理学部数学科2000前期 計算数学1[1]

担当：辻下　徹[2]

目次

[1] URL:http://fcs.math.sci.hokudai.ac.jp/doc/announce/cs00.html
質問提出アドレス:cs2000@fcs.math.sci.hokudai.ac.jp

[2] Email:tujisita@math.sci.hokudai.ac.jp,
Homepage:http://fcs.math.sci.hokudai.ac.jp/tjst/

6　第６回：形式言語と文法

きょうやること

- 形式言語
- 書き換え規則
- 文法

6.1　概要

「具体的な」数学的対象は、文字の並びであらわされることが多い。数学で文字列自身に注目することは少ないが、計算理論では「語」「リスト」などと呼ばれ基本的な研究対象となる。

一定の方法で選び出した語の全体を**形式言語**という。「一定の方法」としては、「**文字列変換**」（文法）によるものと、**オートマトン**によるものとがある。

文法のクラスとして主に４種類あり、それに応じて、形式言語の階層が定義される。これを**チョムスキー階層**という。

以下、基本概念を定義する。

6.2　語のモノイド

文字集合と呼ばれる有限集合 Σ があるとする．この集合の要素を有限個並べたものを、Σ の要素を文字とする語あるいは、単に Σ 上の語という．

例 1

$\Sigma = \{ 0, 1 \}$ のとき、

$$0, 1, 00, 01, 10, 11, 000, 001, \cdots$$

などが $\{ 0, 1 \}$ 上の語の例。語は、**リスト**と呼ばれることもあり、そのときは、

$$[0], [1], [0, 0], [0, 1], [1, 0], [1, 1], [0, 0, 0], [0, 0, 1], \cdots$$

などと書くことがある。

例 2

$\Sigma = \{ 0, 1, \cdots, 9, a, b, c, \cdots, z, A, B, C, \cdots, Z, \text{“.''}, \text{“,''}, \text{“!''}, \text{“?''}, \text{“''} \}$ のとき、通常の英文は、Σ 上の語とみなせる。

例 3

$\Sigma = \{$ 英語の辞書にある単語 $\} \bigcup \{$ 句読点 $\}$ のとき、Σ 上の語は、普通の言い方をすると単語を並べたものである。普通の文も Σ 上の語であるが、逆は正しくなく、ほとんどは文としてナンセンスなものとなる。$the, a, dog, man \in \Sigma$ だから、$[a, the, dog, aman]$ も Σ-語である。

この場合のように、「文字」自身が他の文字による語であるときには、リストとして書かないと混乱する場合がある。例えば、文字集合が $\{ 0, 01, 10, 11 \}$ のとき、0110 は $[01, 10]$ と、$[0, 11, 0]$ のいずれかかはわからない。

Σ 上の（空語も含む）語の全体を Σ^* と書く．空語を除いたものを Σ^+ と書く．

２つの語 w_1, w_2 を並べると新しい語 $w_1 w_2$ となる．これにより，語に**併置演算**と呼ばれる２項

演算が定義され**結合法則**

$$(w_1w_2)w_3 = w_1(w_2w_3)$$

を満たす．空語は他の語に並べても何も変わらない（と考える），すなわち

$$\Lambda w = w, \quad w\Lambda = w,$$

が任意の語 w についてなりたつ．

以上のことを，Σ^* は**モノイド**(monoid) であるという．

例 文字が 1 個しかないとき，その語のモノイドは，自然数とゼロの集合に加算を与えたモノイド $(\mathbf{N} \cup \{ 0 \}, +)$ と同型である。

用語 $w = v_1v_2v_3$ であるとき、v_1 を w の接頭語 (prefix), v_2 を w の部分語 (subword), v_3 を w の接尾語 (suffix). v_2 が空でないとき、v_2 は w に出現するという。その出現する場所を、v_2 の w における出現箇所 (occurence) という。

例 01010 の接頭語は 6 個、接尾語は 6 個、部分語は

$$\{ \Lambda, 0, 1, 01, 10, 010, 101, 0101, 1010, 01010 \}$$

の 10 個ある。010 は 2 回出現するが、その出現箇所は重なっている。

6.2.1 形式言語の定義

Σ^* の部分集合のことを Σ をアルファベットとする形式言語という。形式言語を具体的に与えるには次の 2 通りの方法がある：

1. 語がその言語の語彙であるかどうかを判定する方法を与える。(「計算可能な」特性関数を与える。)

   例 $\left\{ w \in \{ a, b \}^* \,\middle|\, a, b \text{ が同じ個数 } w \text{ に出現する} \right\}$ この場合は、w が語彙かどうかは、単に a, b の個数を数えれば判定できる。

   例 $\left\{ 1^n \in \{ 1 \}^* \,\middle|\, n \text{ は素数} \right\}$ これも、自然数が素数かどうかを判定するプログラムが書けるので、「具体的に」特性関数が与えられている。

2. その言語の語彙をすべて列挙する具体的方法を与える。今日は、以下これについて述べる。

後者のように与えられる言語は**帰納的に枚挙可能**(recursively enumerable) であるという。

6.3 書換規則と文法

6.3.1 書換規則

$\Sigma^+ \times \Sigma^*$ の有限部分集合を Σ 上の書換システムという。書換システム R の元を書換規則 (rewriting rule) あるいは生成規則 (production) という。

- $(v, w) \in R$ であるとき、$R : v \to w$ と書く。

- 書換システム R は Σ^* 上の 2 項関係 $\to_R$ (直接書換) を定める:

$$W \to_R V \overset{def}{\iff} W = w_1 v w_2 \wedge V = w_1 w w_2.$$

これにより、Σ^* を頂点集合とする有向グラフが定まる。

- 更に書換システム R は Σ^* 上の 2 項関係 $\to_R^*$ (書換) を定める:

$$w \to_R^* v \overset{def}{\iff} w \text{ から } v \text{ への道がある}.$$

$w \to_R^* v$ のとき、w は**書換システム R に従って v にに書き換えられる**という。

6.3.2 書換システムの例

例

$\Sigma = \{\ 0, 1\ \}$, $R : 1 \to 10, 0 \to 01$ のとき

$$0 \to 01 \to 0110 \to 01101001 \to 0110100110010110 \to \cdots.$$

例

$\Sigma = \{\ 0, 1\ \}$, $R : 010 \to 001, 101 \to 011$ のとき
$011\underline{010}0 \to_R 011\underline{001}0$,
$01\underline{101}00 \to_R 01\underline{011}00$
$0110100 \to_R^* 0011100$

例

$\Sigma = \{\ S, X, +, *, \text{“}(\text{''}, \text{“})\text{''}\ \}$, R は 4 個の書換規則からなる。

- $S \to (S + S)$,
- $S \to (S * S)$,
- $S \to X$.

$$S \to (\underline{S} + S) \to ((S * \underline{S}) + S) \to ((S * (S + S)) + S) \to ((X * (X + X)) + X).$$

例題

$1, \bullet$ という文字を含む文字集合上の書換システムで次のようなものを作れ。どの自然数 $n > 0$ についても

$$\bullet 1^n \bullet \to_R^* 1^{2n}.$$

解答例 1. $\Sigma = \{\ 1, \bullet\ \}$ とし、

$$11 \bullet \to 1 \bullet 11, \bullet 1 \bullet \to 11.$$

解答例 2. $\Sigma = \{\ 1, \bullet\ \}$ とし、

$$1 \bullet \to \bullet 11, \bullet \bullet \to \Lambda.$$

練習問題 $1, 0, \bullet$ という文字を含む文字集合上の書換システムで、次のようなものを作れ。

$$\bullet w \bullet \to^* u$$

ただし、u は w から 0 を省いたもの。

6.3.3　書換規則系による形式言語の生成

文字集合 Σ 上の書換システム R と Σ 上の語のいくつかの集まり L_0 があるとき，L_0 の元を次々と R で書き換えて得られる語の全体を $\mathcal{L}(\Sigma, R, L_0)$ とかく.

例

$$\boxed{\mathcal{L}(\{\,0,1\,\},\{\,11\to 111,\quad 1111\to 101\,\},\{\,11\,\}) = \left\{\,1w1 \;\middle|\; w \text{ には } 00 \text{ が出現しない}\,\right\}}$$

6.3.4　例題

バランスのとれた括弧式の全体は文字集合 $\{\,``('', ``)''\,\}$ 上の形式言語を成す。これを $\mathcal{L}(\Sigma, R, L_0)$ と表現せよ。

6.4　文法による言語生成

書換変形の過程で補助的な記号を用いる事で、書換システムの言語記述力は増す。

6.4.1　定義

文法とは次のようなものがなす組 $(\Sigma_T, \Sigma_N, R, S)$ のことである．

- Σ_T と Σ_N は有限集合であり，Σ_N の元は**文法記号**とよばれ、Σ_T の元は端末記号 (terminal symbol) と呼ばれる。
- R は 和集合 $\Sigma_N \coprod \Sigma_T$ 上の書換システム。
- S は**始記号**と呼ばれる文法記号。

例　$\boxed{(\{\,0,1\,\},\{\,S,A,B\,\},\{\,S{\to}AS, S{\to}SB, A{\to}0S1, B0A{\to}000\,\},S)}$

6.4.2　文法の定める形式言語

文法 $G=(\Sigma_T,\Sigma_N,R,S)$ は次のようにして Σ_T 上の形式言語を決める．語 $S{\to}^*_R w \in \Sigma_T^*$ のとき、w は**文法** G **に則した語**という．G に則した語の全体を $\mathcal{L}(G)$ と書き，**文法** G **が決める形式言語**と呼ぶ.

ある文法 G によって $\mathcal{L}(G)$ と表示できる形式言語は**帰納的に枚挙可能** (recursively enumerable) であるという。

6.4.3　文法のチョムスキー階層

文法は、生成規則の形によって４つのクラスに階層付けられ、各々は、その下の階層より本当に広いことが知られている。

以下 X,Y は補助記号を、a,b は端末記号,w は端末記号の語, W,W' は任意の語.

1. **正則文法**(regular)：書き換え規則は $X{\to}wY$ または $X{\to}w$ の形に限る．
2. **文脈自由文法**(context-free)： $X{\to}W$.
3. **文脈依存文法**(context-sensitive)：$W{\to}W'$, ただし $\ell(W)\le\ell(W')$.
4. 一般の文法

正則文法・文脈自由文法・文脈依存文法・一般の文法で記述される形式言語を各々、正則言語・文脈自由言語・文脈依存言語・再帰的に枚挙可能な言語という。

6.5　正則言語と有向グラフ

正則言語の文法は $X{\to}wY$ という形だが、$w=a_1a_2\cdots a_m$ のときに補助記号 $Y_1,\cdots,Y_{m-1}$ を増やし、書換え規則 $X{\to}a_1Y_1$, $Y_1{\to}a_2Y_2,\cdots,Y_{m-1}{\to}a_mY$ を追加し、$X{\to}wY$ を削除しても定義される形式言語は変化しない。従って、書換規則としては $X{\to}aY$ または $X{\to}a$ という形のものだけであるとしてよい。このとき、正則文法に対して次のようなラベル付きのグラフが定まる。

- 頂点は補助記号と新しい頂点 F.
- 書換規則 $X{\rightarrow}aY$ に対しては、辺 $X\overset{a}{\rightarrow}Y$.
- 書換規則 $X{\rightarrow}a$ に対しては、辺 $X\overset{a}{\rightarrow}F$.

このとき、$S{\rightarrow}^{*}w$ という書換過程は、このラベル付きグラフ上の S から F への道と対応し、w はその道の辺についているラベルを並べたものになる。

6.6　文脈自由言語

6.6.1　文脈自由文法の BNF 表示

(Backus-Naur Form)：

$$X ::= w_1 \mid w_2 \mid \cdots \mid w_n,$$

は $X{\rightarrow}w_i \quad i=1,\cdots,n$ が生成規則であることを表す。

6.6.2　文脈自由言語の例

例1：　$\Sigma=\{\ a,b\ \},\ \Sigma_N=\{\ S\ \}$,

$$S ::= \Lambda \mid aSb \mid SS.$$

例2： (項)：

- $\alpha:\Sigma{\rightarrow}\mathbf{N}\cup\{\ 0\ \}$:arity 関数
- $S ::= v(v\in V)$,$S ::= \sigma S^{\alpha\sigma}\ (\sigma\in\Sigma)$
- この文法に則した語、V を変数、Σ を関数記号とする**項** (term) という。Σ_0 の元は定数 (constant) という。

6.7　演習問題

[問題 (6-1)] $\{\ 1,\bullet,B\ \}$ を含む文字集合上の書換系 R で

$$B1^n\bullet 1^m B{\rightarrow}^{*}_{R}1^{nm}$$

となるものを作れ。

[問題 (6-2)] 次の文法の定める言語は何か？

(a) $(\{\ a,b\ \},\{\ S\ \},\{\ S ::= aSb|ab\ \},\{\ S\ \})$

(b) $(\{\ a,b\ \},\{\ S,A,B\ \},R,\{\ S\ \})$ ただし R は

$$\begin{array}{rcl} S & ::= & aB|bA \\ A & ::= & a|aS|bAA \\ B & ::= & b|bS|aBB \end{array}$$

6.8 前回の質問表

[**Q6-1**](22950035) 状態空間は「観測できるもの」、「状態遷移規則」は「物理法則」に当てはまるものと考えていいのでしょうか？」

[**A6-1**] 観測できるものの中に物理法則が見出されたときには、力学系的な記述が使えるようになります。しかし、世界は因果律で支配されている、というのはもっと根本的な作業仮説であると思います。

[**Q6-2**](22960018) オートマトンが有限力学系の族であるならば、わざわざ区別して定義する必要があるのですか。

[**A6-2**] パラメータを動かすので違う働きをします。

[**Q6-3**](22960115) $\mathbf{Z}/n\mathbf{Z}$ のときのサイクルの周期は n というような図がありましたが、その意味がわかりません。$\mathbf{Z}/13\mathbf{Z}$ の時の周期は１３ではなかったです。

[**A6-3**] ２つの力学系が出てきています。$(\mathbf{Z}/13\mathbf{Z}, \tau_1)$,$(\mathbf{Z}/13\mathbf{Z}, \tau_2)$ $\tau_1(n) := n + 1 \mathrm{mod} 13$, $\tau_2(n) := n^2 \mathrm{mod} 13$ です。前の方の力学系の周期は１３、後者はこの前に図を書いたものです。

[**Q6-4**](22970002)$\tau^i x$ の意味がわかりません。

[**A6-4**] $\tau^i x$ は $\tau(\tau(\tau(\cdots(\tau(x)))\cdots))$($\tau$ が i 個) の略です。

[**Q6-5**](22970010) オートマトンの具体的な例として、バネばかりにおもりを下げるという問題を考えた場合、どういうグラフ表示になりますか？

[**Q6-6**](22970037) グラフを書けとはどういうことですか。やはり、グラフというものがよくわかりません。

(質問理由：2page 5.2 の例の演習問題で、$p = 23$ のグラフを書くことは、（周期２２のサイクルと固定点 0 を書いた図がここに書いてある）を書くことですか。それともこれは、単にグラフ表示したにすぎないのですか。)

[**A6-6**] 数学的対象としてのグラフは (V, E) $(E \subseteq V \times V)$ という対のことです。上の力学系のグラフは $(\{\, 0, 1, 2, \cdots, 22 \,\}, \left\{ (i, 10i \mathrm{mod} 23) \;\middle|\; i = 0, 1, 2, \cdots, 22 \right\})$ です。この「グラフ」を書くというのは、君が書いたような図を書くことです。

[**Q6-7**](22970041)Collaz の問題のようにサイクルがありそうで、パッと思いつかず１から書いていくしかないものは、順番に書いていくようなグラフしかないのでしょうか？

[**A6-7**] Collaz の問題は無限力学系ですので図に書き尽くすことはできません。この問題は未だに未解決です。

[**Q6-8**](22980007) 有限力学系があるということは、無限力学系もあるのですか？
(質問理由：Xが有限のとき、有限力学系と定義しましたが、Xが無限のときにも、状態遷移規則が設定されているものや、設定してやったりできるものがあると思います。そして、もし無限の力学系があったとして、その性質というのは、有限力学系によって得られた性質を、そのまま適用したりすることができるのかなと思い、質問しました。また、無限の力学系では、新たな性質があるのでしょうか？)
[**A6-8**] 今研究されているものは大抵無限です。カオス理論の中核をなるのが [0,1] を状態空間とする力学系（区間力学系）です。

[**Q6-9**](s980047) 質問：オートマトンはいつでも数え上げることができるのですか。
(質問理由：｜Q｜や｜Σ｜がmとかnであらわされた時でも、一般にmとnで数を表せますか。) そういう結果があります。単純な式ではありませんが。

[**Q6-10**](s980008) オートマトンが退化ということがよく分かりません。
(質問理由：テキストに書いてなくないですか。)
[**A6-10**] 「テキスト」ではなく「補助資料」です。口頭で言いました。$\tau(q,\sigma)$ が σ によらないときに退化する、と言います。

[**Q6-11**](s980030) ２つの離散力学系が共通部分を持っていた時の直和もとなり同士に並べたものになるのですか？
(質問理由： ２つの離散力学系で、共通するところがあったときは、その部分を重ね合わせたものにするのですか？)
[**A6-11**] そうです。集合の直和と同じで 1 と 1 の直和は 1,1' となります。（後半への答え：そういう演算（張りあわせ）も別にありますが、この講義では扱いません。

[**Q6-12**](s980038) Mathematica では、4 次元以上のベクトルの図示などはできるのでしょうか？もしできるなら、どのようになるのでしょうか？
[**A6-12**] いろいろ工夫が要ります。ベクトルを関数と見て（４次元ベクトルは 1,2,3,4 から実数への関数）グラフ表示することが多い（これは、いわゆる棒グラフと同じ）。

[**Q6-13**](s980038) それと、計算数学１のホームページ `http://fcs.math.sci.hokudai.ac.jp/doc/announce/cs00.html` とか、そのページにある「講義関係の情報」`http://fcs.math.sci.hokudai.ac.jp/doc/announce/00cs/index.html` は更新されていないのですが、いつ更新されるのでしょうか？
[**A6-13**]「講義関係の情報」> `http://fcs.math.sci.hokudai.ac.jp/doc/announce/00cs/index.html` というページは作っていません。訂正しておきます。御指摘ありがとうございました。> `http://fcs.math.sci.hokudai.a`
にに掲載しています。

[**Q6-14**](22980051) なぜ「退化」というのか。状態が変わっていないのであまり「退化」したという感じはしないのですが。また、退化でない場合を進化といったりはしないのですか？

[**A6-14**] 「退化」は、概念的により単純な構造に帰着される、という意味で使っています。概念的には力学系を進化させたものの一つがオートマトンですが、進化のさせかたは沢山ありますので専門用語にはなりにくいと思います。〔マルコフ連鎖なども「進化」の一例になります。)

[**Q6-15**](22990019) オートマトンでいう同型は写像でいう同型写像みたいなものですか？どういう分野をやっているのかわかりません。（集合とか群とかと関連してますか）

[**A6-15**] 集合は空気のように使っています。オートマトンの同型は、２つの全単射〔集合の同型写像）から成っています。どういう分野か分類する必要はありません。内容は集合論を知っていれば理解できますし、集合論がわからない人は素朴な有限集合だけで十分わかりようにしています。

[**Q6-16**](22007806) カオスが有限で考えられるというのがわからない

[**A6-16**] カオスの現象を計算機実験で見るときは、皆有限近似で観察している、という程度に考えてください。

[**Q6-17**](22990045) 世界が因果関係で構成されていることを前提として考えるのならば、世界にはじまりはあるのか、という命題については言及しないという事ですか因果で世界の始まりは説明できない、と何かの本で読んだので

[**A6-17**] そうです。力学系は初期状態を説明する方法はありませんし、そもそも、状態空間がどうしてそれなのか、を説明することもできません。その説明は「我々側にある」というべきものです。

[**Q6-18**](22980034)$\tau(q,\sigma)$ の値はどういう意味を持つのですか

[**A6-18**] $q' := \tau(q,\sigma)$ のとき $q \xrightarrow{\sigma} q'$ と書くと分かりやすいと思います。状態 q のとき入力 σ を受取ると q' に変わるということを表します。

[**Q6-19**](22980032)5.4 の定理１と、既約であることの定義は同値ではないですか？

[**A6-19**] 正確には、定理１は既約であるための必要条件を与えていると言うことができます。君の問いは、定理１は既約であることの十分条件にもなっていないか、というものです。実際に十分条件になっています。

どちらを定義にしてもいいですが、それが同値であることは証明する必要があります。

[**Q6-20**](22980031) 一方通行化問題は一筆書きの問題と同じものですか

(質問理由：一方通行化問題は一筆書きの問題と同様に各頂点について奇点か偶点かどうかを調べればできそうな気がします。)

[**A6-20**] 一筆書きができれば一方通行化問題も解決でています。しかし、逆は正しくありません。

[**Q6-21**](22980020) オートマトンの数え上げは具体的に何に使うのですか？またこういうのは計算機では計算できないのですか？

[**A6-21**] 数え上げによって、普通には思いつかないような面白いものが見出されることがあります。一般に分類はそういう効果があります。なお、計算機で計算することも可能なはずですが、そうしてしまうと、今言った意味の効用がなくなってしまいます。

[**Q6-22**](22980013)「システム」て何ですか

[**A6-22**] 半独立的な部分が協働して統一性のある存在を成しているときシステムと言いう言葉を使います。

[**Q6-23**](22980012) 感情の起伏や記憶の変化のようなこともオートマトンで表せるのではないか

[**A6-23**] 表したいと思えば、ある程度の所まではできる。

[**Q6-24**](22980011) 力学系の直和は、他の科目の既習の直和とどのような関係がありますか？

[**A6-24**] 直和と呼ばれるものは、全く異なる数学的構造に使われる場合でも同じ性質で特徴付けられます。カテゴリー理論では皆定義が同じになります。

[**Q6-25**](22980002) 乗算器を逆回しに、入力・出力を入れ替えることで、素因数分解を効率良く計算できないか。 面白い発想ですが少し的が外れているように思います。オートマトンの構造では、入力と出力の役割は逆転できな構造的非対称性があるので、「入力・出力を入れ替える」という意味は定かではないです。しかし、余り気にせず考えてみてください。

[**Q6-26**](22970049) オートマトンは非常に単純に見えます。多数の組み合わせで複雑なことができそうには思えますが、どのぐらいの組み合わせでどのくらいのことができるのでしょうか。

[**A6-26**] 万能チューリング機械の構造も、とても簡単なものです。