

2011 年 2 月 17 日

自主課題研究

情報システム工学科 4 年 088 太田 隼人

ハフマン符号化とエントロピー

1. 目的

ハフマン符号によるデータ圧縮法の理解および符号化された符号語列に乱数性がみられるか検証を行う。

2. 概要

情報理論において、データ圧縮は符号化により 0 と 1 のビット列によって表すこととされる。データ圧縮とはビット列の長さをできるだけ短くするような効率的な符号化として定式化される。

今回扱う D.A.Huffman が開発したハフマン符号は、情報源のモデルを一旦作成した後、得られたモデルに基づいて入力記号列を符号化する静的符号化方式の 1 つであり、情報源に対して最短な符号を示すコンパクト符号である。

ハフマン符号による符号化および復号化を実現する C 言語によるプログラミングを構成した。ソースファイルは共通ヘッダファイル `c-huff.h`、`bitio.c` のプロトタイプ宣言 `bitio.h`、`e-huff.c`、`d-huff.c` の複数のソースファイルをコンパイルした。

データ圧縮によりハフマン符号で生成された符号語列に乱数性が見られるかを簡単な頻度検定を用いて検証した。研究の対象にしていた符号語列はこの側面からみると乱数と認めてもいいと考えられる。他の乱数検定法を用いて検証を行うことが望まれる。