A Small Tour of Information Theory

Jinfeng Du

Royal Institute of Technology (KTH), Stockholm, Sweden

§Part I of the lecture given at ACROPOLIS Winter School, Barcelona.

2013-02-14

A theory to **quantify** the processing of information.

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X,Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X,Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X;Y) = I(Y;X)$

the amount of information of X you can get by observing Y

A theory to **quantify** the processing of information.
Entropy and Mutual Information

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.

$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.

$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.

$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.

$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.

$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.

$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.

$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.

$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.

$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.

$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

$H(X)$: the amount of information you get by observing X.
$H(X,Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X,Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X;Y) = I(Y;X)$

the amount of information of X you can get by observing Y

What's information theory

Entropy: a quantitative measure of information/uncertainty

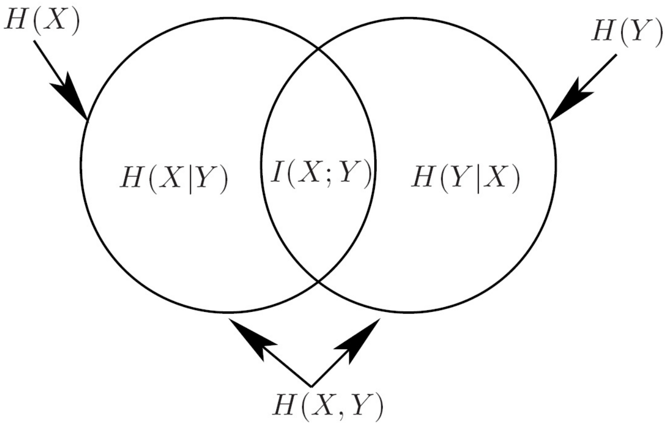
$H(X)$: the amount of information you get by observing X.
$H(X, Y)$: - - - - - - - - - : - - - - - - - - - by observing X and Y.
$H(X|Y)$: - - - - - - - - - : - - - - - - - - - by observing X if you know Y.
$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
$H(X, Y) \leq H(X) + H(Y)$: equality holds iff X and Y are independent.
$H(X|Y) \leq H(X)$: condition reduces entropy!

Mutual information $I(X; Y) = I(Y; X)$

the amount of information of X you can get by observing Y

Relationship among entropy and mutual information

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$$

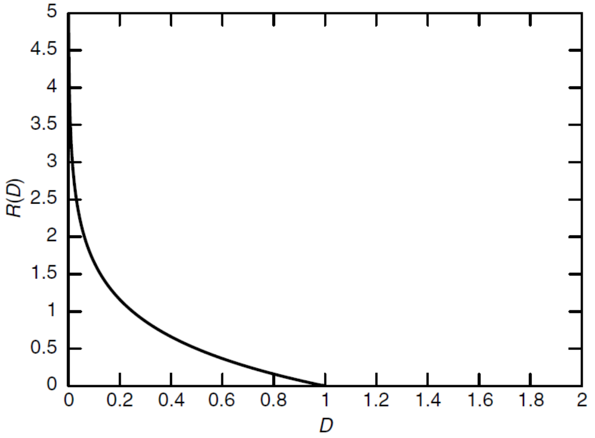


Where's information theory

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding
- Separation theorem II: separate/joint channel-network coding

Where's information theory

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem



[CT'06]

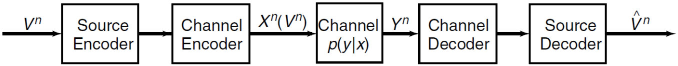
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding

Where's information theory

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
  Examples of Gaussian channel capacity:
  discrete real-valued Gaussian channel: $C = \frac{1}{2}\log(1 + \frac{P}{N})$ [bpcu]
  discrete complex-valued Gaussian channel: $C = \log(1 + \frac{P}{N})$ [bpcu]
  band-limited Gaussian channel: $C = W\log(1 + \frac{P}{WN_0})$ [bps]
- Separation theorem: first source coding, then channel coding
- Separation theorem II: separate/joint channel-network coding

Where's information theory

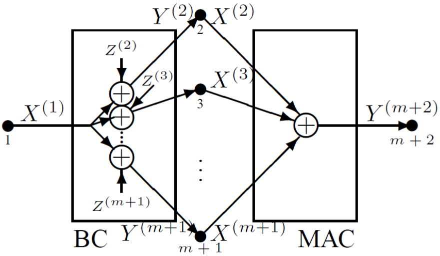
- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding



[CT'06]

- Separation theorem II: separate/joint channel-network coding

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding
- Separation theorem II: separate/joint channel-network coding

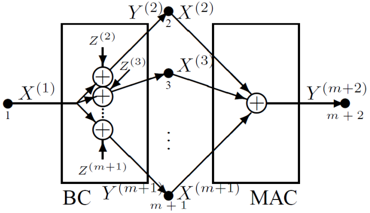


[KEM'10]

TX power P, independent AWGN noise with power N

Where's information theory

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding
- Separation theorem II: separate/joint channel-network coding

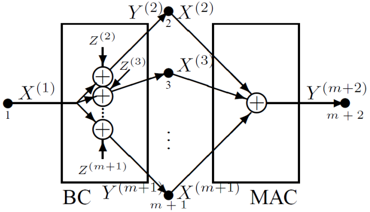


[KEM'10]

Separate channel-network coding: $R = \log(1 + P/N)$

Where's information theory

- Lossless source compression: rate $R > H$
- Lossy source compression: rate-distortion theorem
- Channel coding: $0 < R < C$ is possible (Shannon'48)
- Separation theorem: first source coding, then channel coding
- Separation theorem II: separate/joint channel-network coding



[KEM'10]

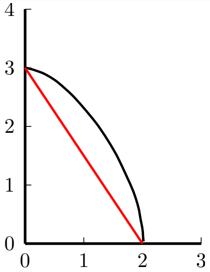
Separate channel-network coding: $R = \log(1 + P/N)$

Joint channel-network coding: $R = \log(1 + mP/N)$

Where's information theory: BC

Two-user broadcast channel: P, $|h_2| > |h_1|$

orthogonal TX: $t \in [0,1]$, $[0,t]x = x_1$; $[t,1]x = x_2$,
$R_1 = t\log(1 + P|h_1|^2)$, $R_2 = (1-t)\log(1 + P|h_2|^2)$
superposition: $x = \sqrt{\alpha}x_1 + \sqrt{1-\alpha}x_2$, $\alpha \in [0,1]$
$R_1 = \log(1 + \frac{\alpha P|h_1|^2}{1+(1-\alpha)P|h_1|^2})$, $R_2 = \log(1 + (1-\alpha)P|h_2|^2)$



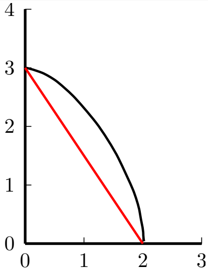
Two-user broadcast channel: P, $|h_2| > |h_1|$

orthogonal TX: $t \in [0,1]$, $[0,t]x = x_1; [t,1]x = x_2$,

$R_1 = t\log(1 + P|h_1|^2)$, $R_2 = (1-t)\log(1+P|h_2|^2)$

superposition: $x = \sqrt{\alpha}x_1 + \sqrt{1-\alpha}x_2$, $\alpha \in [0,1]$

$R_1 = \log(1 + \frac{\alpha P|h_1|^2}{1+(1-\alpha)P|h_1|^2})$, $R_2 = \log(1 + (1-\alpha)P|h_2|^2)$

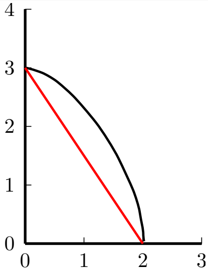


Where's information theory: BC

Two-user broadcast channel: P, $|h_2| > |h_1|$

orthogonal TX: $t \in [0,1]$, $[0,t]x = x_1$; $[t,1]x = x_2$,
$R_1 = t\log(1 + P|h_1|^2)$, $R_2 = (1-t)\log(1 + P|h_2|^2)$

superposition: $x = \sqrt{\alpha}x_1 + \sqrt{1-\alpha}x_2$, $\alpha \in [0,1]$
$R_1 = \log(1 + \frac{\alpha P|h_1|^2}{1+(1-\alpha)P|h_1|^2})$, $R_2 = \log(1 + (1-\alpha)P|h_2|^2)$



Where's information theory: BC

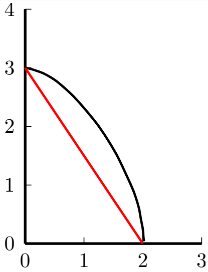
Two-user broadcast channel: P, $|h_2| > |h_1|$

orthogonal TX: $t \in [0,1]$, $[0,t]x = x_1$; $[t,1]x = x_2$,

$R_1 = t\log(1 + P|h_1|^2)$, $R_2 = (1-t)\log(1 + P|h_2|^2)$

superposition: $x = \sqrt{\alpha}x_1 + \sqrt{1-\alpha}x_2$, $\alpha \in [0,1]$

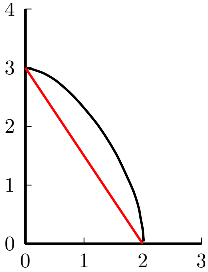
$R_1 = \log(1 + \frac{\alpha P|h_1|^2}{1+(1-\alpha)P|h_1|^2})$, $R_2 = \log(1 + (1-\alpha)P|h_2|^2)$



Where's information theory: BC

Two-user broadcast channel: P, $|h_2| > |h_1|$

orthogonal TX: $t \in [0,1]$, $[0,t]x = x_1$; $[t,1]x = x_2$,
$R_1 = t\log(1+P|h_1|^2)$, $R_2 = (1-t)\log(1+P|h_2|^2)$
superposition: $x = \sqrt{\alpha}x_1 + \sqrt{1-\alpha}x_2$, $\alpha \in [0,1]$
$R_1 = \log(1 + \frac{\alpha P|h_1|^2}{1+(1-\alpha)P|h_1|^2})$, $R_2 = \log(1+(1-\alpha)P|h_2|^2)$



ROYAL INSTITUTE OF TECHNOLOGY

Two-user multiple-access channel: P_1, P_2

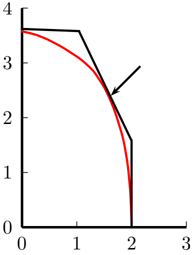
orthogonal TX: $R_1 = t\log(1+P_1/t)$, $R_2 = (1-t)\log(1+P_2/(1-t))$, $t \in [0,1]$

capacity region: $y = x_1 + x_2 + n$, successive interference cancellation

$R_1 = \log(1+\frac{P_1}{1+P_2})$, $R_2 = \log(1+P_2)$ or

$R_1 = \log(1+P_1)$, $R_2 = \log(1+\frac{P_2}{1+P_1})$

Any other points on the boundary can be achieved by time-sharing.



Two-user multiple-access channel: P_1, P_2

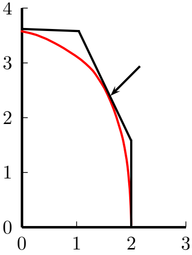
orthogonal TX: $R_1 = t\log(1+P_1/t)$, $R_2 = (1-t)\log(1+P_2/(1-t))$, $t \in [0,1]$

capacity region: $y = x_1 + x_2 + n$, successive interference cancellation

$R_1 = \log(1+\frac{P_1}{1+P_2})$, $R_2 = \log(1+P_2)$ or

$R_1 = \log(1+P_1)$, $R_2 = \log(1+\frac{P_2}{1+P_1})$

Any other points on the boundary can be achieved by time-sharing.



ROYAL INSTITUTE OF TECHNOLOGY

Two-user multiple-access channel: P_1, P_2

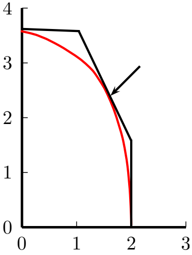
orthogonal TX: $R_1 = t\log(1+P_1/t)$, $R_2 = (1-t)\log(1+P_2/(1-t))$, $t \in [0,1]$

capacity region: $y = x_1 + x_2 + n$, successive interference cancellation

$R_1 = \log(1+\frac{P_1}{1+P_2})$, $R_2 = \log(1+P_2)$ or

$R_1 = \log(1+P_1)$, $R_2 = \log(1+\frac{P_2}{1+P_1})$

Any other points on the boundary can be achieved by time-sharing.



Where's information theory: MAC

Two-user multiple-access channel: P_1, P_2

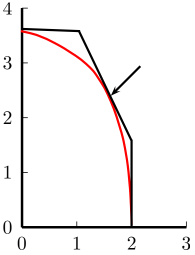
orthogonal TX: $R_1 = t\log(1 + P_1/t)$, $R_2 = (1-t)\log(1 + P_2/(1-t))$, $t \in [0, 1]$

capacity region: $y = x_1 + x_2 + n$, successive interference cancellation

$R_1 = \log(1 + \frac{P_1}{1+P_2})$, $R_2 = \log(1 + P_2)$ or

$R_1 = \log(1 + P_1)$, $R_2 = \log(1 + \frac{P_2}{1+P_1})$

Any other points on the boundary can be achieved by time-sharing.



Where's information theory: MAC

Two-user multiple-access channel: P_1, P_2

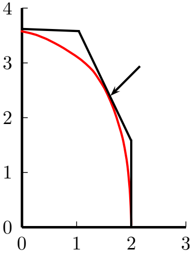
orthogonal TX: $R_1 = t\log(1 + P_1/t)$, $R_2 = (1-t)\log(1 + P_2/(1-t))$, $t \in [0, 1]$

capacity region: $y = x_1 + x_2 + n$, successive interference cancellation

$R_1 = \log(1 + \frac{P_1}{1+P_2})$, $R_2 = \log(1 + P_2)$ or

$R_1 = \log(1 + P_1)$, $R_2 = \log(1 + \frac{P_2}{1+P_1})$

Any other points on the boundary can be achieved by time-sharing.



ROYAL INSTITUTE OF TECHNOLOGY

HARQ: soft combining

repetition redundance: $\log(1 + P_1 + P_2 + \ldots)$

incremental redundance: $\log(1 + P_1) + \log(1 + P_2) + \ldots$

Overheard signal in wireless transmission

interference: has to be mitigated

resource: when cooperation among nodes is allowed

Diversity order, Water-filling, Outage probability, and more ...

Where's information theory: HARQ, and more

KTH VETENSKAP OCH KONST
ROYAL INSTITUTE OF TECHNOLOGY

HARQ: soft combining

repetition redundance: $\log(1 + P_1 + P_2 + ...)$
incremental redundance: $\log(1 + P_1) + \log(1 + P_2) + ...$

Overheard signal in wireless transmission

interference: has to be mitigated
resource: when cooperation among nodes is allowed

Diversity order, Water-filling, Outage probability, and more ...

Where's information theory: HARQ, and more

HARQ: soft combining

repetition redundance: $\log(1 + P_1 + P_2 + ...)$
incremental redundance: $\log(1 + P_1) + \log(1 + P_2) + ...$

Overheard signal in wireless transmission

interference: has to be mitigated
resource: when cooperation among nodes is allowed

Diversity order, Water-filling, Outage probability, and more ...

ROYAL INSTITUTE OF TECHNOLOGY

HARQ: soft combining

repetition redundance: $\log(1 + P_1 + P_2 + ...)$
incremental redundance: $\log(1 + P_1) + \log(1 + P_2) + ...$

Overheard signal in wireless transmission

interference: has to be mitigated
resource: when cooperation among nodes is allowed

Diversity order, Water-filling, Outage probability, and more ...

Where's information theory: HARQ, and more

HARQ: soft combining

repetition redundance: $\log(1 + P_1 + P_2 + \ldots)$
incremental redundance: $\log(1 + P_1) + \log(1 + P_2) + \ldots$

Overheard signal in wireless transmission

interference: has to be mitigated
resource: when cooperation among nodes is allowed

Diversity order, Water-filling, Outage probability, and more ...

Information theory results provide

the performance limit
the existence of schemes to achieve the predicted performance

Insights from information theory results

anything beyond the theoretical limit is impossible
gap from the theoretical limit indicates possibility of improvement
predicted performance can be surprised to up-to-date knowledge
the existence results can guide practical solution design

How to use information theory

Information theory results provide

- the performance limit
- the existence of schemes to achieve the predicted performance

Insights from information theory results

- anything beyond the theoretical limit is impossible
- gap from the theoretical limit indicates possibility of improvement
- predicted performance can be surprised to up-to-date knowledge
- the existence results can guide practical solution design

Want to know more about information theory?

Shannon'48 C. E. Shannon,
"A mathematical theory of communication,"
Bell System Technical Journal,
vol. 27, pp. 379–423 & 623–656, Jul. & Oct. 1948.

CT'06 T. M. Cover and J. A. Thomas,
Elements of Information Theory. New York: Wiley, 2006.

KEM'10 R. Koetter, M. Effros, and M. Médard,
"A theory of network equivalence–part I: point-to-point channels," *IEEE Transactions on Information Theory*,
vol. 57, pp. 972–995, Feb. 2011.