

Kompresja

Kompresja danych (ang. data compression) - polega na zmianie sposobu zapisu **informacji** tak, aby zmniejszyć **redundancję** i tym samym objętość zbioru, nie zmieniając przesyłanych informacji. Innymi słowy chodzi o wyrażenie tego samego zestawu informacji, lecz za pomocą mniejszej liczby bitów. Działaniem przeciwnym do kompresji jest dekompresja.

Kompresja dzieli się na **bezstratną** - w której z postaci skompresowanej można odzyskać identyczną postać pierwotną, oraz **stratną** - w której takie odzyskanie jest niemożliwe, jednak główne właściwości które nas interesują zostają zachowane, np. jeśli kompresowany jest obrazek, nie występują w postaci odtworzonej widoczne różnice w stosunku do oryginału. Pomimo to może się już nie nadawać zbyt dobrze np. do dalszej przeróbki czy do wydruku, gdyż w tych zastosowaniach wymaga się zachowania innych właściwości.

1.4. Reprezentacja informacji

Kompresja

Algorytmy kompresji dzieli się na algorytmy zastosowania **ogólnego** oraz algorytmy do danego typu danych. Z definicji nie istnieją algorytmy kompresji stratnej zastosowania ogólnego, ponieważ dla różnych typów danych konieczne jest zachowanie różnych właściwości. Na przykład kompresja dźwięku używa specjalnego modelu psychoakustycznego, który nie ma sensu w zastosowaniu do obrazu, poza bardzo ogólnymi przesłankami dotyczącymi sposobu postrzegania rzeczywistości przez człowieka.

Kompresja

Większość algorytmów bezstratnych to algorytmy zastosowania ogólnego oraz ich drobne przeróbki, dzięki którym lepiej działają z określonymi typami danych. Nawet drobne poprawki mogą znacząco polepszyć wyniki dla pewnych typów danych.

Algorytmy kompresji stratnej często jako ostatniej fazy używają kompresji bezstratnej. W takim przypadku poprzednie fazy mają za zadanie nie tyle kompresować ile przygotować dane do łatwiejszej kompresji.

Algorytmy kompresji używają modeli [prawdopodobieństwa](#).

Istnieją dwa systemy:

- modele statyczne
- modele adaptywne

Kompresja

Modele statyczne, jeśli nie są znane z góry, są przesyłane przed właściwymi danymi. Koszt przesłania takiego modelu jest bardzo duży i wymusza stosowanie wyłącznie bardzo prostych modeli. To powoduje, że modele statyczne rzadko są stosowane. Kompresory są tutaj zwykle znacznie bardziej złożone niż dekompresory.

Modele adaptywne są tworzone w miarę przetwarzania danych. Kompresor i dekompresor używają tego samego algorytmu do nanoszenia zmian na model w miarę napływu danych. W tym przypadku złożoność kompresorów i dekompresorów jest zwykle, choć nie zawsze, podobna. Wadą modeli adaptywnych jest to, że na początku model ten znacznie odbiega od optymalnego. Jednak możliwość stosowania modeli o dowolnej złożoności, możliwość używania różnych modeli do różnych obszarów kompresowanych danych oraz brak potrzeby przesyłania modelu sprawia, że właściwie całkowicie wyparły one modele statyczne.

1.4. Reprezentacja informacji

Kompresja bezstratna

- Kodowanie Huffmana
- **Kodowanie arytmetyczne**
- Kodowanie Shannona,
- LZ77, LZSS, LZW
- LZ78, LZW, LZMW, LZAP
- LZMA
- PNG
- RLE
- PPM
- Deflate
- BZIP2

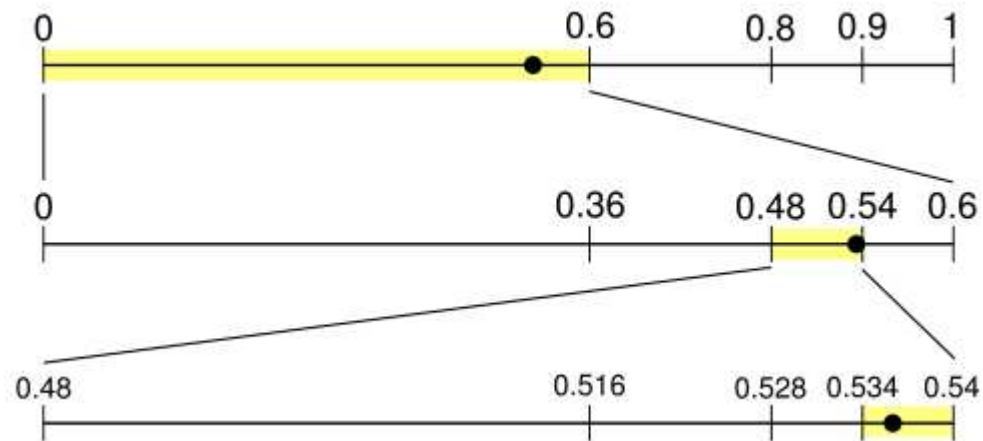
Kompresja arytmetyczna

Dany jest zbiór symboli $S=\{x_1, x_2, x_3, \dots\}$ oraz ich prawdopodobieństwa $p=\{p_1, p_2, p_3, \dots\}$

Przedział $P = [0,1)$, dzielimy jest na podprzedziały o szerokościach równych kolejnym prawdopodobieństwom p czyli $[0, p_1), [p_1, p_1 + p_2), \dots [p_{n-1} + p_n, 1)$

Następnym podprzedziałom odpowiadają wartości ze zbioru S

Przykład: kodowanie znaków z prawdopodobieństwem $p = \{0.6, 0.2, 0.1, 0.1\}$



1.4. Reprezentacja informacji