

情報科学概論A

2015/06/05 第08回目
情報の効率的なデータ表現

データ圧縮

- * 可逆圧縮
 - * 元情報を保ったままデータ量を減らす
 - * 元のデータに戻ることができる
- * 非可逆圧縮
 - * 人間があまり強く認識できない成分を削除しデータ量を減らす
 - * 元のデータに戻ることができない

データ圧縮に利用されている技術

- * シャノン符号
 - * ハフマン符号の方がよい
 - * ほぼ使われていない
- * ハフマン符号
- * ランレングス圧縮

天気予報

	符号
快晴	000
晴れ	001
曇り	010
雨	011
雷	100
雪	101

- * 6種類の予報
 - * 快晴、晴れ、曇り、雨、雷、雪
- * そのまま符号化すると右の表のようになる
- * 1日1回予報を出した場合、100日間予報したとき何bit送信？

シャノン符号

	年間確率
快晴	15%
晴れ	43%
曇り	15%
雨	20%
雷	2%
雪	5%

- * 確率が高いものから順番に並べる
- * 確率が半分になるところで分割する
- * 分割したあとさらに半分になるところで分割する

シャノン符号 - 平均符号長

	年間確率	符号	確率×bit数
快晴	15%		
晴れ	43%		
曇り	15%		
雨	20%		
雷	2%		
雪	5%		
平均符号長			

ハフマン符号

	年間確率
快晴	15%
晴れ	43%
曇り	15%
雨	20%
雷	2%
雪	5%

- * 確率が高いものから順番に並べる
- * 確率が1番低いのと2番目に低いのを結合
- * 結合したものをグループとし確率を合算する

ハフマン符号 - 平均符号長

	年間確率	符号	確率×bit数
快晴	15%		
晴れ	43%		
曇り	15%		
雨	20%		
雷	2%		
雪	5%		
平均符号長			

ラングレス圧縮

- * 連続して現れるデータを繰り返しの回数で置き換えてデータ量を削減

例えば…)

AAAAAABBBBBBCCCCCCCC