

Lesson 2

Pengantar Teori Informasi

Oleh :

Ledya Novamizanti

Astri Novianty

Prodi S1 Teknik Telekomunikasi
Fakultas Teknik Elektro
Universitas Telkom

Review: Logaritma

- Jika $\log_a b = c$, maka $a^c = b$
dengan $a > 0, a \neq 1, b > 0$
- a disebut bilangan pokok logaritma atau basis
- b disebut bilangan yang dilogaritmakan
- c disebut hasil logaritma
- untuk basis atau bilangan pokok 10 boleh tidak ditulis.

Review: Logaritma

- Contoh:
 - $\log 100 = ?$
 - $\log_2 16 = ?$
 - $\log_2 (1/8) = ?$

Review: Logaritma

- Beberapa sifat logaritma yang digunakan di dalam bab ini:

1. $\log_a b^n = n \times \log_a b$

sehingga:

$$\log_a \frac{1}{b} = \log_a b^{-1} = -\log_a b$$

2. $\log_a b = \frac{\log_c b}{\log_c a}$

Beberapa Standar log

$$\log_a b = \frac{\log_c b}{\log_c a}$$

■ Contoh: $\log_2 15 = \frac{\log_{10} 15}{\log_{10} 2} = \frac{\log 15}{\log 2}$

- Beberapa nilai log basis 10 yang standar:
- $\log 2 = 0,301$
 - $\log 3 = 0,477$
 - $\log 5 = 0,698$
 - $\log 7 = 0,845$

Teori Informasi

- Dikembangkan oleh Elwood Shannon, *electrical engineer* di Bell Labs, 1940-an.
- Menjelaskan batas komunikasi menggunakan teori probabilitas
- Menentukan parameter kuantitatif dalam mengukur informasi, yang disebut *self-information*.

Pengukuran Informasi secara Kuantitatif

- Menggunakan *self-information*
- Misalkan ada sebuah event A , dengan $P(A)$ adalah peluang terjadinya event A .
- Maka *self-information* pada event A adalah:

$$i(A) = \log_b \frac{1}{P(A)} = -\log_b P(A)$$

Satuan Informasi

$$i(A) = \log_b \frac{1}{P(A)} = -\log_b P(A)$$

- Tergantung pada basis log yang digunakan
- Jika log base = 2, satuannya bits
- Jika log base = e, satuannya nats
- Jika log base = 10, satuannya hartleys

Self-Information...

$$i(A) = \log_b \frac{1}{P(A)} = -\log_b P(A)$$

- Jika tingkat kepastian terjadinya sebuah event itu tinggi, maka jumlah informasi akan rendah
- Jika tingkat kepastian terjadinya sebuah event itu rendah, maka jumlah informasi akan tinggi

Self-Information...

$$i(A) = \log_b \frac{1}{P(A)} = -\log_b P(A)$$

- Jika peluang terjadinya sebuah event itu rendah, lalu kemudian terjadi, maka self information akan tinggi
- Jika peluang terjadinya sebuah event itu tinggi, lalu kemudian terjadi, maka self information akan rendah

Contoh[1]

1. The barking of a dog during a burglary is a high-probability event, therefore does not contain too much information
 - If the dog did not bark during a burglary, this is a low-probability event and contains a lot of much information.

Contoh[2]

2. X mahasiswa cerdas, rajin, IPK tinggi. Y mahasiswa bermasalah, pemalas, IPK rendah.
 - Peluang X lulus Rekrutasi Telkom tinggi, ketika terjadi, tidak ada berita penting tentangnya.
 - Peluang Y lulus Rekrutasi Telkom rendah, ketika ia ternyata lulus, ada banyak informasi tentangnya.

Contoh[3]

- Arsenal winning the FA Cup is hardly news
- "Bristol Rovers win FA Cup" - That's news!

- Sesuatu yang sudah pasti terjadi (probabilitas 1), tidak mengandung informasi
- Sesuatu yang tidak pasti, mengandung informasi
- Semakin kecil peluang terjadinya, semakin banyak informasi yang terkandung
- Contoh:
 - Gosip infotainment..
 - Ketika Mubarak belum dipastikan mundur, berita di mana-mana.

Entropi

- Adalah rata-rata self-information dari sekumpulan event random.
- Jika terdapat sekumpulan event A_i , maka entropi dinyatakan oleh:

$$H = \sum P(A_i) * i(A_i) = - \sum P(A_i) * \log_b P(A_i)$$

Menurut Shannon..

- Entropi menunjukkan rata-rata banyaknya simbol biner yang diperlukan untuk mengkodekan source data
- Jika informasi dalam bits, maka satuan entropi adalah bits/sample atau bits/symbol.
- nilai entropi dari sebuah source data = kemampuan maksimum lossless compression untuk mengkodekan source data tersebut

Contoh Entropi..

- Terdapat *message* (pesan) yang mengandung 4 simbol a_1, a_2, a_3, a_4 dengan probabilitas masing-masing simbol di dalam pesan tersebut:

$$P(a_1) = 1/2, P(a_2) = 1/4, P(a_3) = P(a_4) = 1/8$$

- Berapa entropi dari *message* tersebut? Apa artinya?

Menentukan Entropi dari Source Data

- Tergantung pada model struktur data yang digunakan
- Entropi dapat dikurangi jika digunakan model struktur data yang berbeda
- Semakin kecil entropi sebuah source data, maka tingkat maksimum kompresi yang dapat dicapai pada source data tersebut akan semakin meningkat
- d. k. l source data tersebut memungkinkan untuk dikompresi dengan lebih baik

Contoh[1]

- Terdapat source data sebagai berikut:

1 2 3 2 3 4 5 4 5 6 7 8 9 8 9 10

- Tentukan entropi data tersebut.

Jawaban

- Source data dapat dipandang **setidaknya** oleh 2 pendekatan model struktur data.

Jawaban

- Model 1:

- Kita pandang setiap simbol sebagai data yang saling independen satu sama lain, sehingga source data

1 2 3 2 3 4 5 4 5 6 7 8 9 8 9 10

dipandang mengandung 10 simbol, dengan peluang kemunculan masing-masing diestimasi sebagai berikut:

- $P(1) = P(6) = P(7) = P(10) = 1/16$
- $P(2) = P(3) = P(4) = P(5) = P(8) = P(9) = 2/16$

Jawaban-Lanj.

- Sehingga entropi dapat dihitung sbb:

$$H = - \sum_{i=1}^{10} P(i) * \log_2 P(i)$$

Jawaban-Lanj.

- Diperoleh entropi 3.25 bits, artinya pengkodean yang dilakukan dapat mencapai 3.25 bit per sampel, maka totalnya 16×3.25 bit = 52 bit

Jawaban-Lanj.

- Model 2
 - Lihat aspek korelasi antar sampel yang bertetangga, mulai dari sampel pertama yaitu '1'.
 - Jika data residu dari setiap sampel adalah nilai selisih dengan sampel sebelumnya, maka diperoleh rangkaian data residu sbb:

Data awal : 1 2 3 2 3 4 5 4 5 6 7 8 9 8 9 10

Data residu : 1 1 1 -1 1 1 1 -1 1 1 1 1 1 -1 1 1

Jawaban-Lanj.

- Jadi, representasi source data

1 2 3 2 3 4 5 4 5 6 7 8 9 8 9 10

yang terdiri atas 10 simbol, dapat diubah representasinya ke dalam bentuk model persamaan matematis

$$x_n = x_{n-1} + r_n$$

dengan data residu (r_n):

1 1 1 -1 1 1 1 -1 1 1 1 1 1 -1 1 1

Jawaban-Lanj.

- Maka entropi dapat dihitung dari data residu (r_n):

1 1 1 -1 1 1 1 -1 1 1 1 1 1 -1 1 1

yang terdiri atas 2 simbol (1 dan -1) dengan $P(1) = 13/16$, dan $P(-1) = 3/16$.

Jawaban-Lanj.

- Sehingga diperoleh entropi 0.70 bits/sampel untuk source data yang sama

Tugas-2

- Terdapat source data:

1 2 1 2 3 3 3 3 1 2 3 3 3 3 1 2 3 3 1 2

Tentukan entropi data dengan menggunakan 2 pendekatan!

Petunjuk:

- Pendekatan 1: data dipandang mengandung 3 simbol (1, 2, 3)
- Pendekatan 2: data dipandang mengandung 2 simbol (1 2, 3 3)

Thank you. 😊