

1 アルゴリズム演習 9 モンテカルロ法による の計算

223 頁

ある問題を数値計算で解くのではなく、確率を用いて解くことをモンテカルロ法といいます。円周率の をこの方法で求めてみましょう。

`rand()/32767.0` で発生させる一様乱数は 0 から 1 の範囲ですから、2 組の乱数を x 、 y に対応させると、 1×1 の正方形の中に均一にばらまかれると考えられます。

したがって、正方形の面積と $1/4$ 円の面積の比は、そこにばらまかれた乱数に比例するはずでず。今、 $1/4$ 円の中にばらまかれた乱数の数を a 、円外にばらまかれた乱数の数を b とすると、次の関係が成り立ちます。

$$\begin{aligned} \pi/4 : 1 &= a : a + b \\ &= 4a / (a + b) = 4a / n \end{aligned}$$

なお、 $a + b$ は発生した乱数の総和数 n です。

1.1 モンテカルロ法による円周率の計算 monte.c

入力した文字列の小文字は大文字に、大文字は小文字に変換しなさい。

```
/*
   モンテカルロ法による円周率の計算
   monte.c
*/

#include <stdio.h>
#include <stdlib.h>

int main()
{
    int NUM = 2000;
    int in = 0;
    int i;

    double x;
    double y;
    double pai;
```

```
for(i = 0; i < NUM; i++){
    x = rand() / 32767.0;
    y = rand() / 32767.0;

    if(x * x + y * y < 1.0){
        in++;
    }
}

pai = (double)4 * in / NUM;
printf("pai = %f\n", pai);

return 0;
}
```

2 AIによる大規模データ処理入門 8月分の演習と同じ

2.1 言語処理アルゴリズム

自然言語で記述された大規模データから、その特徴を抽出する方法を考えます。人工知能の歴史においては、自然言語処理の研究は、機械翻訳や自動要約などへの応用を念頭に置いて、さまざまな側面から取り組まれてきました。

2.2 単語レベルでの特徴抽出

n-gram による処理では、文書の自然言語としての特徴は全く利用していませんが、それでも文書のある特徴を表現することが可能です。そこで次に、文書が単語の連なりで記述されているという自然言語としての特徴を使った解析方法を考えます。

日本語の文書では、単語と単語の区切りは明確ではありません。このため、文から単語を切り出す形態素解析には一般に複雑な処理を行いません。これに対して英語では、単語と単語は空白で区切られていますから、文から単語を取り出すのは容易です。そこで、空白を手掛かりに英文から単語を取り出す処理を考えます。(AIによる大規模データ処理入門

215 頁)

2.3 単語を抽出するプログラム getword.c

```
/*
*****
/* データ前処理プログラム          */
/*      getword.c                  */
/* 単語の取り出し                  */
*****

#include <stdio.h>

/*
*****
/*      main() 関数          */
*****
int main()
{
    char chr ; /*入力文字*/

    while((chr=getchar())!=EOF){
        /*アルファベット以外なら改行*/
        if(('A'<=chr) && (chr<='z'))
            putchar(chr); /*一文字出力*/
        else
            putchar('\n') ; /*改行*/
    }

    return 0 ;
}
```

2.4 getword の出力例

```
Alice
was
beginning
to
get
very
tired
of
```

sitting
by
her
sister
on
the
bank

and
of
having
nothing

2.5 getword.exe の使い方

```
./getword < alice.txt
```

```
./getword < alice.txt | sort | uniq -c | sort -rg
```