یک تعمیم دو پارامتری آنتروپی رنی و میانگین جدید طول کدواژه مرتبط با آن

علیرضا حسین‌زاده[۱]، مهدی یعقوبی اول ریابی*[۲]

چکیده

آنتروپی رنی به عنوان یک تعمیم تک پارامتری برای آنتروپی شانون توسط آلفرد رنی در سال ۱۹۶۱ معرفی شد. این اندازه اطلاع تعمیم یافته خواص مشابهی با آنتروپی شانون دارد.

در مقاله حاضر ابتدا آنتروپی شانون و رنی و برخی ویژگی‌های مهم آن‌ها را مرور می‌کنیم. سپس یک تعمیم دو پارامتری برای آنتروپی رنی با میانگین جدید طول کدواژه مربوطه، که توسط بهات و دیگران (۲۰۲۳) مطرح شده، و همچنین ویژگیهای مهم آن را مورد بررسی قرار می‌دهیم. به ویژه برای مقادیری از پارامترها کارایی کد هافمن، نسبت به این آنتروپی تعمیم یافته و میانگین جدید طول کدواژه نظیر آن، را مورد ارزیابی قرار می‌دهیم. نتایج بیانگر آن است که کارایی کد هافمن با آنتروپی و میانگین طول کد واژه جدید نیز با انتخاب مناسب پارامترها تقریبا بهینه باقی می‌ماند.

.

کلمات کلیدی: آنتروپی شانون، آنتروپی رنی، میانگین طول کدواژه، نامساوی کرافت، کد هافمن.

دریافت مقاله: ۲۳/۱۱/۱۴۰۳

پذیرش مقاله: ۲۶/۱۲/۱۴۰۳

۱-مقدمه

شروع تحقیقات در زمینه نظریه اطلاع اولین بار توسط نایکوئیست [۱] در مقاله‌ای تحت عنوان "عوامل خاصی که سرعت تلگراف را تحت تاثیر قرار می‌دهند" آغاز شد. او در این مقاله چگونگی ارسال پیام‌ها توسط یک کانال تلگراف با ماکزیمم سرعت ممکن ولی بدون دگرشکلی را مورد بررسی قرار داد.

هارتلی [۲] برای اولین بار کوشید تا اندازه اطلاع را به صورت زیر تعریف کند:

فرض کنید هر نماد یک پیام را بتوان به s طریق ممکن انتخاب کرد. اکنون با در نظر گرفتن پیام‌های l نمادی می‌توان s^l پیام متمایز تشخیص داد. اینک مقدار اطلاع این منبع اطلاعاتی به صورت لگاریتم تعداد پیام‌های قابل تشخیص یعنی $\log_D s^l$ تعریف می‌شود.

حال چون $\log_D s^l$ با $l\log_D s$ برابر است، این نکته با درک ذهنی این مطلب که میزان اطلاع پیامی به طول l، l برابر میزان اطلاع پیامی به طول یک است، سازگار می‌باشد. این نکته حضور موجه تابع لگاریتم را در تعریف هارتلی بیان می‌کند. زیرا می‌توان ثابت نمود که تنها تابع حقیقی که در معادله تابعی:

$$f(s^l) = lf(s)$$

صدق می‌کند تابع لگاریتمی است.

توجه داریم که افزایش تعداد نمادها افزایش مقدار اطلاع را تضمین می‌کند که با درک مستقیم تطابق دارد. در رهیافت هارتلی هیچ فرضی در این مورد که امکان دارد s نماد ممکن با شانس‌های نابرابر در پیام ظاهر شوند یا اینکه

* پست الکترونیک نویسنده مسئول:
m_yaghoobiawal@yahoo.com
۱. استادیار، دانشکده مهندسی برق، دانشگاه آزاد اسلامی واحد گناباد
۲. استادیار، دانشکده مهندسی برق، دانشگاه آزاد اسلامی واحد گناباد

ممکن است یک نوع وابستگی بین نمادهای متوالی وجود داشته باشد در نظر گرفته نشده است.

کلود شانون[۳] مقاله‌اش را تحت عنوان "نظریه ریاضی ارتباطات" منتشر نمود. دستاورد بزرگ شانون این است که او نظریه‌های نایکوئیست و هارتلی را توسعه داد و نظریه اطلاع امروزی را با مرتبط ساختن مفهوم اطلاع با عدم قطعیت و با بهره‌گیری از مفاهیم نظریه احتمال پایه‌گذاری نمود. شانون در حالت کلی اندازه اطلاع را بر این اساس که توزیع احتمال نمادها غیر یکنواخت باشند توسعه داد. او در مورد معیار هارتلی پیشنهاد نمود با فرض این که همه نمادها با احتمال برابر رخ دهند آنگاه این معیار را نیز می‌توان به عنوان یک اندازه اطلاع تفسیر نمود. با این توصیف می‌توان اندازه اطلاع هاتلی را به عنوان حالتی خاص از اندازه اطلاع شانون در نظر گرفت.

یکی از زمینه‌های کاربردی نظریه اطلاع مطالعه قوانین علمی حاکم بر هر گونه انتقال و تغییر حالت اطلاعات نظیر فشرده‌سازی، کدینگ، ظرفیت کانال و غیره است. در این عرصه نظریه اطلاع به دو سوال اساسی زیر در علوم ارتباطات پاسخ می‌دهد.

اول اینکه حد نهایی فشرده‌سازی داده‌ها چقدر است، که پاسخ آن آنتروپی شانون است.

دوم اینکه حد نهایی نرخ انتقال اطلاعات چه اندازه می‌تواند باشد، که پاسخ آن ظر فیت کانال ارتباطی است.

از این رو در ابتدا چنین به نظر می‌رسید که نظریه اطلاع شاخه‌ای از نظریه ارتباطات است. در حقیقت دامنه کاربردهای نظریه اطلاع به مراتب فراتر از نظریه ارتباطات است و زمینه‌های تحقیقاتی مشترکی با برخی علوم دیگر نیز دارد. در یک دورنمای کلی می‌توان گفت که نظریه اطلاع نقش اساسی در عرصه‌های فیزیک آماری (تبیین قوانین ترمودینامیک)، علوم کامپیوتر (نظریه اطلاع الگوریتمی و پیچیدگی کولموگروف)، آمار (اطلاع فیشر و آزمون فرض)، نظریه ارتباطات (فشرده‌سازی اطلاعات، کدینگ و ظرفیت کانال‌های ارتباطی)، اقتصاد (نرخ رشد سرمایه و نرخ آنتروپی سرمایه در یک بازار سهام) ایفا می‌کند.

برای ملاحظه جزئیات بیشتر وابستگی میان نظریه اطلاع و سایر علوم، خواننده را به مطالعه کتاب ارزشمند و غنی کاور و توماس [۴] توصیه می‌کنیم.

در ادامه آنتروپی شانون و رنی و برخی ویژگی‌های مهم آن‌ها را مرور می‌کنیم. سپس با الهام از بهات و دیگران (۲۰۲۳) یک تعمیم دو پارامتری برای آنتروپی رنی با میانگین جدید طول کدواژه مربوطه و همچنین ویژگیهای مهم آن را مورد بررسی قرار می‌دهیم.

۲-آنتروپی شانون

منطقی به نظر می‌رسد که بگوییم میزان تعجّب حاصل از وقوع یک پیشامد به احتمال رخداد آن پیشامد بستگی دارد. به عنوان مثال در آزمایش تصادفی پرتاب سه تاس اگر بشنویم مجموع اعداد ظاهر شده زوج است زیاد تعجّب نمی‌کنیم، زیرا احتمال وقوع این پیشامد برابر $\frac{1}{2}$ است. در حالی که اگر بشنویم مجموع اعداد ظاهر شده در پرتاب سه تاس ۱۸ است بیشتر متعجّب می‌شویم زیرا احتمال وقوع این پیشامد برابر $\frac{1}{216}$ است. اینک سعی می‌کنیم میزان تعجّب را کمّی نماییم. ابتدا روی این اصل توافق می‌کنیم که میزان تعجّب از وقوع یک پیشامد به احتمال رخداد آن بستگی دارد. بنابراین اگر p احتمال رخداد آن پیشامد باشد آنگاه میزان تعجّب از وقوع آن پیشامد تابعی از p مانند $S(p)$ خواهد بود .

حال سعی می‌کنیم شکل تابعی $S(p)$ را بر اساس توافق روی مجموعه‌ای از شرایط منطقی تعیین نموده و سپس ثابت کنیم که این اصول برای این که $S(p)$ شکل معینی داشته باشد لازم هستند.

فرض می‌کنیم $S(p)$ برای مقادیر $0 < p \leq 1$ تعریف شده و برای پیشامدهایی که احتمال رخداد آنها صفر است، $S(p)$ تعریف نشده باشد.

اولین شرط بیان این واقعیت است که میزان تعجّب ما از وقوع یک پیشامد قطعی (پیشامدی با احتمال رخداد ۱) برابر صفر است. یعنی :

اصل اول : $S(1) = 0$.

دومین شرط بیان می‌کند که میزان تعجّب از رخداد پیشامدی که شانس وقوع کمتری دارد بیشتر از میزان تعجّب برای رخداد پیشامدی با شانس وقوع بیشتر است. به عبارت دیگر:

اصل دوم : $S(p)$ تابعی اکیدا نزولی از p است. یعنی اگر $p < q$ آنگاه $S(p) > S(q)$.

شرط سوم یک ویژگی ریاضی تابع $S(p)$ است به گونه‌ای که انتظار می‌رود هر تغییر کوچک در مقدار p باعث ایجاد تغییر کوچکی در مقدار $S(p)$ خواهد بود. بنابراین:

اصل سوم : $S(p)$ تابعی پیوسته از p است.

برای بیان آخرین شرط دو پیشامد مستقل A و B را به گونه‌ای که $P(A) = p$ و $P(B) = q$ در نظر می‌گیریم. حال چون پیشامدهای A و B مستقل هستند لذا داریم، $P(A \cap B) = pq$. بنابراین میزان تعجّب حاصل از وقوع هم زمان A و B برابر $S(pq)$ خواهد بود. حال فرض کنید ابتدا مطلع شویم که A اتفاق افتاده و پس از آن پیشامد B نیز رخ داده است. چون $S(p)$ میزان تعجّب وقوع پیشامد A است بنابراین $S(pq) - S(p)$ نشان دهنده افزایش تعجّب است وقتی مطلع شویم که B نیز اتفاق افتاده است. بعلاوه چون A و B مستقل از یکدیگر و اطلاع از وقوع یا عدم وقوع پیشامد A تاثیری در احتمال وقوع پیشامد B ندارد بنابراین میزان افزایش تعجّب باید دقیقا برابر با $S(q)$ باشد. یعنی :

اصل چهارم : برای $0 < q \leq 1$ و $0 < p \leq 1$ داریم:

$$S(pq) = S(p) + S(q).$$

حال می‌توانیم قضیه زیر را که بیان کننده ساختار تابع $S(p)$ است بیان نماییم.

قضیه۱: اگر تابع $S(p)$ در اصول چهار گانه فوق صدق نماید آنگاه داریم:

$$S(p) = -c \log_D p.$$

که در آن c یک ثابت دلخواه صحیح و مثبت است و معمولا آن را برابر یک اختیار می‌کنند.

بسته به این که مبنای لگاریتم اعداد ۲ یا e و یا ۱۰ باشد واحد $S(p)$ که آن را آنتروپی پیشامد A می‌نامیم بیت، نت و یا دک خواهد بود. مرسوم است که مبنای لگاریتم را عدد ۲ در نظر می‌گیرند که در این حالت واحد اطلاع بر حسب بیت خواهد بود. $S(p)$ را اندازه اطلاع از وقوع پیشامد A یا اندازه عدم قطعیت A نیز می‌گویند.

بر این اساس شانون در سال ۱۹۴۸ آنتروپی متغیر تصادفی گسسته X را به صورت زیر تعریف نمود.

تعریف۱: فرض کنید متغیر تصادفی گسسته X مقادیر $\chi = \{x_1, x_2, \dots, x_n\}$ را اختیار نماید. اگر توزیع احتمال X به صورت $P = \{p_1, p_2, \dots, p_n\}$ باشد، آنگاه آنتروپی شانون متغیر تصادفی X را با نماد $H(P)$ یا $H(X)$ نشان می‌دهند و به صورت زیر تعریف می‌کنند:

$$H(P) = -\sum_{i=1}^{n} p_i \log_D p_i .$$

$H(P)$، میانگین تعجّب حاصل از وقوع پیشامد $\{X = x\}$ برای مقادیر مختلف $x \in \chi$ را ارزیابی می‌کند. $H(P)$ را متوسط اطلاع حاصله برای مقادیر مشاهده شده متغیر تصادفی X نیز می‌نامند. در آنتروپی شانون معمولا مبنای لگاریتم را عدد ۲ در نظر گرفته و از نوشتن آن صرف نظر می‌کنند. قضیه زیر خواص آنتروپی شانون را بیان می‌کند.

قضیه۲: آنتروپی شانون $H(P)$ دارای خواص زیر است:

الف) $H(P) \geq 0$. تساوی برقرار است اگر و فقط اگر X یک متغیر تصادفی تباهیده باشد.

ب) $H(P) \leq \log n$ ، که در آن n تعداد اعضاء مجموعه متناهی χ است. تساوی برقرار است اگر و فقط اگر متغیر تصادفی X دارای توزیع یکنواخت $p(x) = \frac{1}{n}$ بر تکیه‌گاه متناهی χ باشد.

قسمت دوم قضیه بالا بیان می‌کند که بیشترین میزان عدم قطعیت زمانی حاصل می‌شود که به تمامی برآمدهای متغیر تصادفی گسسته X با تکیه‌گاه متناهی، شانس‌های مساوی

نسبت داده شود. بعبارت دیگر توزیع احتمال متغیر تصادفی یک توزیع یکنواخت باشد.

۳- فشرده‌سازی و کدگذاری

در عصر حاضر بکارگیری روزافزون داده‌ها و نیاز به ذخیره سازی، انتقال و پردازش آن‌ها ما را به استفاده از روش‌های فشرده‌سازی داده‌ها هدایت می‌کند.

در علوم کامپیوتر و نظریه اطلاعات، فشرده‌سازی به معنای کدگذاری اطلاعات به نحوی است که تعداد بیت‌های کمتری نسبت به نسخه‌ی اصلی آن داشته باشد.

کدگذاری در بخش‌های مختلف مخابره یک پیام، کاربرد دارد. در بخش فرستنده کدگذاری به منظور بیان پیام ارسالی با کمترین تعداد بیت، جهت کاهش پهنای باند مورد نیاز برای ارسال پیام، مورد توجه است. در بخش کانال به منظور مقابله با انواع منابع نوفه و مقابله با تغییرات پیام در هنگام ارسال و کاهش احتمال خطا، می‌توان از کدگذاری استفاده نمود. این موارد به ترتیب با نام‌های کدگذاری منبع و کدگذاری کانال شناخته می‌شوند.

برای اولین بار در سال ۱۹۵۰ ریچارد همینگ [۵] نظریه کدگذاری خود را پایه‌گذاری کرد. هم زمان با او، کلود شانون نیز نظریه کدگذاری بدون نوفه و پس از آن مفهوم آنتروپی و نظریه ریاضی ارتباطات را ارائه نمود.

در سال ۱۹۵۲ هافمن [۶] با استفاده از درخت دودویی مرتب شده بر حسب تکرار و بر اساس مفهوم آنتروپی تعریف شده توسط شانون، توانست یک روش دیگر کدگذاری را ابداع نماید. این کد تنها کد بهینه مبتنی بر آنتروپی شانون است.

کدهای فانو، شانون، هافمن، مورس، الفبایی،حسابی از جمله روش‌های کدگذاری منبع هستند که مورد استفاده قرار می گیرند.

منبع اطلاع گسسته منبعی است که دنباله‌ای از نمادها را که گاهی اوقات حروف نیز نامیده می‌شوند، تولید می‌کند. مجموعه نمادهای ممکن را الفبای منبع می‌نامند. حال فرض کنید $U = \{u_1,u_2,\ldots,u_n\}$ الفبای منبع باشد و این نمادها به طور تصادفی هر یک با احتمال معینی بر اساس توزیع احتمال $P = \{p_1,p_2, \ldots,p_n\}$ رخ دهند. این احتمالات با گذشت زمان تغییر نمی‌کنند. به عبارت دیگر دنباله‌ای تصادفی از نمادها از یک توزیع احتمال مانا تولید می‌شوند. علاوه بر این منبع اطلاع مورد نظر بدون حافظه است. یعنی نمادهای تولید شده به طور احتمالی مستقلند.

برای کدگذاری منبع فوق، فرض کنید الفبای کد با مجموعه $S = \{s_1,s_2,\ldots,s_D\}$ مشخص شده باشد. به هر عنصر الفبای منبع ترکیب معینی از عناصر S نسبت می‌دهند و آن را کدواژه می‌نامند.

اگر همه کدواژه‌ها متفاوت باشند آن را کد ناویژه می‌نامند.

اگر هر دنباله کدگذاری شده وقتی رمزگشایی می‌شود یک دنباله ممکن از الفبای منبع را به دست دهد در این صورت کد مربوطه را یکتا گشودنی می‌نامند.

یک کد را فوری (با شرط پیشوندی) گوییم اگر هیچ کدواژه‌ای پیشوند هیچ کدواژه دیگر نباشد. یک کد فوری می‌تواند بدون مراجعه به کدواژه بعدی کدگشایی شود.

قضیه زیر موسوم به "نامساوی کرافت" شرط لازم و کافی را برای وجود یک کد فوری بیان می‌کند.

قضیه۳ (نامساوی کرافت): شرط لازم و کافی برای وجود یک کد فوری آن است که :

$$\sum_{i=1}^{n} D^{-l_i} \leq 1.$$

که در آن D تعداد اعضای الفبای کد و l_i طول کدواژه نظیر u_i است.

قضیه زیر موسوم به "قضیه کدگذاری منبع"، رابطه بین متوسط طول کدواژه و آنتروپی منبع اطلاع را بیان می‌کند.

قضیه۴ (قضیه کدگذاری منبع): فرض کنید الفبای منبع $U = \{u_1,u_2,\ldots,u_n\}$ دارای توزیع احتمال:

$$P = \{p_1,p_2, \ldots,p_n\}.$$

باشد و الفبای کد، مجموعه D عضوی $S = \{s_1,s_2,\ldots,s_D\}$ باشد. اگر نامساوی کرافت برقرار باشد آنگاه داریم:

$$\frac{H(P)}{\log D} \leq L.$$

که در آن $L = \sum_{i=1}^{n} p_i l_i$ متوسط طول کدواژه‌ها و l_i طول کدواژه نظیر u_i است.

نامساوی فوق به تساوی تبدیل می‌شود اگر و فقط اگر برای هر $i = 1.2 \dots . n$ داشته باشیم $p_i = D^{-l_i}$.

قضیه بالا بیان می‌کند که متوسط طول کدواژه‌ها هرگز نمی‌تواند از آنتروپی منبع اطلاع در مبنای D کمتر باشد. حال اگر طول کدواژه‌ها به گونه‌ای انتخاب شوند که نامساوی بالا به تساوی تبدیل شود آنگاه متوسط طول کدواژه‌ها کمترین مقدار خود را اختیار خواهد نمود و این در صورتی است که برای هر $i = 1.2 \dots . n$ داشته باشیم، $p_i = D^{-l_i}$ و یا به طور معادل $l_i = -\log_D p_i$.

از قضیه کدگذاری منبع می‌توان برای ساختن معیاری برای ارزیابی کارایی یک کد استفاده کرد. بر این اساس، برای یک کد مفروض هر چه نسبت $\eta = \frac{H(P)}{L \log D}$ موسوم به "کارایی کد"به عدد یک نزدیک‌تر باشد کد کاراتری خواهیم داشت. برای درک بهتر این موضوع بهینگی کد هافمن را در مثال زیر بررسی می‌کنیم.

مثال۱ : فرض کنید ۸ اسب با شماره‌های ۱ تا ۸ که شانس برنده شدن آن‌ها متفاوت است در یک مسابقه اسب دوانی شرکت می‌کنند. متغیر تصادفی X را شماره اسب برنده در نظر می‌گیریم که دارای توزیع احتمال جدول (۱) است.

جدول ۱ : توزیع احتمال الفبای منبع

x	۱	۲	۳	۴	۵	۶	۷	۸
$p(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$

با پاره‌ای محاسبات جبری ساده، آنتروپی شانون متغیر تصادفی X برابر ۲ بیت به دست می‌آید.

حال فرض کنید بخواهیم پیامی را ارسال نماییم که معرف شماره اسب برنده شده باشد. یک روش می‌تواند ارسال شماره اسب برنده شده، بدون در نظر گرفتن احتمال برنده شدن آن‌ها باشد. در این وضعیت شانس برنده شدن همه اسب‌ها را یکسان و مساوی $\frac{1}{8}$ در نظر می‌گیریم. بنابراین ارسال شماره اسب برنده، متضمّن ارسال ۳ بیت اطلاع خواهد بود. در روش دوم با در نظر گرفتن توزیع احتمال X، شماره اسب‌ها را بگونه‌ای کدگذاری می‌کنیم که طول کد کمتر به اسب با شانس برنده شدن بیشتر اختصاص یابد. این استراتژی کدگذاری منطبق بر کد هافمن است. جدول (۲) کدگذاری هافمن را برای الفبای منبع جدول (۱) مشخص می‌کند. در این جدول x شماره اسب برنده، $c(x)$ کدواژه مربوط به x و $l(x)$ طول کدواژه است.

جدول ۲ : کدگذاری هافمن و طول کدواژه‌ها

x	$c(x)$	$l(x)$
۱	۰	۱
۲	۱۰	۲
۳	۱۱۰	۳
۴	۱۱۱۰	۴
۵	۱۱۱۱۰۰	۶
۶	۱۱۱۱۰۱	۶
۷	۱۱۱۱۱۰	۶
۸	۱۱۱۱۱۱	۶

در این صورت امید ریاضی(متوسط) طول کدواژه‌ها برابر است با:

$$L = \sum_{i-1}^{8} p_i l_i = \sum_{x=1}^{8} p(x) l(x) = 2,$$

که با آنتروپی متغیر تصادفی X برابر است.

توجه کنید که در این مثال الفبای منبع شماره اسب‌ها و الفبای کد { ۰,۱} است و کارایی کد هافمن برابر یک است. لذا مثال فوق بهینگی کد هافمن را تایید می‌کند.

۴- آنتروپی رنی[1]

تا کنون تعمیم‌های متعددی برای آنتروپی شانون توسط محققین ارائه گردیده است. هر کدام از این تعمیم‌ها دارای خواصی هستند که آنها را به کاربردهای وسیعی در زمینه‌های مختلف هدایت می‌کند.

[1] Renyi's Entropy

آنتروپی رنی به عنوان تعمیمی برای آنتروپی شانون توسط آلفرد رنی [۷] معرفی شد. انگیزه اصلی تعمیم رنی این ایده بود که هر انتخاب مناسب برای آنتروپی باید یک میانگین باشد. از این رو او از این ایده ریاضی استفاده کرد که میانگین‌های خطی، تنها راه ممکن محاسبه میانگین نیستند و می‌توان میانگین را نسبت به هر تابع دلخواه دیگری نیز تعریف نمود. در نظریه عمومی میانگین‌ها، میانگین شبه‌خطی[1] متغیر تصادفی X نسبت به توزیع احتمال $P = \{p_1, p_2, \dots, p_n\}$ به صورت زیر تعریف می‌شود.

$$E_\varphi(X) = \langle X \rangle_\varphi = \varphi^{-1}\left(\sum_{i=1}^{n} p_i \varphi(x_i)\right),$$

که در آن φ تابعی پیوسته و اکیداً یکنوا است. در مقالات مربوط به میانگین‌های تعمیم یافته، تابع φ به تابع کولموگروف- ناگومو[2] (KN) موسوم است.

در حالت خاص $\varphi(x) = x$ میانگین شبه‌خطی به میانگین خطی $E(X) = \sum_{i=1}^{n} p_i x_i$ تبدیل می‌شود.

برای $\varphi(x) = x^\alpha$، $\alpha > 0$، میانگین توانی با نمای α و برای $\varphi(x) = \ln x$، با توزیع احتمال یکنواخت، میانگین هندسی حاصل می‌شود.

قضیه ۵ : فرض کنید φ یک تابع KN و c یک ثابت حقیقی باشد. آنگاه شرط لازم و کافی برای آن که:

$$\langle X + c \rangle_\varphi = \langle X \rangle_\varphi + c,$$

آن است که φ یک تابع خطی یا نمایی باشد.

حال اگر در تعریف آنتروپی شانون به جای میانگین استاندارد $H(P) = -\sum_{i=1}^{n} p_i \log_D p_i$، از میانگین شبه خطی $E_\varphi(X)$ استفاده نماییم در این صورت یک اندازه اطلاع تعمیم یافته نسبت به تابع φ به صورت زیر به دست خواهد آمد:

$$S_\varphi(P) = \varphi^{-1}\left(\sum_{I=1}^{n} p_i \varphi(-\log_D p_i)\right).$$

$S_\varphi(P)$ می‌تواند به عنوان یک آنتروپی شبه‌خطی مورد توجه قرار گیرد.

حال یک سوال طبیعی که ممکن است مطرح شود این است که، عمومی‌ترین کلاس توابع φ که برای آنها آنتروپی شبه خطی فوق یک اندازه اطلاع جمع‌پذیر را فراهم آورد چه کلاسی است؟ پاسخ این سؤال با توجه به قضیه (۵) تنها دو کلاس خطی و نمایی است. کلاس خطی منجر به آنتروپی شانون می‌شود و کلاس نمایی به فرم، $\varphi(x) = D^{(1-\alpha)x}$، $\alpha > 0$، $\alpha \neq 1$، خانواده‌ای تک پارامتری از اندازه‌های اطلاع به فرم :

$$R^\alpha(P) = \frac{1}{1-\alpha}\log_D\left(\sum_{i=1}^{n} {p_i}^\alpha\right),\ \alpha>۰,\ \ \alpha\neq۱,$$

را به دست می‌دهد که توسط رنی معرفی گردید.

از آنجایی که وقتی $\alpha \longrightarrow 1$، داریم، $R^\alpha(P) \longrightarrow H(P)$ لذا این اندازه اطلاع، یک تعمیم تک پارامتری برای اندازه اطلاع شانون به شمار می‌رود.

در قضیه زیر مشابه آنتروپی شانون، خواصی برای آنتروپی رنی بیان شده است.

قضیه ۶ : آنتروپی رنی دارای خواص زیر است:

الف) $R^\alpha(P) \geq 0$.

ب) برای هر $0 < \alpha < 1$، $R^\alpha(P)$ نسبت به توزیع احتمال $P = \{p_1, p_2, \dots, p_n\}$ تابعی اکیداً مقعر است.

ج) $R^\alpha(P)$ تابعی نزولی از α است.

د) برای هر $0 < \alpha < 1$ داریم:

$$H(P) < R^\alpha(P) < \log_D n.$$

همچنین برای هر $\alpha > 1$ داریم:

$$R^\alpha(P) < H(P) < \log_D n.$$

هـ) برای متغیرهای تصادفی مستقل X و Y با توزیع‌های احتمال $P = \{p_1, p_2, \dots, p_n\}$ و $Q = \{q_1, q_2, \dots, q_n\}$ داریم:

$$R^\alpha(P,Q) = R^\alpha(P) + R^\alpha(Q)$$

این تساوی به ویژگی جمع پذیری آنتروپی رنی موسوم است که در مورد آنتروپی شانون نیز برقرار است.

۵- تعمیم آنتروپی رنی

پس از رنی با تلاش‌های محققین دیگری از جمله هاوردا و چاروات [۸]، شارما و میتال [۹]، بهات و دیگران[۱۰]، اندازه‌های اطلاع تعمیم‌یافته مختلفی به عرصه نظریه اطلاع

[1] Quasilinear mean

[2] Kolmogorov-Nagumo

وارد شده و کاربردهای متعدد آنها در مقالات مختلف علمی مورد بررسی قرار گرفته است.

اخیرا در مقاله‌ای بهات و دیگران یک تعمیم دو پارامتری جدید برای آنتروپی رنی معرفی نموده‌اند. بعلاوه آن‌ها یک میانگین جدید برای طول کدواژه‌ها معرفی و از جنبه قضایای کدگذاری منبع مورد بررسی قرار داده‌اند.

تعریف زیر توسط بهات و دیگران به منظور تعمیم آنتروپی رنی مطرح شده است.

تعریف۲: آنتروپی رنی دو پارامتری متغیر تصادفی گسسته X را با نماد $R_\beta^\alpha(P)$ نشان می‌دهند و به صورت زیر تعریف می‌کنند:

$$R_\beta^\alpha(P) = \frac{\beta}{\beta-\alpha}\log_D\left(\sum_{i=1}^{n} p_i^{\frac{\alpha}{\beta}}\right),\ \alpha \neq \beta\ \ \alpha,\beta > 0\,.$$

وجود پارامترهای α و β را از نقطه نظر کاربردی می‌توان به صورت زیر تفسیر نمود.

فرض کنید $[x_i,p_i]$ یک سیستم سایبرنتیک باشد به گونه‌ای که x_iها پیشامدها و p_iها احتمالات نظیر آن‌ها باشند. در این صورت α و β را می‌توان به عنوان پارامترهای انعطاف‌پذیر و یا به عنوان اعداد از پیش تعیین شده‌ای در نظر گرفت که به چندین سیستم سایبرنتیک مرتبط شده‌اند. حال دو سیستم سایبرنتیک که دارای مجموعه‌های یکسان $[x_i,p_i]$ هستند اما ممکن است اطلاعات متفاوتی (نسبت به هدفی یکسان) داشته باشند را در نظر می‌گیریم. آنگاه پارامترهای α و β می‌توانند به عنوان عوامل محیطی نظیر دما و رطوبت و غیره مورد توجه باشند. بعلاوه مجموعه‌ای از عوامل بر تنوع هزینه‌ها تاثیر می‌گذارند. لذا می‌توانیم α و β را عواملی از این دست در نظر بگیریم که اطلاعات یک سیستم سایبرنتیک به آنها بستگی دارد.

گزاره‌های زیر روابط بین آنتروپی جدید، آنتروپی رنی و آنتروپی شانون را برای حالت‌های خاص پارامترها نشان می‌دهد.

- برای $\beta = 1$ داریم:

$$R_{\beta=1}^{\alpha}(P) = R^{\alpha}(P).$$

- برای $\alpha = 1$ داریم:

$$R_{\beta}^{\alpha=1}(P) = R^{\frac{1}{\beta}}(P).$$

- برای $\beta = 1$ و $\alpha \longrightarrow 1$ داریم:

$$R_{\beta=1}^{\alpha\to 1}(P) = H(P).$$

- برای $\alpha = 1$ و $\beta \longrightarrow 1$ داریم:

$$R_{\beta\to 1}^{\alpha=1}(P) = H(P).$$

- برای $\alpha \longrightarrow \beta$ داریم:

$$R_{\beta}^{\alpha\to\beta}(P) = H(P).$$

- برای توزیع یکنواخت $p_i = \frac{1}{n}$، $i = 1.2.\ldots.n$ و برای هر $\alpha \neq \beta\ \ \alpha,\beta > 0$ داریم:

$$R_\beta^\alpha\left(\frac{1}{n}\right) = H\left(\frac{1}{n}\right) = \log_D n\,.$$

در قضیه زیر مشابه آنتروپی رنی، خواصی برای تعمیم آنتروپی رنی پیشنهادی بیان شده است.

قضیه۷: آنتروپی رنی دو پارامتری پیشنهادی دارای ویژگی‌های زیر است:

الف) برای هر $\alpha \neq \beta\ \ \alpha,\beta > 0$، $R_\beta^\alpha(P) > 0$.

ب) بیشترین مقدار برای $R_\beta^\alpha(P)$ زمانی حاصل می‌شود که توزیع احتمال یک توزیع یکنواخت باشد و مقدار این ماکزیمم برابر $\log_D n$ است.

ج) $R_\beta^\alpha(P)$ دارای ویژگی جمع‌پذیری است. یعنی برای متغیرهای تصادفی مستقل X و Y با توزیع‌های احتمال $P = \{p_1,p_2,\ldots,p_n\}$ و $Q = \{q_1,q_2,\ldots,q_n\}$ داریم:

$$R_\beta^\alpha(P,Q) = R_\beta^\alpha(P) + R_\beta^\alpha(Q).$$

۶- میانگین جدید طول کدواژه‌ها

قضیه کدگذاری (بدون نوفه) شانون بیان می‌کند که اگر یک کد در نامساوی کرافت صدق نماید آنگاه متوسط طول کدواژه‌ها، $L = \sum_{i=1}^{n} p_i l_i$ در نامساوی زیر صدق می‌کند:

$$H(P) < L < H(P) + 1.$$

کمپبل [۱۱] با تعریف میانگین طول کدواژه ها در یک کانال بدون نوفه گسسته به صورت :

$$L^\alpha(P) = \frac{\alpha}{1-\alpha}\log_D\left[\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right].$$

که در آن، $\alpha > 0$ و $\alpha \neq 1$، قضیه کدگذاری شانون را برای آنتروپی رنی به صورت زیر تعمیم داد:

$$R^{\alpha}(P) < L^{\alpha}(P) < R^{\alpha}(P) + 1.$$

متناظر با تعمیم دو پارامتری $R_{\beta}^{\alpha}(P)$ ، میانگین جدیدی نیز برای طول کدواژه‌ها در یک کانال بدون نوفه گسسته توسط بهات و دیگران به صورت زیر ارائه شده است:

$$L_{\beta}^{\alpha}(P) = \frac{\alpha}{\beta - \alpha}\log_D\left[\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\alpha-\beta}{\alpha}\right)}\right].$$

که در آن $\alpha,\beta > 0$ و $\alpha \neq \beta$.

گزاره‌های زیر روابط بین L و $L^{\alpha}(P)$ و $L_{\beta}^{\alpha}(P)$ یعنی میانگین طول کدواژه‌ها در آنتروپی شانون، آنتروپی رنی و تعمیم جدید رنی، را برای حالت‌های خاص پارامترها نشان می‌دهد.

- برای $\beta = 1$ داریم:

$$L_{\beta=1}^{\alpha}(P) = L^{\alpha}(P).$$

- برای $\alpha = 1$ داریم:

$$L_{\beta}^{\alpha=1}(P) = L^{\frac{1}{\beta}}(P).$$

- برای $\beta = 1$ و $\alpha \longrightarrow 1$ داریم:

$$L_{\beta=1}^{\alpha\to 1}(P) = L.$$

- برای $\alpha = 1$ و $\beta \longrightarrow 1$ داریم:

$$L_{\beta\to 1}^{\alpha=1}(P) = L.$$

- برای $\alpha \longrightarrow \beta$ داریم:

$$L_{\beta}^{\alpha\to\beta}(P) = L.$$

قضیه زیر رابطه بین کمیت‌های دو پارامتری $R_{\beta}^{\alpha}(P)$ و $L_{\beta}^{\alpha}(P)$ را از دیدگاه قضیه کدگذاری منبع بیان می‌کند.

قضیه ۸ : فرض کنید الفبای منبع $U = \{u_1,u_2,\ldots,u_n\}$ دارای توزیع احتمال $P = \{p_1,p_2,\ldots,p_n\}$ بوده و بعلاوه برای کدگذاری منبع فوق، الفبای کد توسط مجموعه متناهی $S = \{s_1,s_2,\ldots,s_D\}$ مشخص شده باشد. اگر برای طول کدواژه‌ها نامساوی کرافت برقرار باشد آنگاه داریم:

$$R_{\beta}^{\alpha}(P) \leq L_{\beta}^{\alpha}(P).$$

تساوی برقرار است اگر و فقط اگر:

$$l_i = -\log_D\left[\frac{p_i^{\frac{\alpha}{\beta}}}{\sum_{i=1}^{n} p_i^{\frac{\alpha}{\beta}}}\right].$$

که در آن l_i طول کدواژه نظیر u_i است.

اکنون قضیه کدگذاری شانون را در مورد اندازه اطلاع دو پارامتری جدید مورد بررسی قرار می‌دهیم.

قضیه۹ : اگر برای کد مفروضی نامساوی کرافت برقرار باشد آنگاه متوسط طول کدواژه‌ها یعنی $L_{\beta}^{\alpha}(P)$ در نامساوی زیر صدق می‌کند:

$$R_{\beta}^{\alpha}(P) < L_{\beta}^{\alpha}(P) < R_{\beta}^{\alpha}(P) + 1.$$

اثبات قضایای فوق در بهات و دیگران آمده است.

۷- یافته ها:

در این بخش کارایی کد هافمن با معیارهای $R_{\beta}^{\alpha}(P)$ و $L_{\beta}^{\alpha}(P)$ بررسی می شود. به استناد قضیه کدگذاری منبع (قضیه۴) یک کد یکتا گشودنی کاملا بهینه، کدی است که کارایی آن برابر یک باشد. به عبارت دیگر:

$$L = \frac{H(P)}{\log D}.$$

یعنی یک کد یکتا گشودنی، کاملا بهینه است اگر متوسط طول کدواژه‌ها با آنتروپی شانون (در مبنای D) برابر باشد.

هافمن یک روش کدگذاری با طول متغیر را برای یک منبع بدون نوفه به گونه‌ای طراحی نمود که متوسط طول کدواژه‌های آن با آنتروپی شانون (در مبنای D) برابر باشد. بنابراین او یک روش کدگذاری کاملا بهینه ارائه داد.

بر این اساس ارزیابی کارایی کد هافمن با معیارهای $R_{\beta}^{\alpha}(P)$ و $L_{\beta}^{\alpha}(P)$ برای مقادیر مختلف پارامترها شایان توجه است. با توجه به خواص مطرح شده برای این دو معیار می‌توان ادعا نمود که وقتی مقدار یکی از پارامترها یک و دیگری نزدیک به یک باشد آنگاه کارایی کد هافمن به عدد یک نزدیک خواهد بود. همچنین وقتی فاصله پارامتر خیلی کم باشد باز هم کارایی کد نزدیک یک خواهد بود.

بجز وضعیت‌های فوق‌الذکر باید کارایی کد هافمن محاسبه و مورد ارزیابی قرار گیرد. بنابراین با انتخاب مقادیری برای پارامترها کارایی کد هافمن را با معیارهای جدید مورد ارزیابی قرار می‌دهیم.

به این منظور با در نظر گرفتن کد هافمن مثال (۱) جداول زیر را برای مقادیر مختلف پارامترها تهیه نموده و در آن مقادیر آنتروپی جدید، میانگین طول کدواژه‌ها و کارایی کد هافمن را محاسبه نموده‌ایم.

جدول ۳ : کارایی کد هافمن برای $\boldsymbol{\alpha} = ۱$

l_i	p_i	β	R_β^α	L_β^α	η
۱	$\frac{1}{2}$	0.8	۱,۸۵۶۶	۱,۸۸۲۴	۰,۹۸۶۲۷
۲	$\frac{1}{4}$				
۳	$\frac{1}{8}$	0.9	۱,۹۳۱۷	۱,۹۳۸۲	۰,۹۹۶۶۵
۴	$\frac{1}{16}$				
۶	$\frac{1}{64}$	0.95	1.9667	1.9683	0.99917
۶	$\frac{1}{64}$				
۶	$\frac{1}{64}$	0.99	۱,۹۹۳۴	۱,۹۹۳۵	۰,۹۹۹۹۷
۶	$\frac{1}{64}$				

جدول ۴ : کارایی کد هافمن برای $\boldsymbol{\beta} = ۱$

l_i	p_i	α	R_β^α	L_β^α	η
۱	$\frac{1}{2}$	0.8	2.1439	2.1845	0.98142
۲	$\frac{1}{4}$				
۳	$\frac{1}{8}$	0.9	2.0684	2.0764 1	0.99613
۴	$\frac{1}{16}$				
۶	$\frac{1}{64}$	0.95	2.0333	2.0351 3	0.99911
۶	$\frac{1}{64}$				
۶	$\frac{1}{64}$	0.99	2.0065	2.0066	.099996
۶	$\frac{1}{64}$				

جدول ۵ : کارایی کد هافمن برای $\boldsymbol{\alpha}$ **و** $\boldsymbol{\beta}$ نزدیک به یکدیگر

l_i	p_i	α	β	R_β^α	L_β^α	η
۱	$\frac{1}{2}$	0.5	0.51	2.01 2	2.01 3	0.9998
۲	$\frac{1}{4}$					
۳	$\frac{1}{8}$	3	3.1	2.02 1	2.02 2	0.9994
۴	$\frac{1}{16}$					
۶	$\frac{1}{64}$	25	24	1.97 4	1.97 3	0.9995
۶	$\frac{1}{64}$					
۶	$\frac{1}{64}$	70	67	1.97 1	1.97 2	0.9994
۶	$\frac{1}{64}$					

جدول ۶ : کارایی کد هافمن با $\boldsymbol{\alpha}$ ، $\boldsymbol{\beta}$ با فاصله زیاد و بزرگتر از ۱

l_i	p_i	α	β	R_β^α	L_β^α	η
۱	$\frac{1}{2}$	۲۰۰	۱۵۰	۱,۸۱۶	۱,۸۵۶	۰,۹۷۸۳
۲	$\frac{1}{4}$					
۳	$\frac{1}{8}$	۱۰	۱۵	۲,۲۵	۲,۴۱	۰,۹۳۴
۴	$\frac{1}{16}$					
۶	$\frac{1}{64}$	۱۵	۲۰	۱,۵۱	۱,۷۱۷	۰,۹۶۸۱
۶	$\frac{1}{64}$					
۶	$\frac{1}{64}$	۱۵	۱۰	۱,۷۴	۱,۸۱	۰,۹۶۰۷
۶	$\frac{1}{64}$					

با توجه به سه جدول اول ملاحظه می‌شود که انتخاب مقادیر α و β با سه استراتژی:

۱) $\alpha = 1$ و β نزدیک به ۱،

۲) $\beta = 1$ و α نزدیک به ۱،

۳) α و β نزدیک به یکدیگر،

کارایی بسیار نزدیک به ۱ را برای کد هافمن ابقا می‌کند.

در جدول آخر نیز ملاحظه می‌شود که برای مقادیر بزرگتر از یک برای α و β کارایی کد هافمن باز هم نزدیک به ۱ واقع است.

کارایی کد هافمن زمانی که یکی از پارامترها کمتر از ۱ و دیگری بیشتر از یک است دستخوش نوسان بیشتری می شود. جدول زیر موید این مطلب است.

جدول ۷: کارایی کد هافمن برای $(\boldsymbol{\beta > 1 ، \alpha < 1})$ یا $(\beta < 1 ، \alpha > 1)$

l_i	p_i	α	β	R_β^α	L_β^α	η
۱	$\frac{1}{2}$	۳	۰,۵	۱,۱۹	۱,۶۳	۰,۷۳۲۵
۲	$\frac{1}{4}$					
۳	$\frac{1}{8}$	۰,۵	۳	۲,۶۲	۵,۱۲	۰,۵۱۱
۴	$\frac{1}{16}$					
۶	$\frac{1}{64}$	۰,۱	۵	۲,۹۶	۵,۹	۰,۵۰۱
۶	$\frac{1}{64}$					
۶	$\frac{1}{64}$	۵	۰,۱	۱,۰۲	۱,۵۹	۰,۶۴۱
۶	$\frac{1}{64}$					

۸-نتیجه گیری

به دلیل اهمیت اندازه‌های اطلاع و کاربرد آنها در نظریه ارتباطات، در این مقاله تعمیم جدیدی برای آنتروپی رنی، که خود تعمیم آنتروپی شانون می‌باشد، معرفی و ویژگی‌های آن مورد بررسی قرار گرفته است. همچنین در مبحث کدگذاری منبع، میانگین جدیدی برای محاسبه طول کدواژه‌ها تعریف و ارتباط آن با اندازه اطلاع تعمیم یافته جدید بررسی شده است.

بخصوص کارایی کد هافمن برای مقادیر مختلف پارامترهای آنتروپی جدید، مورد ارزیابی قرار گرفت و مشخص شد که کارایی کد هافمن با آنتروپی و میانگین طول کد واژه جدید نیز با انتخاب مناسب پارامترها تقریبا بهینه باقی می‌ماند.

استفاده از سایر آنتروپی‌های پارامتری از قبیل آنتروپی تی‌سالیس و تعمیم‌های آن و به کارگیری میانگین طول کدواژه خوش تعریف متناظر با آن می‌تواند در کارهای آینده مورد توجه باشد.

فهرست منابع

[۱] H. Nyquist, Certain Factor Affecting Telegraph Speed. Bell System Technical Journal, vol. 3, No. 2, 324-346, 1924.

[۲] R. V. L. Hartley, Transmission of information. Bell System Technical Journal, 7, 535, 1928.

[۳] C.E. Shannon, A mathematical theory of communication. Bell System Technical Journal, 27, 379-423, 1948.

[۴] T. Cover, J. Thomas, Elements of Information Theory. ISBN 0-471-06259-6, 1991.

[۵] R. W. Hamming, Error detecting and error correcting codes. **29** (2): 147–160, 1950.

[۶] DA. Huffman, A method for the construction of minimum redundancy codes, in: Proceedings of the IRE.; 40:1098-1101, 1952.

[۷] A. Renyi, On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press 4, 547-562, 1961.

[۸] J. Havrda and F. Charvat, Quantification method of classification processes. Concept of structural a-entropy. Kybernetika 3(1), 30-35, 1967.

[۹] B. D. Sharma and D. P. Mittal, New non-additive measures of entropy for discrete probability distributions. J. Math. Sci.,10, 28-40, 1975.

[۱۰] A. H. Bhat, N. A. Siddiqui, I. A. Mageed, S. Alkhazaleh, V. R. Das and M. A. K Baig, Generalization of Renyi's Entropy and its Application in Source Coding. Appl. Math. Inf. Sci. 17, No. 5, 941-948, 2023.

[۱۱] L. L. Campbell, A coding theorem and Renyi's entropy. Information and control 8(4), 423-429, 1965.

A Two-Parametric Generalization of Renyi's Entropy and its New Mean Cod-word Length

Alireza Hosseinzadeh[1], Mahdi yaghoobi avval Riabi[2*]

1. Department of Statistics, Gonabad Branch, Islamic Azad University, Gonabad, Iran
2. Department of Statistics, Gonabad Branch, Islamic Azad University, Gonabad, Iran

*Corresponding Author: Mahdi yaghoobi avval Riabi

ABSTRACT

Renyi's entropy was introduced as a one-parameter generalization of Shannon entropy by Alfred Renyi in (1961). This generalized information measure has similar properties to Shannon's entropy.

In this paper, we first review Shannon and Renyi entropies and some of their important properties. Then, a two-parameter generalization of Renyi's entropy with the corresponding new mean code-word length, that was proposed by Bhatt et. al (2023), and its important properties are studied. In particular, for some values of the parameters, we evaluate the efficiency of the Huffman code with respect to this generalized entropy and its new mean code-word length. The results indicate that the efficiency of the Huffman code with entropy and new mean code-word length remains almost optimal with appropriate selection of parameters.

Keywords: Shannon's entropy, Renyi's entropy, Mean Code-word length, Kraft's inequality, Huffman code.