自由モノイドから生成される粒状階層構造から見た有限マルチ集合・部分集合

Finite Multisets and Subsets from a Point View of Granular Hierarchical Structure Generated from a Free Momoid.

村井哲也（Tetsuya MURAI）
Tel & Fax: 0123-27-6049　　E-mail: t-murai@photon.chitose.ac.jp

Abstract： In this report, we make some consideration on the concepts of finite multisets and finite subsets based on free monoids. For the purpose, we reformulate finite subsets and multisets as some quotient sets induced from free monoids as the set of finite sequences, i.e, strings based on equivalence relations. As a result we have a commutative diagram we call granular hierarchical structure.

例えば「集合{3, 3}」と書くと，誤りであろうか？この問いは次のいずれを採るかを迫る:

(1) 実際に要素の重複があろうがなかろうか，気にしない(無視する)から正しい．

(2) 集合には，要素の重複はあってはならないから誤りである．

一般には，公理的集合論の外延性公理に基づいて，前者の解釈が採られていると考えられる．有限集合の外延的記法$\{x_0, \cdots, x_{n-1}\}$に対して，メタ変数 x_k ($k = 0,\cdots, n-1\in\mathbb{N}$)は異なる要素を表す，という注意書きは(少なくとも筆者は)見たことがない．ここで，自然数全体の集合は 0 を含むとする：□ ={0, 1, 2, …}．以上から，集合の記述として重複は許容するが，しかし，無視する，という見方が妥当と思われる．しかし，後者の立場もある．例えば，Codd の関係データベースの理論(cf.[4])において，タプルの重複が許されない，という説明がある．一方，SQL(cf.[1])では重複が許されるべきであり，マルチ集合(cf.[8])で理論化するのが本来の姿である，という議論が展開される．以上のように，最初の問いは微妙な性格を孕んでいる．本報告では，自由モノイド(cf.[8])の立場から，列の概念(cf.[8])に基づいて，粒状階層構造(cf.[5])を図１に示す可換ダイアグラムとして構成し，有限マルチ集合と有限部分集合の表現について考察する．図中の記号の意味は以下で説明される．

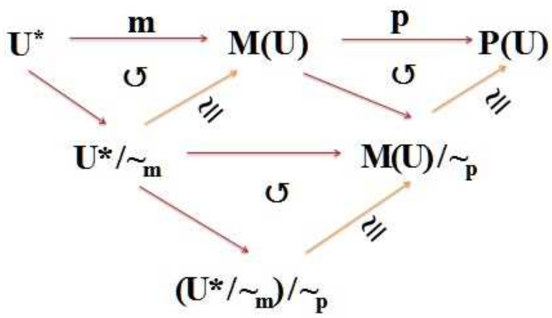

図１．粒状階層構造

まず，基本概念を準備する．U を記号の有限集合とする．U のベキ集合

$$P(U)=\{A \mid A\subseteq U\}$$

は集合演算∪ (合併)と∩ (共通部分), ^{C} (補集合)に空集合∅と全体集合Uを加えて，ブール代数$\langle P(U), \cup, \cap, {}^C, \emptyset, U\rangle$をなす．空集合∅と全体集合 U は包含関係⊆を順序と考えた時の，それぞれ，最小元と最大元である．べき集合の各要素 A∈P(U)に対して，特性関数$\chi_A : U\rightarrow 2$ (ここで，2＝{0,1})を $\chi_A(x)=1\Leftrightarrow x\in A$ によって定義でき，以下，部分集合とその特性関数を同一視する：$P(U) \equiv \{A \mid A : U\rightarrow 2\}$.

次に，集合 U 上のマルチ集合とはカウント関数 $\mu : U\rightarrow \mathbb{N}$ として定義される．要素が何回

出現するかを考慮する集合概念である．集合 $\mathrm{supp}(\mu)=\{x\in U \mid \mu(x) > 0\}$を μ の台と呼ぶ．マルチ集合 μ, μ'間の相等性と包含関係は $\mu=\mu' \Leftrightarrow \forall x\in U\ (\mu(x)=\mu'(x))$, $\mu\subseteq\mu' \Leftrightarrow \forall x\in U(\mu(x) \le \mu'(x))$ によって定義される．また，$\cup$(合併)および$\cap$(共通部分)に加え，$+$(加法)の演算も要素毎に定義される：$(\mu\cup\mu')(x) = \max(\mu(x),\mu'(x))$, $(\mu\cap\mu')(x) = \min(\mu(x),\mu'(x))$, $(\mu+\mu')(x) = \mu(x)+\mu'(x)$. U 上のマルチ集合全体を M(U)で表す：

$$M(U)=\{\mu \mid \mu : U\to\mathbb{N}\}.$$

構造$\langle M(U), \cup, \cap\rangle$は分配束をなす．マルチ集合 0 を $0(x) = 0\ (\forall x\in U)$で定義すると，加法+に関する単位元になるから，構造$\langle M(U), +, 0\rangle$は可換モノイドになる．よって，両者を併せた$\langle M(U), \cup, \cap, +, 0\rangle$は束順序可換モノイドである．また，$\langle M(U),\subseteq\rangle$は順序集合であり，$0(\in M(U))$はこの順序$\subseteq$に関する最小元である．M(U) は最大元を持たないので，一般に補元演算は定義できない．

有限記号集合 U から生成される自由モノイド U*の要素を文字列，記号列，あるいは単に，列と呼ぶ．列をつなげる連接演算$\bullet$で生成される記号列全体を U*と書く．演算$\bullet$は結合則$(s\bullet(s'\bullet s'')=(s\bullet s')\bullet s''\ (\forall s,s',s''\in U^*))$を満たし，空列 ε は連接演算に関する単位元である$(s\bullet\varepsilon=\varepsilon\bullet s=\varepsilon\ (\forall s\in U^*))$．よって，構造$\langle U^*, \bullet, \varepsilon\rangle$はモノイドをなし，U 上の自由モノイドと呼ばれる．自由性の意味はここでは割愛する．連接演算はしばしば省略される：$st=s\bullet t$．各列 $s(\in U^*)$に関して，各記号 x の s 中の出現回数を与える関数 $ct^* : U\times U^*\to\mathbb{N}$ を仮定する．$ct^*(x,s)$は x が s 中に出現する回数である．列 s の台は次式で定義される：$\mathrm{supp}(s)=\{x\in U \mid ct^*(x,s) > 0\}$．列 $s=x_1x_2\cdots x_n$, $s'=y_1y_2\cdots y_m$ に関して，s が s'の部分列である($s \sqsubseteq s'$と書く)とは，順序保存単射 $\varphi: n\to m$ が存在して，任意の $k\in n$ に対して，$s(k)=s'(\varphi(k))$となる時である．ここで，自然数 n (≥ 1)に対して，$n=\{0, 1, \cdots, n-1\}$である．部分列の関係 $\sqsubseteq$ は U*上の順序であり，$\langle U^*, \sqsubseteq\rangle$ は順序集合である．しかし，任意の列 s, s'に関して，上限と下限は必ずしも存在しないので，一般に，束にはならない．

では本題に入る．まず，写像 $m : U^*\to M(U)$を $m(s)(x)=ct^*(x,s)$で定義すると，明らかに準同型である． U*上の二項関係$\sim_m$を $s\sim_m s' \Leftrightarrow \forall x\in U\ (ct^*(x,s)=ct^*(x,s'))$で定義すると，これは同値関係であり，交換律で一致する列を互いに同値とする．商集合 $U^*/\sim_m$ と M(U)との関係を実際に確かめる．$U^*/\sim_m$ の要素(同値類)を$[s]_M$と書く．任意の$[s]_m\ (\in U^*/\sim_m)$ について，$s'\in[s]_m \Rightarrow \forall x\in U\ (ct^*(x,s')=ct^*(x,s))$ が成り立ち，$ct^*(x,s)$の値は代表元の取り方に依らないので，$[s]_m$に関するカウント関数 $ct_m : U\times(U^*/\sim_m)\to\mathbb{N}$ を $ct_m(x,[s]_m)=ct^*(x,s)$. で定義できる．相等性に関して $[s]_m=[s']_m \Leftrightarrow \forall x\in U(ct_m(x,[s]_m)=ct_m(x,[s']_m))$が成り立つ．$[s]_m$ の台を次式で定義する：$\mathrm{supp}_m([s]_m)=\{x\in U \mid ct_m(x,[s]_m) > 0\}$．同値類 $[s]_m, [s']_m\in U^*/\sim_m$ に対する演算$\cup_m$, $\cap_m$ を以下によって定義する：

(1) $[s]_m\cup_m[s']_m=[t]_m$ ($t\in U^*$は $ct^*(x,t)=\max(ct_m(x,[s]_m),ct_m(x,[s']_m))$でその選び方に依らない),

(2) $[s]_m\cap_m[s']_m=[t]_m$ ($t\in U^*$は $ct^*(x,t)=\min(ct_m(x,[s]_m),ct_m(x,[s']_m))$でその選び方に依らない).

すると，べき等律，交換律，結合律，吸収律，分配律を満たすので，$\cup_m$と$\cap_m$はそれぞれ join と meet であり，構造$\langle U^*/\sim_m,\cup_m,\cap_m\rangle$は分配束である．商集合$U^*/\sim_m$の要素 $[s]_m$, $[s']_m$に対して，列間の連接演算を利用して，同値類間の演算$+_m$ を$[s]_m+_m[s']_m=[s\bullet s']_m$ で定義すると，$ct_m(x,[s]_m+_m[s']_m)=ct_m(x,[s]_m)+ct_m(x,[s']_m)$が成り立つ．よって，$+_m$ は同値類間のカウント関数 ct_mに関する加法であり，交換律と結合律を満たす．演算$+_m$に関するゼロ(単位)元を$0_m=[\varepsilon]_M$で定義でき，実際，$[s]_m+_m0_m=0_m+_m[s]_m=[s]_m$ である．よって，構造$\langle U^*/\sim_m,+_m,0_m\rangle$は可換モノイドである．更に，加法と join および meet との間に，それぞれ分配性が成り立つので，構造$\langle U^*/\sim_m,\cup_m,\cap_m,+_m,0_m\rangle$ は束順序可換モノイドである．前述の join(または meet)から，$U^*/\sim_m$上の順序を$[s]_m\subseteq_m[s']_m \Leftrightarrow [s]_m\cup_m[s']_m=[s']_m$ $(\Leftrightarrow [s]_m\cap_m[s']_m=[s]_m)$によって定義できるので，構造$\langle U^*/\sim_m,\subseteq_m\rangle$は順序集合である．$0_m$ はこの順序の下で最小元である．明らかに$[s]_m\subseteq_m[s']_m \Leftrightarrow \forall x\in U(ct_m(x,[s]_m) \le ct_m(x,[s']_m))$が成り立つ．混乱が生じない限り構造$\langle U^*/\sim_m,\cup_m,\cap_m,+_m,0_m\rangle$を単に $U^*/\sim_m$ と書く．これと M(U)との関係を調べるために，写像 $\psi: M(U)\to U^*/\sim_m$を各 $\mu\in M(U)$に対して，$\psi(\mu)=[t]_m$で定義する．ここで，列 t は任意の $x\in U$ に対して，$ct_m(x, t)=\mu(x)$を満たすように選ばれ，選び方に依存しない．この時，ψ は全単射であり，M(U)

の演算∪および∩，+を保存する：$\psi(\mu\cup\mu')=\psi(\mu)\cup_m\psi(\mu')$，$\psi(\mu\cap\mu')=\psi(\mu)\cap_m\psi(\mu')$，$\psi(\mu+\mu')=\psi(\mu)+_m\psi(\mu')$．つまり，ψ は M(U)から $U^*/\sim_m$ への同型写像である：
【命題】$M(U)\cong U^*/\sim_m$
よって，次を得る：$\forall\mu\in M(U)\,\exists s\in U^*(\mu=[s]_m)$.

最後に，写像 $p:M(U)\to P(U)$を $p(\mu)=supp(\mu)$で定義すれば，これも明らかに準同型である．$M(U)(=U^*/\sim_m)$ 上の二項関係$\sim_p$を $\mu\sim_p\mu'\Leftrightarrow supp(\mu)=supp(\mu')$によって定義すると，明らかに同値関係であり，マルチ集合上へのべき等性の導入を意味する．以下，P(U)と商集合 $M(U)/\sim_p$ との関係を実際に確かめる．同値類$[\mu]_p$ の台は代表元の取り方に依らずに $supp_p([\mu]_p)=supp(\mu)$で定義できる．同値類$[\mu]_p$, $[\mu']_p$ の相等性について次が成り立つ：$[\mu]_p=[\mu']_p\Leftrightarrow supp_p([\mu]_p)=supp_p([\mu']_p)$. $[\mu]_p$ の特性関数 $ct_p:U\times(M(U)/\sim_P)\to 2$ を $ct_p(x,[\mu]_p)=1\Leftrightarrow x\in supp_p([\mu]_p)$で定義する．任意の$[\mu]_p$, $[\mu']_p\in M(U)/\sim_p$ に対して，演算$\cup_p$ と$\cap_p$ をそれぞれ，$[\mu]_p\cup_p[\mu']_p=[\mu\cup_m\mu']_p$, $[\mu]_p\cap_p[\mu']_p=[\mu\cap_m\mu']_p$ で定義する時，特性関数に関して，以下の性質が成り立つ：$ct_p(x,[\mu]_p\cup_p[\mu']_p)=\max(ct_p(x,[\mu]_p),ct_p(x,[\mu']_p))$, $ct_p(x,[\mu]_p\cap_p[\mu']_p)=\min(ct_p(x,[\mu]_p),ct_p(x,[\mu']_p))$. これから，べき等律，交換律，結合律，吸収律，分配律が容易に導かれる．よって，$\cup_p$, $\cap_p$ はそれぞれ join, meet であり，構造$\langle M(U)/\sim_p,\cup_p,\cap_p\rangle$は少なくとも分配束である．部分集合の加法 $+_p$はマルチ集合の加法 $+_m$から，任意の$[\mu]_p$, $[\mu']_p\in M(U)/\sim_p$に対して，$[\mu]_p+_p[\mu']_p=[\mu+_m\mu']_p$ によって定義できるが，演算$\cup_p$ に縮退する：$[\mu]_p+_p[\mu']_p=[\mu]_p\cup_p[\mu']_p$. $M(U)/\sim_p$上の順序を$[\mu]_p\subseteq_p[\mu']_p\Leftrightarrow[\mu]_p\cup_p[\mu']_p=[\mu']_p$ $(\Leftrightarrow[\mu]_p\cap_p[\mu']_p=[\mu]_p)$で定義すると，台に関して，$[\mu]_p\subseteq_p[\mu']_p\Leftrightarrow supp(\mu)\subseteq supp(\mu')$が成り立つ．順序集合$\langle M(U)/\sim_p,\subseteq_p\rangle$において，$0_p=[0_m]_p$ は順序$\subseteq_p$に関する最小元となる：$[\mu]_p\cap_p0_p=0_p\cap_p[\mu]_p=0_p$. 一方，$supp(\mu)=U$ となる μ があり，その選び方に依らず，$1_p=[\mu]_p$ を定義でき，$\subseteq_p$ に関する最大元となる：$[\mu]_p\cap_p1_p=1_p\cap_p[\mu]_p=[\mu]_p$. 構造$\langle M(U)/\sim_p,\subseteq_p\rangle$に最大元が存在するので，任意の$[\mu]_p$ に対して，$supp(\mu)=(supp(\mu'))^C$を満たす μ, μ'が存在し，それらの選び方によらずに，演算^{C_P}を$([\mu]_p)^{C_P}=[\mu']_p$で定義できて，次が成り立つ：$([\mu]_p)^{C_P}\cup_p[\mu]_p=1_p$, $([\mu]_p)^{C_P}\cap_p[\mu]_p=0_p$. 従って，構造$\langle M(U)/\sim_p,\cup_p,\cap_p,{}^{C_P},1_p,0_p\rangle$はブール代数である．以下，混乱が生じない限り，構造$\langle M(U)/\sim_p,\cup_p,\cap_p,{}^{C_P},1_p,0_p\rangle$を $M(U)/\sim_p$ と略記する．べき集合 P(U)と $M(U)/\sim_p$ との関係について述べる．写像 $\xi: P(U)\to M(U)/\sim_p$ を定義するために，$A\in P(U)$に対し $ct_p(x,[\mu]_p)=A$ を満たす $\mu\in M(U)$が存在することに注意し，その選び方に依らずに$\xi(A)=[\mu]_p$で定義できる．ξ は全単射であり，演算∪および∩，^{C}，最大元，最小元を保存する：$\xi(A\cup B)=\xi(A)\cup_p\xi(B)$, $\xi(A\cap B)=\xi(A)\cap_p\xi(B)$, $\xi(A^C)=\xi(A)^{C_P}$, $\xi(U)=1_p$, $\xi(\emptyset)=0_p$. 従って，ξ は P(U)から $M(U)/\sim_p$への同型写像である：
【命題】$P(U)\cong M(U)/\sim_p$
よって，以下が成り立つ：$X\in P(U)\Leftrightarrow\exists\mu\in M(U)\,(X=[\mu]_p)\Leftrightarrow\exists s\in U^*(\mu=[[s]_m]_p)$.

なお，以上の議論はファジィ集合(cf.[2])やファジィマルチ集合などに拡張できる(cf.[6,7]).

文献：

[1] 赤間世紀 (2007): 初歩の SQL，技報堂出版.
[2] 赤間世紀，宮本定明 (2008): ソフトコンピューティングのロジック，工学社.
[3] S.Miyamoto (2004): Generalizations of Multisets and Rough Approximations, International Journal of Intelligent Systems, Vol.19, pp.639-652.
[4] 村井哲也 (2004): 初歩のデータベース，昭晃堂.
[5] T.Murai, S.Miyamoto, M.Inuiguchi, S.Akama (2012): Granular Hierarchical Structures of Finite Naïve Subsets and Multisets Based on Free Monoids and Homomorphisms. International Journal of Reasoning-based Intelligent Systems, Vol.4, pp.118-128.
[6] T.Murai, S.Miyamoto, M.Inuiguchi, Y.Kudo, S.Akama (2014): Crisp and Fuzzy Granular Hierarchical Structures Generated from a Free Monoid. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.18, pp.929-936.
[7] T.Murai, S.Miyamoto, M.Inuiguchi, Y.Kudo, S.Akama (2015): Fuzzy Multisets in Granular Hierarchical Structures Generated from Free Monoids. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.19, pp.43-50.
[8] 小野寛晰 (1994): 情報代数，共立出版.