

アルゴリズムとデータ構造III

15回目

期末試験



問題1 文脈自由文法

- 文脈自由文法と文脈依存文法の違いを200文字以内で説明せよ.
- 文脈依存文法の生成規則は $u\alpha v \rightarrow u\beta v$ (α は非終端記号, β, u, v は終端または非終端記号列) の形で表される. これは非終端記号 α の前後の記号 u, v により α から β の導出が制限される事を意味する. $u\alpha v \rightarrow u\beta v$ で u と v が ε であるとき $\alpha \rightarrow \beta$ となり, 前後の記号(文脈)は α から β の導出に影響を与えない. そのため $\alpha \rightarrow \beta$ のような生成規則だけを持つ文法を前後の文脈(記号)に対して影響を受けないという意味で文脈自由文法と呼ぶ.

問題2 CYK法

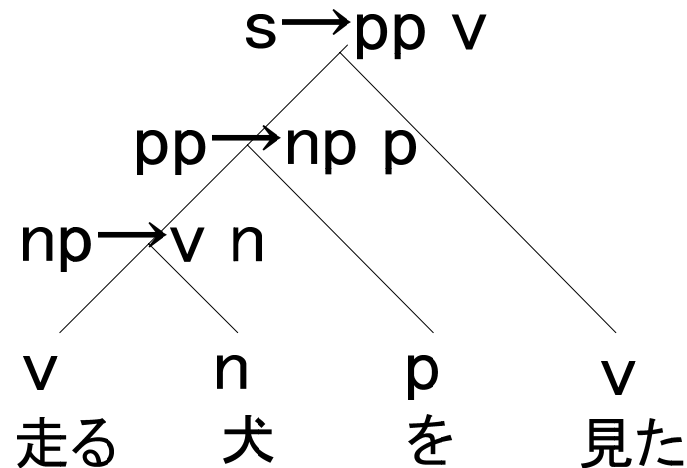
- 下の図は「走る犬を見た」を構文解析中のCKY表である。
- 図中の①, ②, ③, ④, ⑤には何が入るか答えよ。
- CKY表から得られる「走る犬を見た」の構文木を描け。

	走る	犬	を	見た
走る	v → 走る	①	③	⑤
犬		n → 犬	②	④
を			p → を	
見た				v → 見た

書き換え規則

- (1) $s \rightarrow pp\ v$
- (2) $vp \rightarrow pp\ v$
- (3) $np \rightarrow vp\ n$
- (4) $np \rightarrow v\ n$
- (5) $pp \rightarrow np\ p$
- (6) $pp \rightarrow n\ p$
- (7) $n \rightarrow 犬$
- (8) $p \rightarrow を$
- (9) $v \rightarrow 走る$
- (10) $v \rightarrow 見た$

- ①: (4)
- ②: (6)
- ③: (5)
- ④: (1), (2)
- ⑤: (1), (2)



問題3 ダイクストラ法, Aアルゴリズム, A*アルゴリズム

- ダイクストラ法, Aアルゴリズム, A*アルゴリズムの特徴を説明せよ.
- ダイクストラ法, Aアルゴリズム, A*アルゴリズムともに最短経路探索問題に利用される.
- ダイクストラ法は始点から各ノードまでの最短距離を求めながらゴールノードまでの距離を調べる
- Aアルゴリズムは各ノードからゴールノードまでの推定距離を用いて, 始点からそのノードを經由してゴールノードにいたる最短距離が短い順に探索を進める.
- Aアルゴリズムで各ノードからゴールノードまでの推定値が実際の距離よりも長くなることがないことが保障されている場合A*アルゴリズムと呼ばれる.
- A*アルゴリズムで各ノードからゴールまでの推定距離がすべて0の場合, ダイクストラ法と同じ動作になる.

問題4 動的計画法

■ 重さ5kgまで詰め込めるナップサックと3種類の商品(A,B,C)があったとする。各商品の重さと単価は下の表のとおりである。合計金額が最大になるように3種類の商品をナップサックに詰め込むためには各商品を何個詰め込めばよいか答えよ。またどのように考えたかを簡単に説明せよ。

商品名	A	B	C
重さ	1kg	2kg	3kg
単価	100円	300円	700円

■ B:1個, C:1個

■ ナップサック問題の説明をすればよい

■ Aだけを詰め込んだときの各重さの最大金額を求める

重さ	1kg	2kg	3kg	4kg	5kg
合計金額(円)	100	200	300	400	500
最後に入れた物	A	A	A	A	A

■ AとBを詰め込んだときの各重さの最大金額を求める

重さ	1kg	2kg	3kg	4kg	5kg
合計金額(円)	100	300	400	600	700
最後に入れた物	A	B	B	B	B

■ ABCを詰め込んだときの各重さの最大金額を求める
5kgのとき最後にCを入れたのでCを取り除くと2kgになる
2kgの時は最後にBをいれたのでBを取り除く、これをナップサックがからになるまで続ける

重さ	1kg	2kg	3kg	4kg	5kg
合計金額(円)	100	300	700	800	1000
最後に入れた物	A	B	C	C	C

← $2-2=0$ Bを取り出す
← $5-3=2$ Cを取り出す



問題5 KMP法

- KMP法 (Knuth-Morris-Pratt法) の特徴を200文字程度で説明せよ
- 文字列検索アルゴリズム
- シンプルサーチの計算量は最悪の場合 $O(mn)$ になるが KMP法は $O(n)$: n は文字列長, n はキーワード長
- 文字列照合時にバックトラックしない
- 照合失敗時に次の照合でキーワードをどのくらいずらすかを前もって調べておく. どのくらいずらすかは文字列にはよらずキーワードだけで決まる.



問題6 BM法

- BM (Boyer-Moore) 法での検索に適した文字列とキーワードの特徴を述べよ。またその理由を説明せよ。
- 文字列にキーワードに含まれない文字を多く含む
- キーワードが長い
- 文字列にキーワードに含まれない文字列があった場合、キーワードの長さ分だけキーワードをずらすことができ、照合回数を減らすことができる。



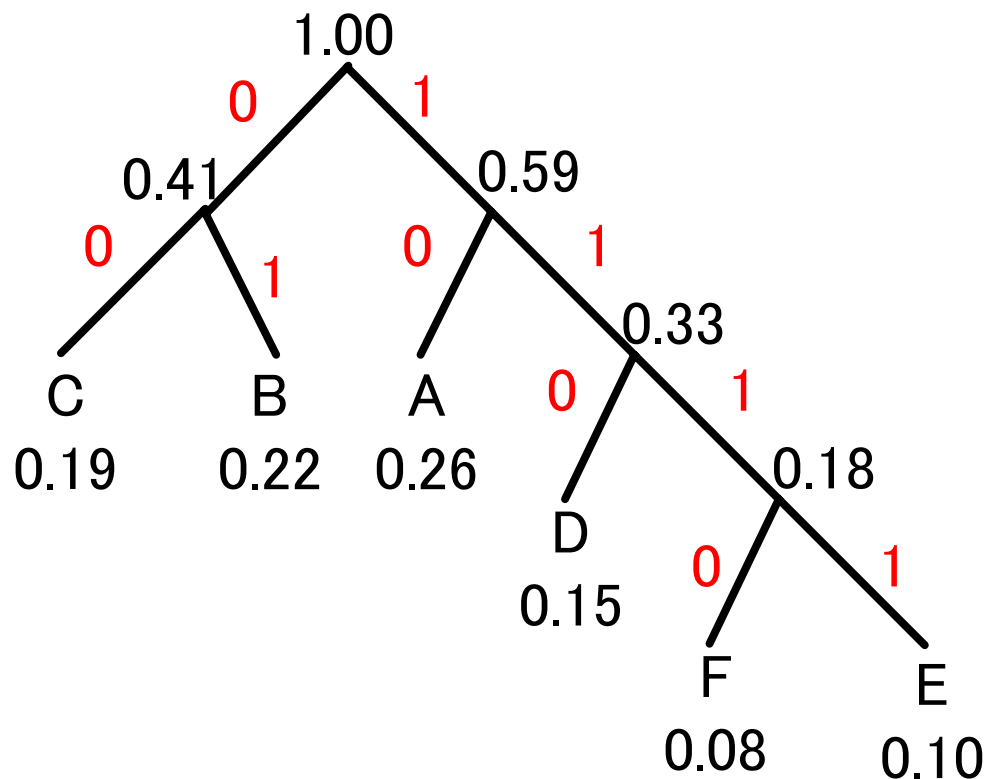
問題7 Aho-Corasick法

- Aho-Corasick法を100文字程度で説明せよ.
- 複数のキーワードを同時に検索できる文字列検索アルゴリズム
- KMP法を複数キーワードに適用した方法
- 複数キーワードより有限オートマトンを作成し, 文字列を入力とすることで文字列照合を行う.
- バックトラックを行わずに複数キーワードを同時に検索できる

問題8 ハフマン符号化

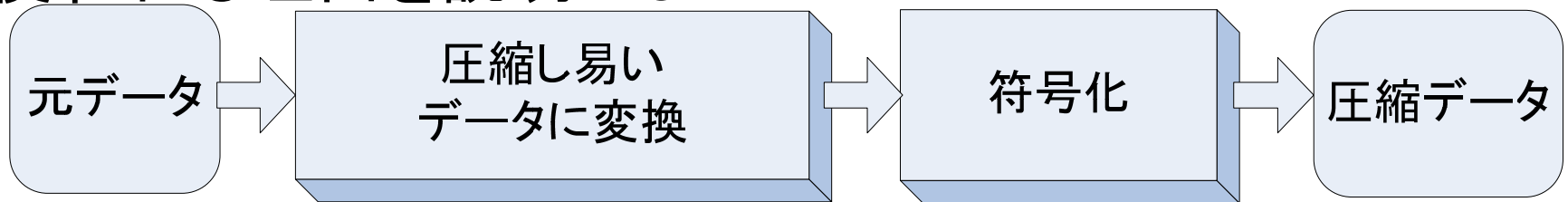
下の表のような記号の出現確率のとき、ハフマン符号を作りなさい。但し、ハフマン符号作成のための二分木も書くこと。

記号	頻度	符号
A	0.26	10
B	0.22	01
C	0.19	00
D	0.15	110
E	0.10	1111
F	0.08	1110



問題9 データ圧縮

- 下の図は一般的なデータ圧縮処理の流れを示している。
- ①「圧縮しやすいデータ」の特徴を2つ挙げよ。
- ②符号化でよく使われる手法は何か。またその手法がよく使われる理由を説明せよ。



- ①:
 - ・出現する事象の確率の偏りが大きい
 - ・出現する事象の数が少ない
- ②: ハフマン符号化
これ以上圧縮できない限界点である平均情報量を簡単な計算で近似できるから

問題10 エントロピー

- アルファベットのAからDまでの4文字が全て等確率で出現すると仮定した場合の平均情報量(エントロピー)求めよ. また各文字の出現確率が下の表である場合, 平均情報量(エントロピー)を求めよ.

文字	A	B	C	D
出現確率	1/2	1/4	1/8	1/8

$$\text{等確率: } H = -4 \times \frac{1}{4} \log_2 \frac{1}{4} = -\log_2 2^{-2} = 2 \text{ビット(シャノン)}$$

$$\begin{aligned} \text{表確率: } H &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{8} \log_2 \frac{1}{8} \\ &= -\frac{1}{2} \log_2 2^{-1} - \frac{1}{4} \log_2 2^{-2} - \frac{1}{8} \log_2 2^{-3} - \frac{1}{8} \log_2 2^{-3} \\ &= \frac{1}{2} + \frac{1}{2} + 2 \times \frac{3}{8} = \frac{7}{4} \text{ビット(シャノン)} \end{aligned}$$



問題11 圧縮(音声)

- ADPCMについて200字以内で説明せよ.
- Adaptive Differential Pulse Code Modulation
- 適応的差分パルス符号変調
- 音声波形は連続的に変化しているので前回のサンプリングからの差分を記録するだけなら量子化ビット数を抑えられる(例えばPCMを用いて16ビットで表していたデータを音質を落とさずに12ビットに圧縮できる)



問題12 圧縮(画像)

■ ①「静止画データの圧縮符号化に関する国際標準」はどれか。下の選択肢から1つ選びなさい。

■ ア: BMP, イ: MP3, ウ: JPEG, エ: MPEG

■ ② 圧縮された情報を伸張しても、完全には元の情報を復元できない場合がある圧縮方式はどれか。下の選択肢から1つ選びなさい。

■ ア: GIF イ: JPEG ウ: MH エ: MR

■ 答え ①: ウ, ②: イ