある。

このような意義を有するものとして、最初に情報エントロピーが取り上げられたのは、秤量問題においてであった[3]。そこで、この問題を再び取り上げ、解決の手掛かりを獲得する過程を考察する。特に分枝構造が重要な役割を果たすことから、その性質を検討したところ、相補的分枝構造により相互情報量が簡単に求まることも分かった。

これらは、方法論としての有意性もさることながら、情報量の意味をグラフを通じて理解できる利点がある、と思われるので報告する。

2．情報の測定について

2．1情報量の直接定義

今その成り行きに関心を払っている、或る事象があり、それが生起して、どういう結果になったかを教える通報が伝達されたとする。ゴールドマンによる[4]と、この通報によって受信した情報量は、次式で定義される（これを直接情報量と呼ぶことにする）。

$$\text{通報から得られた情報量} = \log_2 \frac{\text{通報受信後の事象の受信側の確率}}{\text{通報受信前の事象の受信側の確率}} \qquad (1)$$

この式の意味を考えるために、具体例としてしばしば引用される「病院にいる妻の出産の報せを待ち詫びる夫の話」を取り上げよう。

夫は「男の子」を望んでいるとして、その事前確率は０．５である。そこへ「男の子」が生まれたという通報が、職場にいる夫に電話で伝えられたとすれば、（１）式で計算される情報量は１ビットである。又、この時の通報システムを、１（男の子に対応）、０（女の子に対応）によって符号化しておいたとすれば、伝送システムとしては最も簡単な符号通信系により、完全な情報伝送がおこなわれるのであるから、１ビットが最も単純な情報伝達系の運ぶ情報量と見なせる。換言すれば、１ビットの情報伝達系が所謂<u>情報原器</u>とみてよいであろう。（距離にメートル原器、重さにキログラム原器があるように、情報原器なる概念を導入して、情報量の基準値を与えるべきだと思う。）

しかしながら、この説明には若干の問題がある。もし「男の子」を望んでいる夫のところに「女の子」が生まれた、という通報が届いたとする。期待した事象の事後確率は０となる為、（１）式から情報量は－∞となる。これは、夫の失望の度合いを表すにはふさわしいかもしれないが、情報量としては変である。情報の意味を考える立場からは、反対方向の結論が得られたのであるから－１ビットなる答えが得られるべきであろう。しかし、（１）式をどのように解釈しようとも、そのような結論を導くことは出来ない。

従って、事前には「女の子」が生まれる可能性も０．５の確率であり得たのであるから

それを知らせた情報量は、最初の場合と同様１ビットである、と説明することになる。これは、やや苦しい言い逃れであろう。

情報について、万人が抱いているイメージからすれば、（１）式は直観的にぴったりする算式であるが、適用は事後確率が１になる場合に限るほうが無難である。そこで、（２）式が得られる。

通報により事象が確定した時の情報量＝－ｌｏｇ₂(通報受信前の事象の受信側の確率) ……（２）

しかし、この式には更に問題があり、不完全な情報や誤った情報（つまり雑音の影響がある場合）によって、事後確率が１に満たない通報系の情報量を計算できない。

２．２情報エントロピーによる定義

シャノンが提案した情報エントロピーは、次のように定義される。

事象の集合を $\{A_1, A_2, A_3 \cdots A_n\}$ とし、それぞれの出現確率を $\{P_1, P_2, P_3 \cdots P_n\}$ としたとき事象系Ｘの表現と情報エントロピーを次式で与える。

$$X = \begin{Bmatrix} A_1 & A_2 & A_3 \cdots\cdots & A_n \\ P_1 & P_2 & P_3 \cdots\cdots & P_n \end{Bmatrix} \quad (3) \qquad H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i \quad (4)$$

（１）相互情報量について

（３）式で表される事象系に、他の有限離散事象系（次式）

$$Y = \begin{Bmatrix} B_1 & B_2 & B_3 \cdots\cdots & B_n \\ Q_1 & Q_2 & Q_3 \cdots\cdots & Q_n \end{Bmatrix} \quad (5) \qquad H(Y) = -\sum_{i=1}^{n} Q_i \log_2 Q_i \quad (6)$$

が関連している複合事象系Ｘ・Ｙを考える。

この場合、相互情報量は次式により与えられる。

$$I(X;Y) = H(X) - H(X \mid Y) \quad (7)$$

ここでＨ（Ｘ｜Ｙ）は、事象系Ｙが生起した時のＸに関する条件付平均エントロピー。

上式の意味を考察するため、前述の「赤ちゃん誕生」に於ける情報量の問題を再考する。

事象系Ｘ・Ｙは、それぞれ２個の事象しかないので（Ｍは男子、Ｆは女子を示す）

$$X = \begin{Bmatrix} M & F \\ P_1 & P_2 \end{Bmatrix} \quad (8) \qquad Y = \begin{Bmatrix} M' & F' \\ Q_1 & Q_2 \end{Bmatrix} \quad (9)$$

Ｘは通報前の事象、Ｙは通報後の事象を表しているとすれば、「男の子」が生まれたという通報があった時（８）、（９）式は次のようになる。

$$X = \begin{Bmatrix} M & F \\ 0.5 & 0.5 \end{Bmatrix} \quad (10) \qquad Y = \begin{Bmatrix} M' & F' \\ 1 & 0 \end{Bmatrix} \quad (11)$$

上の（１０）式からＨ（Ｘ）＝-0.5log0.5-0.5log0.5 ＝１ビットが得られる。

条件付平均エントロピーは、Q_1、Q_2 を通報後のＸに関する周辺確率とすると

$$H(X \mid Y) = Q_1 H(X \mid M') + Q_2 H(X \mid F') \quad (12)$$

更に、Ｈ（Ｘ｜Ｍ'）とＨ（Ｘ｜Ｆ'）については

$$H(X \mid M') = -P(M \mid M')\log_2 P(M \mid M') - P(F \mid M')\log_2 P(F \mid M') \quad (13)$$

$$H(X \mid F') = -P(M \mid F')\log_2 P(M \mid F') - P(F \mid F')\log_2 P(F \mid F') \quad (14)$$

ここで、 $P(M \mid M')$ $P(F \mid M')$ $P(M \mid F')$ $P(F \mid F')$ はベイズの定理を使って導かれる。これらの

確率値１、０、０、１を順次上式に代入すると

$$H(X|Y)=-0.5(1*\log_2 1+0*\log_2 0)-0.5(0*\log_2 0+1*\log_2 1)=0$$

となる。但し $0*\log_2 0=0$ としている。従って（７）式の相互情報量Ｉ（Ｘ；Ｙ）は１ビットとなり、（１）式から得られる値と同じになる。これは、（１１）式が（１，０）ではなく（０，１）となっても同様である。

情報量の意味は考えないとすれば、通報によって確率事象の変化が起こる場合の情報量は、相互情報量によって統一的に与えられることが分かる。

（２）平均情報エントロピーの変化

これまでの考え方に対して、事後確率分布（男の子）｛１，０｝や（女の子）｛１，０｝を事前確率分布とは全く独立なものとして、その平均情報エントロピーを計算する方法がある。この値は、言うまでもなく０ビットであり、事前確率分布の平均情報エントロピーは１ビットであるから、両者の差は１ビットである。そして、これが通報によって得られた情報量となる。これは相互情報量と全く同じ値である。

前述したとおり、相互情報量の導出過程はかなり複雑であるが、ここに述べた手法は極めて簡単である。ただ両者が一致するのは、Ｈ（Ｘ・Ｙ）＝Ｈ（Ｘ）＋Ｈ（Ｙ）が成立する時だけである。しかしこの制約は、通報による情報量の問題を考える事例ではそれほど強いものではない。

以上の考え方の論拠は、後述する情報エントロピーの分枝構造の相補性により明らかにされる。

２．３通報に誤りがある場合

雑音が通報系の正確な伝達を邪魔し、受信側における事後確率分布が｛１－ｅ，ｅ｝或いはその逆｛ｅ，１－ｅ｝になったとする。この時の相互情報量を求めると[5)]

$$I(X;Y)=1+e*\log_2 e+(1-e)\log_2(1-e) \qquad (15)$$

この結果は、平均情報エントロピーの変化として求めても全く同じである。

ｅ＝0.1 とすると、（１）式の値が 0.85 であり、（１５）式で計算される値が 0.53 となる。前者が後者より大きいのは、雑音により負の情報が発生した分を考慮していないためと思われる。

３．情報エントロピーの分枝構造

３．１分枝分解性

平均情報エントロピーを与える次式は分枝構造のグラフと完全な対応性を有する[6)]。

$$H(P_1,P_2,P_3 \cdots\cdots P_n)=-\sum_{i=1}^{n} P_i \log_2 P_i \qquad (16)$$

換言すれば、（１６）式がもつ性質を分枝構造によって表現することが出来る。

又この分枝構造の性質を数式として表現する次の公式がある。

$$H_{nm}(P_1Q_1,P_1Q_2,\cdots\cdots P_1Q_m,P_2Q_1,P_2Q_2,\cdots\cdots P_2Q_m,\cdots\cdots$$
$$\cdots\cdots P_nQ_1,P_nQ_2,\cdots\cdots P_nQ_m)$$
$$=H_n(P_1,P_2,P_3\cdots\cdots P_n)+H_m(Q_1,Q_2,Q_3\cdots\cdots Q_n) \qquad (17)$$

$$H_n(P_1,P_2,P_3\cdots\cdots P_n)=H_{n-1}(P_1,P_2,P_3\cdots\cdots P_n)$$
$$+H_2(P_1/(P_1+P_2),\ P_2/(P_1+P_2)) \qquad (18)$$

$$H(m/l,m/l,n/l)=H(2m/l,n/l)+2m/l*H(1/2,1/2) \qquad (19)$$

（１８）、（１９）式は簡単な場合を表しているが、２個以上の要素をまとめる公式として一般化できる。なお、これらの式は（１６）式から導出することが出来、分枝構造のグ

ラフから直観的に理解することも出来る。

即ち、 $H(P_1, P_2, P_3 \cdots\cdots P_n)$ が木構造の分枝分解性をもつのは明らかであり、数式としての性質は加法性（(17)式）と漸化性（(18),(19) 式）で十分に表現されている。

3．２秤量問題と分枝構造

有名なクイズに、贋硬貨の問題がある。それは「１２個の硬貨があって、その中に１個だけ重さの異なる（軽重は不明）ものがある。天秤を使い、出来るだけ少ない秤量回数で異常な硬貨を見つけよ」という課題である。

これを解くのに、情報エントロピーが有効なヒューリスティクス情報となることが報告されている[3]。その要点を述べると

①予想される状態は１２＊２＝２４通りある。

②２４状態の中の１状態を確定するのに必要な情報量は、log24=4.585 ビットである。

③天秤による測定１回で（ⅰ）右側に下がる（ⅱ）左側に下がる（ⅲ）平衡する、の３通りの状態のどれかであることが分かるから、log3=1.585ビットの情報量が得られる。

④従って②と③からlog24/log3=2.89 回の測定が必要となる。整数でなければならないから、３回の秤量で見つかるだろう、と予想される。

⑤天秤の能力を十分活用する為、１回の秤量でなるべく３通りの等価な状態が得られるようにする。

⑥あらゆる場合に３回の秤量が必要とすれば、情報量に剰余が出るので、③の２状態しか使わない秤量があり得る。

---

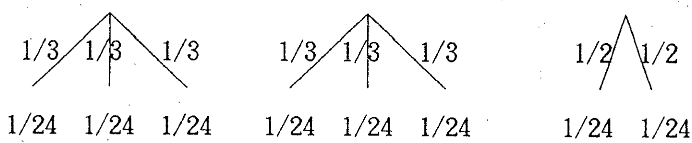
①から⑥を念頭に置いて、以下のステップにより分枝構造を組み立てていく。

---

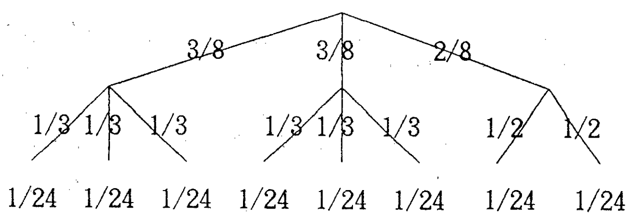
⑦葉にあたる終端確率分布は1/24が２４個並んでいる。

⑧これを３個のグループに分けると、８個ずつになる。

⑨その８個は、下記の分枝をとらざるを得ない。

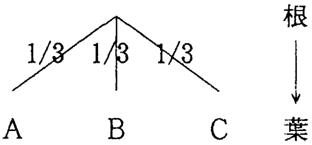


⑩これら３個の分枝を集めると、図－１のようになる。



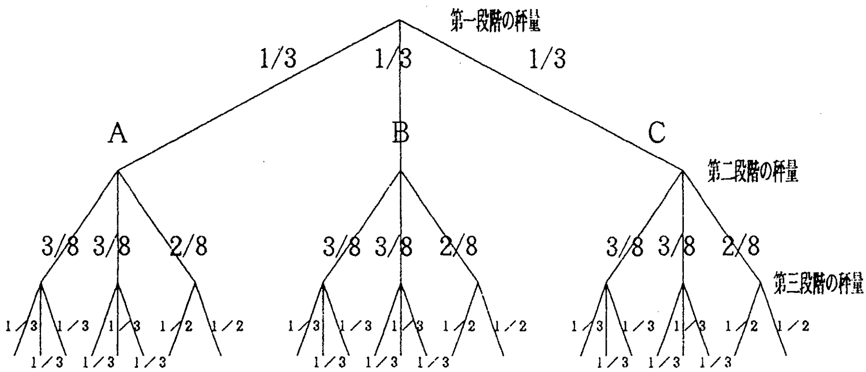
図－１秤量問題に於ける下部分枝構造

⑪根のところを考えると、図－２となる。



図－２出発点に於ける分枝構造

⑫図－２のA、B、Cの下に、図－１を接続すれば、全体構造が完成する。つまりA、B Cは、⑧の３グループに該当するものである。図－３にそれを示す。



図－３贋硬貨を発見する為のステップを示す分枝構造図

図－３の分枝構造が示す手順に沿って、秤量計画をたててゆけばよい。

＜秤量のステップ＞

（ｉ）１２個の硬貨を４個ずつの３グループに分け、２グループを天秤にかける。

（ii）平衡しない場合、その８個の中にあることが分かるから、残りの４個は正常である。これを１個借りて９個とし、又３グループに分ける。例えば軽い群を①^L ②^L ③^L ④^L 、重い群を⑤^H ⑥^H ⑦^H ⑧^H 、正常群を⑨* ⑩* ⑪* ⑫* とする。そして、次の３群とする。

グループⅠ：①^L ⑤^H ⑫*　グループⅡ：②^L ③^L ⑥^H　グループⅢ：④^L ⑦^H ⑧^H

（iii）ⅠとⅡを秤量する。（ａ）Ⅰが軽いとすると、①^L か⑥^H である。そこで①^L と⑫* を比較すれば、どちらであるかが分かる。（ｂ）Ⅰが重いとすると、⑤^H か {②^L か③^L } である。そこで②^L と③^L を比較すれば、結果が得られる。平衡すれば⑤^H が解である。（ｃ）平衡したら、④^L ⑦^H ⑧^H の中に解がある。そこで⑦^H と⑧^H を比較すれば、結果が得られる。平衡すれば④^L が解である。

上例と逆の①^H ②^H ③^H ④^H か⑤^L ⑥^L ⑦^L ⑧^L の場合もほぼ同じ手順で結果が出る。

（iv）初め平衡した場合⑨⑩⑪⑫の中に解があることが分かる。正常な①* を借りて次の３グループ　Ⅰ：⑨ ⑩　Ⅱ：⑪ ①*　Ⅲ：⑫　に分ける。（ａ）Ⅰが軽いとすると⑨^L ⑩^L か⑪^H であるから、⑨^L と⑩^L を比較すれば解が得られる。平衡すれば⑪^H が解である。（ｂ）Ⅰが重いとすると、⑨^H ⑩^H か⑪^L であるから、⑨^H と⑩^H を比較すれば解が得られる。平衡すれば⑪^L が解である。（ｃ）平衡したら⑫と①* を比較する。⑫^L か⑫^H の解が得られる。

このように、分枝構造に対応して秤量計画が得られることが分かる。つまり、事象の出

現確率が予め知られており（秤量問題では1/24）、区分けの手段（秤量では３状態に分ける）が与えられた問題については、先ず分枝構造を考えてみることが問題解決の手順を与える。例えば、或る機械の故障原因の統計的分布が過去のデータから確率値として分かっている時、故障時の点検方法を能率よく組み立てるのに役立つであろう。

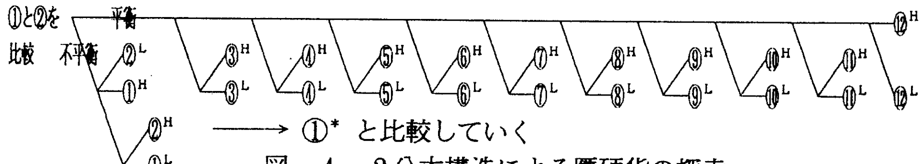
３．３分枝構造と情報量の計算

図－３に関する情報エントロピーは、漸化性から次のように計算される。

$$H(1/24, 1/24, \cdots\cdots 1/24)= H(1/3, 1/3, 1/3)+3*1/3*H(3/8, 3/8, 2/8)+ 3* \{1/3*3/8 *H(1/3, 1/3, 1/3)+1/3*3/8*H(1/3, 1/3, 1/3)+1/3*2/8*H(1/2, 1/2)\} \qquad (20)$$

上式は、平均情報エントロピーが第一段階の秤量計画のエントロピー、第二段階の秤量計画のエントロピーと第三段階の秤量計画のエントロピーの和になっていることを示す。それぞれの値を求めると、1.58、1.56、1.44ビットとなり、合計は4.58ビットである。

次に、秤量の各段階における確率の変化からも情報量の計算が出来ることを説明する。初めに贋硬貨を１２個から選び出し、その軽重をあてる可能性（探索確率）は、一様性を仮定して1/24である。１回目の秤量により、４個ずつのグループのどれかに贋硬貨があるのが分かるから、探索確率は1/8 となる。２回目の秤量の結果は幾通りかに分かれるが、（ⅲ）－（ａ）の場合について考えると、この結果①^{L} か⑥^{H} のどちらかが分かるから探索確率は1/2 となる。即ちこの探索過程での確率変化は1/24→1/8 →1/2 →1 を辿る。従って（１）式を３回使って1.58、２、１ ビットが得られ、計4.58ビットの情報量である。他の探索過程についても同様であるので省略する。なお、図－４の分枝構造に対応した秤量計画でも、解に至るまでの平均情報エントロピーは変わらない。このことから情報エン

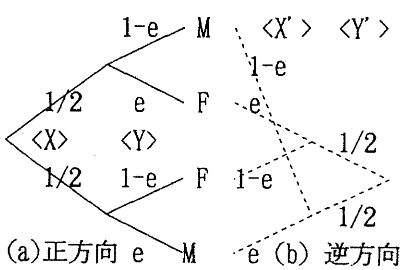


図－４　２分木構造による贋硬貨の探索

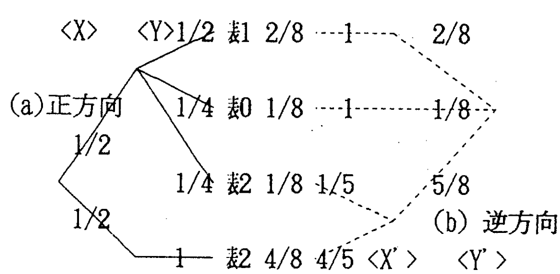
トロピーだけでは問題解決法のヒューリスティクスとは成り得ず、適切な分枝構造を作ることがポイントであることが分かる。

４．相補的分枝構造

最初に例としてあげた「赤ちゃん誕生」の問題を再考する。２．３の通報に誤りがある場合の分枝構造は、図－５（ａ）のようになる。この図に点線で付加した、逆方向の分枝



図－５誕生問題の分枝構造



図－６やや複雑な事象の分枝構造

構造をもつものを相補的分枝構造という。平均情報エントロピーについて、（ａ）で(21)式、（ｂ）で(22)式が成立する。

$$H(X,Y)=H(X)+H(Y|X) \quad (21) \qquad H(X,Y)=H(Y)+H(X|Y) \quad (22)$$

両式を等置することにより、次式が得られる。左側が正方向の相互情報量で、右側が逆

$$H(X)-H(X|Y)=H(Y)-H(Y|X) \quad (23)$$

方向の相互情報量である。従って(15)式は、図－５の相補的分枝構造を利用して直ちに得られる。前述した複雑な計算は必要としない。

別のやや複雑な事象の場合を検討する。「貨幣が２枚あり、ひとつは両面とも表であり他方は表裏があるとする。どちらかを取り出すという試行をＸとし、取り出した貨幣を２回投げるという試行をＹとする[7]」。この複合事象を硬貨の表が何回でるか、という確率で考えた相補的分枝構造は図－６に示される。この時の相互情報量は、同図から直ちに

$$I(X;Y)=H(1/2,1/2)-5/8*H(1/8,4/8) \quad (24)$$

と書き下ろされる。ここで H(1)=0 である為、簡単になっている。

相互情報量を求めるのに、いずれの類書もベイズの定理を使っているが、ここに説明したごとく相補的分枝構造図を用いれば、簡単に求められる。

５．おわりに

秤量問題について考察する傍ら、情報量の定義について根本的に検討してみる必要を感じた。そこで、直接情報量は直観に訴えるものがあるが、限定された範囲でしか(1) 式が使えないことが分かったので、相互情報量との比較検討を試みた。

情報処理系の評価値の性質として、絶対値そのものの意義を見出せなくとも、系の変化や異なる系の評価に有用でなければならない。この立場から相互情報量が直接情報量より勝ると考える。ただ物差しとしての有効範囲の広さは重要であるが、分枝構造としてグラフ化できる点と相補的分枝構造図によって容易に計算できる点に着目し、今後その範囲を広げる必要があろう。

また、秤量問題の解決に際して、分枝構造の検討が秤量計画のヒントとなることを示唆した。この手法は、もっと一般的な問題解決のプロセスを探すのにも役立つ可能性があると主張したい。

なお、本論文では触れる余地が無かったが、ファジィ情報エントロピーの概念をファジィシステムの最適化に利用する為の理論と手法も今後開拓すべき分野であろう。

最後に、ここで述べた手法はグラフに基づいて理解し易くしたものであるから、情報量の学習に役立つ、と思われる。

参考文献

1)情報処理学会編：ファジィ情報処理とその応用（小特集）、Vol.30,No.8(1989/8)
2)本田訳、Klir&Folger:ファジィ情報学、日刊工業新聞（1993/1）
3)P.J.&D.J.Kellogg:Entropy of Information and the Odd Ball Problem,J.App.Phy., Vol.25,pp.1438-1439(1954)
4)関訳、ゴールドマン：情報理論、近代科学社(1956/9)
5)有本：現代情報理論、電子通信学会刊、P.18(1978)
6)堀部：情報エントロピー論、森北出版(1989/10)
7)西田、竹田：ファジィ集合とその応用、森北出版(1978/11)