情報理論06

藤田 一寿

津山工業高等専門学校情報工学科 講師
電気通信大学先進理工学科 協力研究員

ver. 20160511

相互情報量

- 事象系AとBが関連していれば、Aが何かを知るとBが何であるかの情報を知ることができる。

- 熱があるかどうかが分かれば、風邪を引いているかどうかの情報が得られる。

  - 熱のあるかどうか分かった状態での条件付きエントロピーは、熱と風邪の結合エントロピーより少なかった。

前回用いた熱と風邪の例

風邪をひいたか、ひかないかに関するエントロピー

$$
\begin{aligned}
H(B) &= -0.65 \log 0.65 - 0.35 \log 0.35 \\
&= 0.93
\end{aligned}
$$

熱の状態を知ったときの風邪かどうかのに関する条件付きエントロピー

$$
\begin{aligned}
H(B|A) &= -\sum_i \sum_j p(A_i) p(B_j|A_i) \log p(B_j|A_i) \\
&= -0.55 \times \log 0.92 - 0.05 \times \log 0.08 \\
&\quad - 0.1 \times \log 0.25 - 0.30 \times \log 0.75 \\
&= 0.57
\end{aligned}
$$

熱があるかどうか分かることによって得られた情報量は

$$
\begin{aligned}
I(A,B) &= H(B) - H(B|A) \\
&= 0.36 \qquad \text{ビット}
\end{aligned}
$$

相互情報量

$$I(A,B) = H(B) - H(B|A)$$

- 相互情報量とは、Aの情報を知ることで得られる、Bに関する情報の量である。
- AとBの関係性の度合いとも言える。
- AとBが無関係（AとBが独立）なら相互情報量は０である。

性質

性質1

$$\begin{aligned} I(A,B) &= H(B) - H(B|A) \\ &= H(A) + H(B) - H(A,B) \\ &= H(A) - H(A|B) \\ &= I(B,A) \end{aligned}$$

相互情報量に順番は関係ない！！

熱と風邪の例では、逆に風邪が分かれば、熱かどうかの情報が0.36ビット得られるということになる。

性質２

$$I(A,B) \le H(A), H(B)$$

定義から明らか

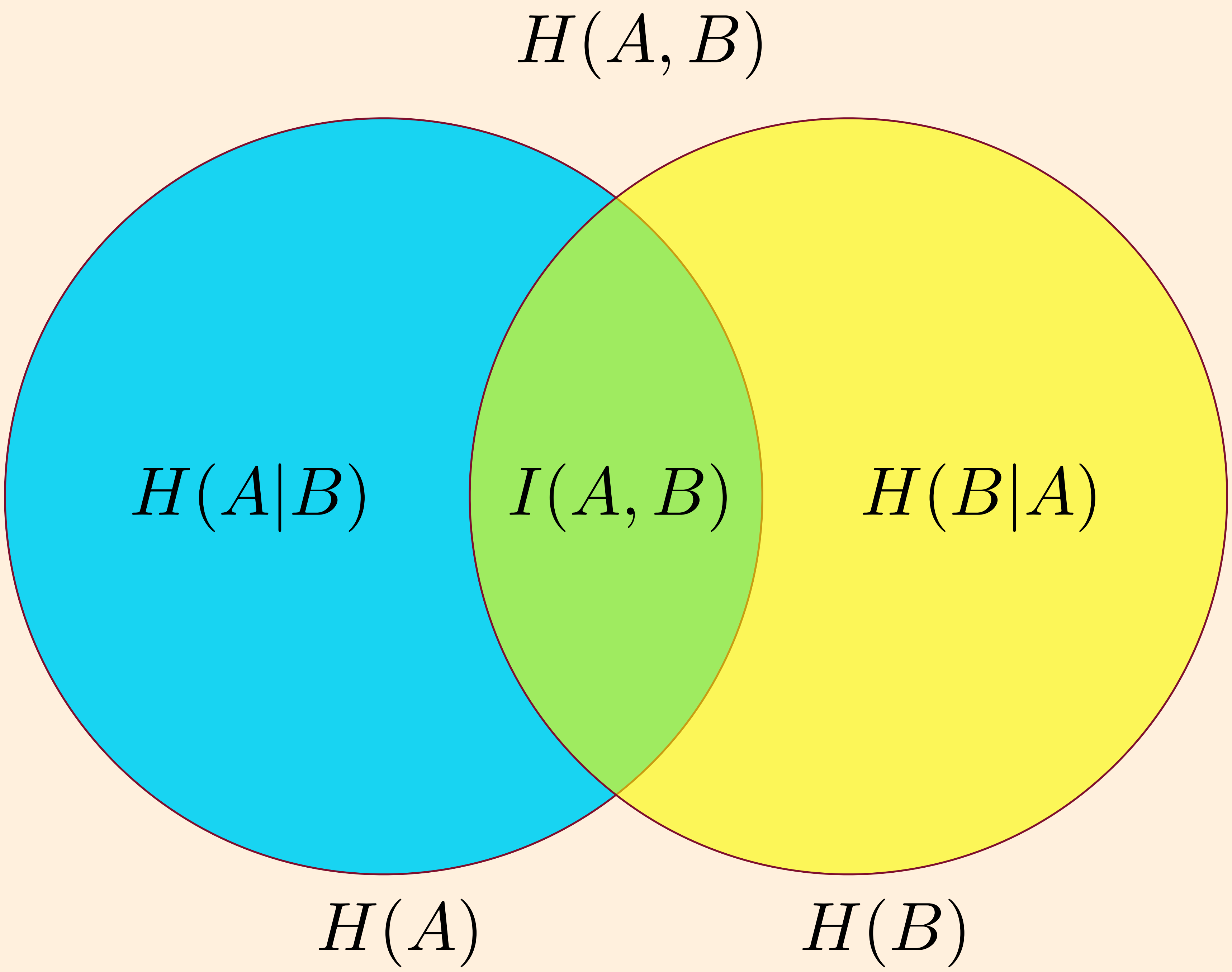
性質３

$$I(A,B) \ge 0$$

それぞれの量の関係

$$
\begin{aligned}
H(A,B) &= H(A) + H(B|A) \\
&= H(B) + H(A|B)
\end{aligned}
$$

$$
\begin{aligned}
I(A,B) &= H(A) - H(A|B) \\
&= H(B) - H(B|A)
\end{aligned}
$$



KLダイバージェンス(情報量)

カルバック-ライブラー(Kulback-Leibler: KL)ダイバージェンス、もしくはKL情報量

$$D(p,q) = \sum_i p_i \log \frac{p_i}{q_i}$$

$D(p,q) \geq 0$　前回証明した

KLダイバージェンスのlogの底はeにします。

KLダイバージェンスの意味

$$D(p,q) = \sum_i p_i \log \frac{p_i}{q_i}$$

$$= \sum_i p_i \log p_i - \sum_i p_i \log q_i$$

$$= \boxed{-\sum_i p_i \log q_i} - \left(\boxed{-\sum_i p_i \log p_i}\right)$$

データが確率分布qから生じたと思って計算したエントロピー（クロスエントロピー）

観測されたデータから求めたエントロピー

想定した確率分布qと実際に観測された確率分布pとの差
砕けた言い方をすると、想像(予想)と現実のギャップ度

逆に考えることもできます

例 1

２人の野球解説者AとBがある球団の勝つ確率をそれぞれ７割、５割と予想した。もし、真の勝つ確率が６割だった場合どちららより正しい予想だといえるか。

	勝つ確率	負ける確率
A氏	0.7	0.3
B氏	0.5	0.5

実際は
勝つ確率0.6
負ける確率0.4

A氏の予想と真の勝敗の割合とのKLダイバージェンス

A氏の予想が想定した確率分布となる。

$$
\begin{aligned}
D_A &= -(0.6 \log 0.7 + 0.4 \log 0.3) + 0.6 \log 0.6 + 0.4 \log 0.4 \\
&= 0.0226
\end{aligned}
$$

B氏の予想と真の勝敗とのKLダイバージェンス

$$
\begin{aligned}
D_B &= -(0.6 \log 0.5 + 0.4 \log 0.5) + 0.6 \log 0.6 + 0.4 \log 0.4 \\
&= 0.0201
\end{aligned}
$$

B氏の予想のほうが比較的あたっている。

例2

サイコロAとBがある。それぞれの目が出る確率がfA、fBであった。どちらのサイコロが理想的なサイコロに近いか。

$$f_A = \{0.20, 0.12, 0.18, 0.12, 0.20, 0.18\}$$
$$f_B = \{0.18, 0.12, 0.14, 0.19, 0.22, 0.15\}$$

理想的なサイコロの確率分布

$$g = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

それぞれのサイコロの確率分布と理想的なサイコロの確率分布のKLダイバージェンスを求める。

D(fA, g) = 0.045

D(fB, g) = 0.041

KLダイバージェンスが小さいほうが、理想と近いと考えられるので、サイコロBの方がより理想的なサイコロといえる。

注意

- 予想と実際の差を計算するので距離と言いたくなる。
    - D(p, q)とD(q, p)は同じ値にならない。
    - 距離の公理に反する。距離ではない。

KLダイバージェンスと相互情報量

I(A, B)をpを使って表すと

$$I(A,B) = \sum_{ij} p(A,B) \log \frac{p(A,B)}{p(A)p(B)}$$

となる。これは、p(A, B)とp(A)p(B)とのKLダイバージェンスとなっている。AとBがi.i.d.の時p(A, B) = p(A)p(B)が成り立つ。つまり、相互情報量は事象AとBが独立に近いかどうかを表す量と言える。

中間試験

- 中間試験は今日までの範囲とする。
- 筆記用具以外の持ち込み可能物品は関数電卓のみとする。
- レポートは中間試験のテスト返却日に提出する。
- 評価は試験70%、レポート30%とする。

次回は情報源をやる予定