

## アルゴリズムとデータ構造III

### 15回目

期末試験

### 問題1 動的計画法

- 動的計画法を100文字程度で説明せよ。
- 解くのに時間のかかる問題を、複数の部分問題に分割することで効率的に解くアルゴリズム
- 動的計画法の適用例として、最短経路検索のためのダイクストラ法、パターンマッチングのためのDPマッチングがある。

### 問題2 文字列検索(KMP)

- simple search法と比較した時のKMP法(Knuth-Morris-Pratt法)の特徴を100文字程度で説明せよ
- 照合回数: Simple searchはテキスト中の各文字に対して何度も照合を繰り返す必要があるが、KMP法はテキスト中の各文字に対する照合回数は1回で済む。
- 前処理: Simple searchは照合を行う前に前処理は必要ないが、KMPは照合に失敗した場合の次の処理を予め調べておく必要がある(Next関数)

### 問題3 文字列検索(Aho-Corasick)

- 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4 Zipfの法則

- Zipfの法則を100文字程度で説明せよ。
- 「あるタイプの現象が生起する確率はその現象の生起する順位に反比例する」という経験則
- 都市の人口、コンピュータのコマンドの使用頻度、Webページのアクセス頻度、アルファベットの使用頻度などがZipfの法則に当てはまると言われている。

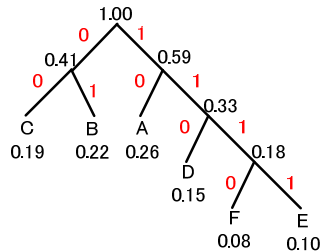
### 問題5 シーザー暗号

- シーザー暗号について以下の問いに答えよ。
- ① シーザー暗号について100文字程度で説明せよ。
  - 換字式暗号の一種。平文の各文字を辞書順に固定数だけずらすことで暗号を作ることができる。
  - 平文の文字を規則に従ってずらすだけなので、文字頻度などを手がかりに、比較的簡単に解読できる。
  - 映画「2001年宇宙の旅」に登場するコンピュータ「HAL」はIBMの各文字を一文字ずらして名付けられたと言われている。
- ② シーザー暗号の解読方法について150文字程度で説明せよ。
  - 暗号文が比較的長ければ各文字の出現頻度、接続文字の出現頻度の情報、単語の出現頻度を利用し、文全体として尤度の高い文字列を見つけ、それを解読文とする。
  - Unixのコマンドcaesarでシーザー暗号のエンコードとデコードができる。
  - 解読(方法)を題材にした小説にポーの黄金虫やコナン・ドイルの「踊る人形」などがある。

## 問題6 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7 ハフマン符号

- モールス信号と比較した時のハフマン符号の特徴を150文字程度で説明せよ。
- 区切り文字: モールス信号は文字の区切り、単語の区切りを表すために区切り文字が必要であるが、ハフマン符号は文字の区切りを表す区切り文字を挿入する必要は無い。
- ノイズに対するロバスト性: モールス信号は区切り文字があるため一部のコードの受信が困難でもそれ以外の部分は正しくデコードできるがハフマン符号はノイズ部分以降は正しくデコードできない可能性がある。

## 問題8 エントロピー

- アルファベットのAからPまでの16文字が全て等確率で出現すると仮定する。平均情報量(エントロピー)を求めよ。

$$H = -16 \times \frac{1}{16} \log_2 \frac{1}{16} = -\log_2 2^{-4} = 4 \text{ bit/symbol}$$

## 問題9 データ圧縮

- 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。



- ・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
- 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 音声圧縮

- PCMとADPCMの違いを100字程度で説明せよ。

- PCM (Pulse Code Modulation) はパルス符号変調、ADPCM (Adaptive Differential Pulse Code Modulation) は適応的差分パルス符号変調
- Differential: PCMは電圧をそのままコード化しているのに対し、ADPCMは前のデータとの差分をコード化している。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
- Adaptive: PCMは電圧の値をそのままコード化しているのに対し、ADPCMは大きな変化の部分では量子化幅を荒くすることで効率よく圧縮できる