1b: Information Theory lecture 1

COMSM0075 Information Processing and Brain

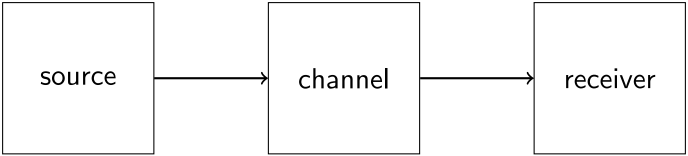
comsm0075.github.io

September 2020

Information Theory

The theory of information is a theory of communication.

a theory of communication

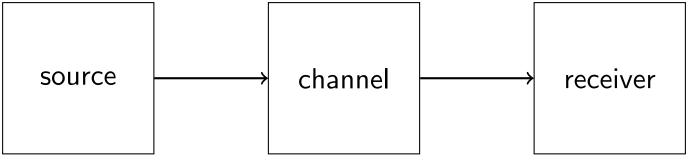


randomness

Image from wikipedia.

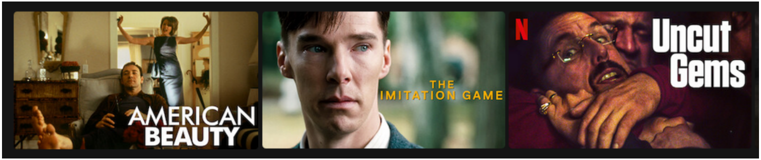
2020

a theory of communication



film recommendations

film recommendations are bad



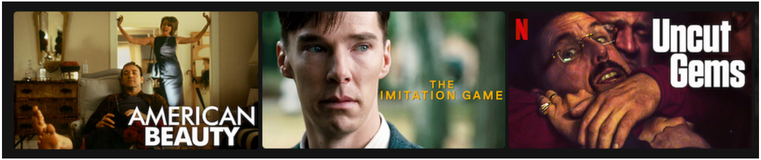
a theory of communication

a theory of communication

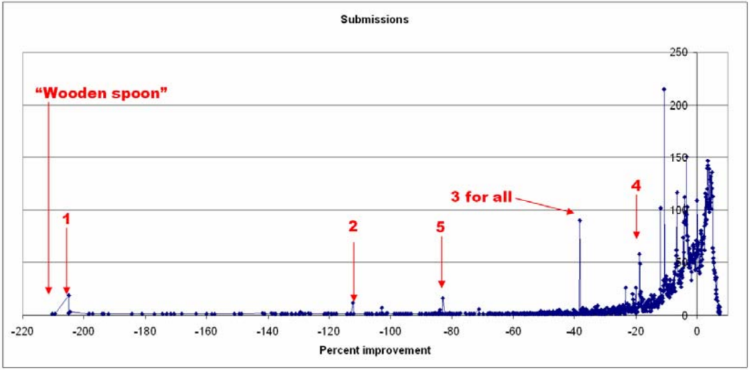
films → 

→ me

film recommendations are bad

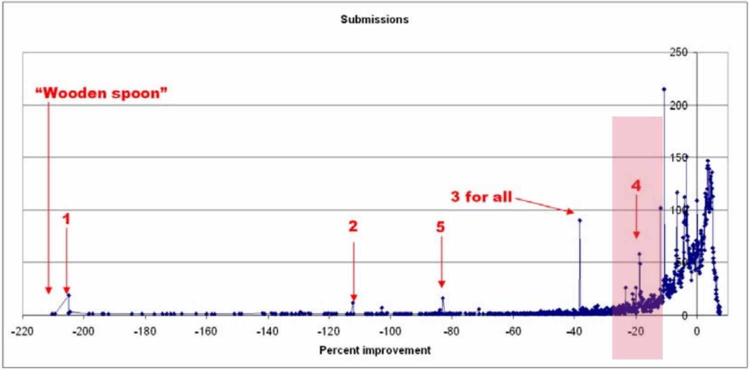


Netflix Prize



Bennett, James, and Stan Lanning. "The netflix prize." Proceedings of KDD cup and workshop. Vol. 2007. 2007.

Netflix Prize



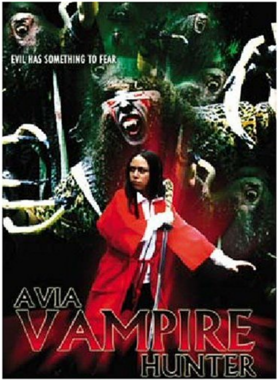
Bennett, James, and Stan Lanning. "The netflix prize." Proceedings of KDD cup and workshop. Vol. 2007. 2007.

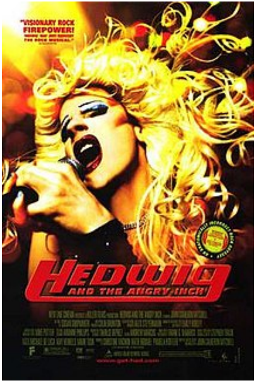
film recommendations

average star ratings

1 star	0.016
2 star	0.310
3 star	0.627
4 star	0.057

the average star ratings mean something

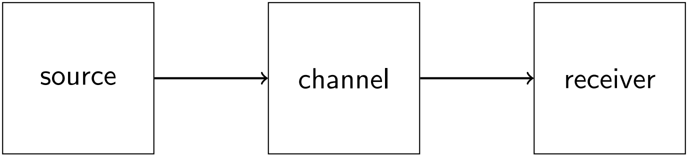




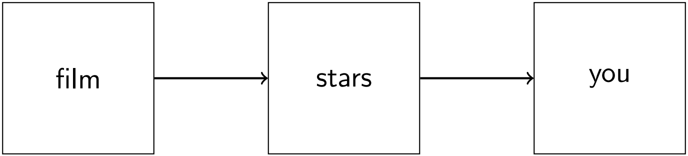
mostly though they tell you it’s an ‘ok’ film

1 star	0.016
2 star	0.310
3 star	**0.627**
4 star	0.057

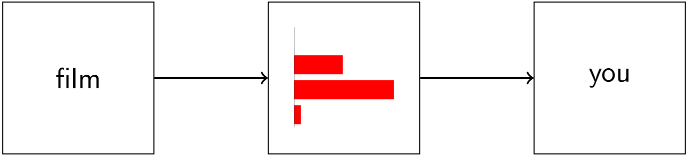
a theory of communication



a theory of communication

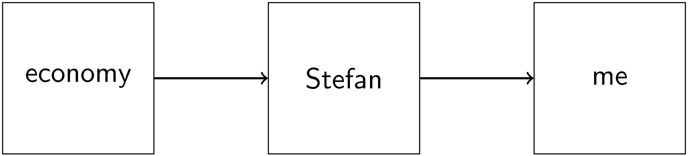


a theory of communication



the fable of Stefan

a theory of communication



a theory of communication

The theory of information starts with an attempt to allow us to quantify the informativeness of information, but not its salience or validity.

Shannon's entropy

For a finite discrete distribution with random variable X, possible outcomes $\{x_1, x_2, \ldots x_n\} \in \mathcal{X}$ and a probability mass function p_X giving probabilities $p_X(x_i)$, the entropy is

$$H(X) = -\sum_{x_i \in \mathcal{X}} p_X(x_i) \log_2 p_X(x_i)$$

example calculation - netflix

1 star	0.016
2 star	0.310
3 star	0.627
4 star	0.057

$$
\begin{aligned}
H(X) = -0.016 \log_2 0.016 - 0.31 \log_2 0.31 \\
- 0.627 \log_2 0.627 - 0.057 \log_2 0.057 \approx 1.28
\end{aligned}
$$

example calculation - netflix

Imagine instead all rankings are equally likely

1 star	0.25
2 star	0.25
3 star	0.25
4 star	0.25

$$H(X) = -4 \times 0.25 \log_2 0.25 = 2$$

example calculation - netflix

Imagine instead everything gets one stars, the Stefan-like case

1 star	1
2 star	0
3 star	0
4 star	0

$$H(X) = -\log_2 1 = 0$$

example calculation - netflix

- deterministic $H(X) = 0$
- actual $H(X) \approx 1.28$
- completely random $H(X) = 2$

0 or 1.28 or 2