به نام خدا

مروری بر مفاهیم تئوری اطلاعات

در این بخش می خواهیم در مورد عملکرد ملدیک کدکننده ی منبع و مباحث مربوط به ظرفیت کانال صحبت کنیم. به همین دلیل نیاز داریم که مفاهیمی از تئوری اطلاعات را در این بخش مطرح کنیم.

به طور کلی تئوری اطلاعات به سوالات اساسی زیر در مخابرات، پاسخ می دهد.

۱- مفهوم اطلاعات چیست و چگونه می توان آن را به صورت کمی اندازه گیری کرد؟

۲- حدنهایی اطلاعاتی که یک منبع به دست می دهد چیست و چگونه اندازه گیری می شود؟

۳- تا چه حدی می‌توانیم اطلاعات یک منبع را فشرده کنیم (اضافات را حذف کنیم) و چگونه؟

۴- حد نهایی اطلاعاتی که روی یک کانال مخابراتی می‌توان ارسال کرد چیست و چگونه می‌توان به آن رسید؟

* مفهوم اطلاعات (Information)

به طور کلی میزان اطلاعات یک پیام (یا گزاره یا پیش‌آمد) با احتمال وقوع آن ارتباط دارد. به عبارت دیگر هرچه احتمال پیش‌آمد کمتر باشد، اطلاعات آن بیشتر است. علاوه بر این به طور شهودی می‌دانیم که اطلاعات مقداری مثبت است. در ادامه به دنبال تابعی هستیم که در مورد هر پیش‌آمد دلخواه A با احتمال P_A، میزان اطلاعات آن را به دست بدهد.

خصوصیات این تابع به صورت زیر است.

۱ـ اطلاعات یک پیشامد همواره مقداری مثبت است.

۲ـ اطلاعات یک پیشامد تابعی نزولی از احتمال آن پیشامد است.

۳ـ اگر دو پیشامد A و B مستقل از هم باشند و داشته باشیم $C = A \cap B$

$A \perp\!\!\!\perp B$,

در این صورت اطلاعات پیشامد C برابر حاصل جمع اطلاعات دو پیشامد A و B است. یعنی

$$\underbrace{I(C)}_{\text{اطلاعات } C} = \overbrace{I(A)}^{\text{اطلاعات } A} + \overbrace{I(B)}^{\text{اطلاعات } B}$$

اثبات می شود که از نظر ریاضی، تنها تابعی که سه خصوصیت بالا را برآورده می کند، تابع لگاریتم است

به عبارت دیگر اگر پیش آمد A با احتمال P_A باشد، اطلاعاتی که این پیش آمد (ابهام یا گزاره) به دست می دهد به صورت زیر قابل محاسبه است.

$$I(A) = \log \frac{1}{P_A} = -\log P_A \qquad (0 \le P_A \le 1)$$

در ادامه می خواهیم در مورد اطلاعات یک منبع صحبت کنیم. همان طور که می دانیم خروجی منابع اطلاعات، می توانند پیوسته یا گسسته باشند. در این درس بر روی منابع گسسته تمرکز می کنیم.

یک منبع اطلاعات گسسته، منبعی است که پیامهایی از یک مجموعه شمارش پذیر گسسته را مانند $\{A_1, A_2, \dots, A_n\}$ را در خروجی ایجاد می‌کند. احتمال وقوع هر یک از پیامهای $A_1, A_2, \dots, A_n$ را با $P_i = P(A_i)$ نمایش می‌دهیم که جزو خصوصیات آماری منبع اطلاعات است.

یکی دیگر از خصوصیات هر منبع اطلاعاتی، نرخ منبع است که برحسب $bit/second$ یا $sym/second$ مشخص می‌شود.

با این دیدگاه می‌توانیم خروجی منبع را با یک متغیر تصادفی X مدل سازی کنیم که در آن

(متغیر تصادفی گسسته)

$$X \in \{A_1, A_2, \dots, A_n\} = \mathcal{X}$$

$$P_r\{X = A_i\} = P(A_i) = P_i$$

$\Leftarrow$ تابع جرم احتمال (pmf) را در مورد متغیر تصادفی X در اختیار داریم.

Probability mass function

به این ترتیب می توان اطلاعات یک منبع را به صورت زیر تعریف کرد

$$\text{اطلاعات منبع} = E\left\{ \text{اطلاعات پیامهای تولید شده توسط منبع} \right\}$$

بنابراین می توان اطلاعات یک منبع که خروجی آن را با متغیر تصادفی X مدل سازی کردیم را به صورت زیر به دست آورد.

$$\underbrace{H(X)}_{\text{اطلاعات منبع یا آنتروپی}} = E\left\{ \overbrace{\log \frac{1}{P_X}}^{I} \right\} = -E\left\{ \log P_X \right\} = E\left\{ I_A \right\}$$

Entropy

$$= \sum_{i=1} P_i \log \frac{1}{P_i} = -\sum_{i=1} P_i \log P_i$$

یادآوری مفهوم امید ریاضی Expectation

برای یک متغیر تصادفی X با تابع چگالی احتمال $f_X(x)$، امید ریاضی هر تابع دلخواه $g(\cdot)$ از متغیر تصادفی X (یعنی $g(X)$) به صورت زیر تعریف می‌شود:

$$E\{g(X)\} = E\,g(X) = \int_{-\infty}^{\infty} \underbrace{g(x)}_{\text{مقادیری که } g(X) \text{ اختیار می‌کند}} \, \underbrace{f_X(x)\,dx}_{\text{معیاری از احتمال}} \equiv$$

معیاری از میانگین مقادیری که $g(X)$ اختیار می‌کند.

برای یک متغیر تصادفی گسسته X با تابع جرم احتمال $P(x)$ ، امید ریاضی $g(X)$
به صورت زیر خواهد بود.

$$E\,g(X) = \sum_{x} g(x)\,p(x)$$

که در آن

$$p(x) = P_r\{X = x\} \quad : \quad \text{pmf}$$

بنابراین درمورد آنتروپی منبع نیز داریم

$$H(X) = E\{I_A\} = E\{\underbrace{\log \frac{1}{P_X}}_{g(X)}\}$$

$$= \sum_{i=1} \underbrace{P_i}_{p(x)} \underbrace{\log \frac{1}{P_i}}_{g(x)} = -\sum_{i=1} P_i \log P_i$$

$X \in \{A_1, \dots, A_n\}$

$P_r\{X = A_i\} = P(A_i) = P_i$: pmf of X

⇐ آنتروپی یک منبع اطلاعات گسسته همواره غیرمنفی است یعنی $H(X) \geq 0$

زیرا $H(X)$ میانگین اطلاعات پیامهای منبع است که مقادیر غیرمنفی دارند. بنابراین حاصل این میانگین نیز غیرمنفی است.

$$0 \leq P_i \leq 1 \quad , \quad \log \frac{1}{P_i} \geq 0$$

$$\Rightarrow E\left\{ \log \frac{1}{P_x} \right\} \geq 0$$

* واحد آنتروپی

واحد آنتروپی بر اساس اینکه لگاریتم را در چه پایه‌ای محاسبه کنیم، متفاوت خواهد بود.

- اگر در محاسبه آنتروپی از تابع لگاریتم با پایه 2 استفاده کنیم، واحد آنتروپی بیت نامیده می‌شود.

bit ≡ binary digit

- اگر در محاسبه آنتروپی از تابع لگاریتم با پایه e استفاده کنیم، واحد آنتروپی نات نامیده می‌شود.

nat ≡ natural digit

این واحدها براساس خصوصیات تابع لگاریتم قابل تبدیل به یکدیگر هستند. می‌دانیم که

$$\log_b^x = \log_b^a \log_a^x \Longrightarrow \underbrace{H_b(X)}_{\text{آنتروپی با } \log_b} = \left(\log_b^a\right)\underbrace{H_a(X)}_{\text{آنتروپی با } \log_a}$$

مثال ۱. منبع باینری

یک منبع باینری به صورت زیر در نظر می‌گیریم.

$$\text{الفبای منبع} = \{0, 1\} = \{A_1, A_2\}$$

$$\text{احتمال هر یک از خروجی‌ها} = \{P, 1-P\} = \{P_1, P_2\}$$

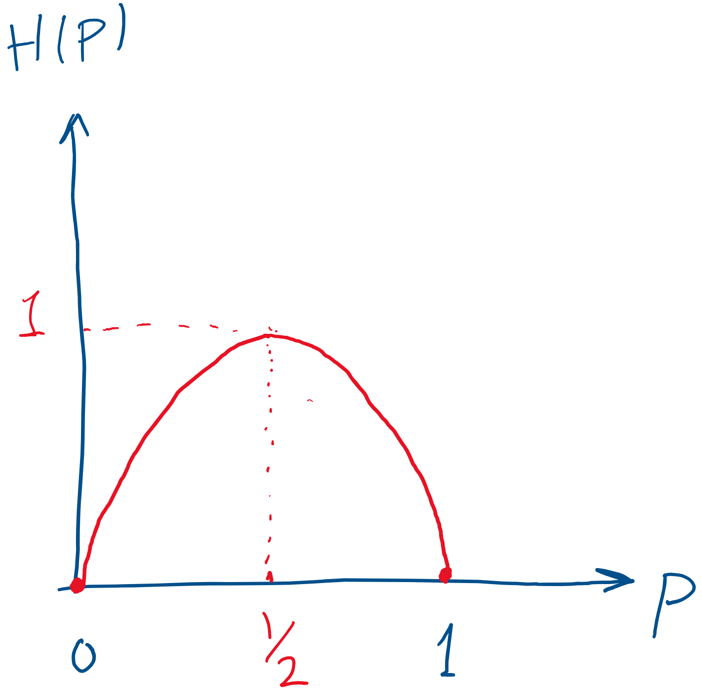
متغیر تصادفی که خروجی منبع را مدل‌سازی می‌کند

$$X \sim \text{Bernolli}(p)$$

$$H(X) = E\left\{\log \frac{1}{P_X}\right\} = -E\left\{\log P_X\right\}$$

$$H(X) = P_1 \log \frac{1}{P_1} + P_2 \log \frac{1}{P_2} = -P_1 \log P_1 - P_2 \log P_2$$

$$H(X) = \underbrace{-p \log p - (1-p) \log (1-p)}_{\triangleq H(p)} \quad \text{bits}$$



همان طور که می‌بینیم وقتی $P=0$ یا $P=1$ برقرار است، آنتروپی یا اطلاعات منبع برابر صفر است. زیرا خروجی منبع در این حالت مشخص است و حالت تصادفی ندارد (ابهامی در خروجی وجود ندارد $\Leftarrow$ اطلاعات برابر صفر است)

* در حالتی که $P=\frac{1}{2}$ برقرار است، بیشترین ابهام را در خروجی منبع داریم و در نتیجه آنتروپی ماکزیمم است.

قضیه (Maximum Entropy)

اگر منبع گسسته X، خروجی هایی از یک مجموعه گسسته $\mathcal{X}$ تولید کند، همواره داریم

$H(X) \leq \log|\mathcal{X}|$ که در آن منظور از $|\mathcal{X}|$، تعداد اعضای مجموعه ی $\mathcal{X}$ است.

تساوی زمانی برقرار است که توزیع متغیر تصادفی X، یکنواخت باشد یا به بیان دیگر سمبل های خروجی منبع، هم احتمال باشند.

$$X \in \mathcal{X} = \{A_1, A_2, \dots, A_n\} \longrightarrow |\mathcal{X}| = n$$

$$H(X) \leq \log|\mathcal{X}|$$

$$P_r\{X = A_i\} = P_i = P(A_i)$$

$H(X)$ زمانی ماکزیمم است که $\{P_i\}_{i=1}^{n}$ ها باهم برابر باشند یا توزیع متغیر تصادفی X یکنواخت باشد یعنی

$$P_1 = P_2 = \dots = P_n = \frac{1}{n}$$

$$\Rightarrow H_{max}(X) = -\sum_{i=1}^{n} P_i \log P_i = \sum_{i=1}^{n} P_i \log \frac{1}{P_i} = n\left(\frac{1}{n} \log n\right)$$

$$\Rightarrow H_{max}(X) = \log n = \log |X|$$

* در مورد منابع اطلاعات پیوسته، این موضوع با در نظر گرفتن محدودیت توان P در خروجی منبع، برای متغیر تصادفی گوسی $N(0, P)$ برقرار است. یعنی با در نظر گرفتن محدودیت توان P در خروجی منبع، منبع گوسی $N(0, P)$ دارای بیشترین آنتروپی است.

مثال 2: منبعی را در نظر بگیرید که یک متغیر تصادفی به صورت زیر تولید می‌کند

$$X = \begin{cases} a & 1/2 \\ b & 1/4 \\ c & 1/8 \\ d & 1/8 \end{cases}$$

آنتروپی منبع را محاسبه کنید.

$$H(X) = \sum_{i=1} p_i \log \frac{1}{p_i} = -\sum_{i=1} p_i \log p_i$$

$$H(X) = \frac{1}{2} \underbrace{\log 2}_{1} + \frac{1}{4} \underbrace{\log 4}_{2} + \frac{1}{8} \underbrace{\log 8}_{3} + \frac{1}{8} \log 8 \qquad \text{bits}$$

$$H(X) = \frac{7}{4} \ \text{bits}$$