نظریه اطلاعات و کدگذاری

بدیدآورندگان:

آ. جانز

ج. [illegible] جانز

برگردان:

مرتضی اسماعیلی

استاد دانشکده علوم ریاضی

دانشگاه صنعتی اصفهان

دانشگاه صنعتی اصفهان
مرکز نشر

شماره کتاب ۸۶ گروه علوم ۳۰

نظریه اطلاعات و کدگذاری

پدیدآورندگان	آ. جانز و ج. م. جانز
برگردان	مرتضی اسماعیلی
صفحه آر	مرکز نشر دانشگاه صنعتی اصفهان
طراح جل	مرکز نشر دانشگاه صنعتی اصفهان
لیتوگرافی، چاپ و صحافی	چاپخانه دانشگاه صنعتی اصفهان
ناشر	مرکز نشر دانشگاه صنعتی اصفهان
چاپ دوم	تابستان ۱۳۹۴
شمارگان	۵۰۰ جلد
شابک	۹۷۸-۹۶۴-۶۰۲۹-۹۷-۳
قیمت	۱۰۰۰۰۰ ریال

جانز، گارت، ۱۹۴۶-م. Jones, Gareth A.
نظریه اطلاعات و کدگذاری/تألیف گ. آ. جانز، ج. م. جانز؛
ترجمه مرتضی اسماعیلی—اصفهان؛ مرکز نشر دانشگاه صنعتی اصفهان، ۱۳۸۳.
IX، ۲۱۱ ص
ISBN 964-6029-97-3
فهرست‌نویسی بر اساس اطلاعات فیپا.
عنوان اصلی: Information and coding theory.
واژه‌نامه.
۱. نظریه اطلاعات. ۲. نظریه
الف. جانز، جوزفین مری، ۱۹۴۶ — م. Jones, Josephine Mary
ب. اسماعیلی، مرتضی، ۱۳۳۶- مترجم.
ج. مرکز نشر دانشگاه صنعتی اصفهان.
د. عنوان
۸ت۹۴ج/Q۳۶۰ ۰۰۳/۵۴
۱۳۸۲
کتابخانه ملی ایران ۸۲-۳۰۱۶۶م

اصفهان: دانشگاه صنعتی اصفهان ـ مرکز نشر ـ کدپستی ۸۴۱۵۶-۸۳۱۱۱، تلفن: ۱۰-۳۳۹۱۲۵۰۹(۰۳۱) دورنگار: ۳۳۹۱۲۵۵۲ (۰۳۱)
برای خرید اینترنتی کلیه کتاب‌های منتشره مرکز نشر می‌توانید به وبگاه http://publication.iut.ac.ir مراجعه و یا مستقیما از کتابفروشی مرکز نشر واقع در کتابخانه مرکزی دانشگاه صنعتی اصفهان (تلفن ۳۳۹۱۲۹۵۲) خریداری فرمائید.

پیش‌گفتار مترجم

نظریه اطلاعات و کدگذاری دانش انتقال دقیق و اقتصادی داده‌ها از نقطه‌ای به نقطه دیگر یا از زمانی به زمان دیگر است. به‌عنوان نمونه می‌توان به مکالمات تلفنی و ارسال تصویر از دیگر سیارات به زمین توسط ایستگاه‌های فضایی و نگهداری داده‌ها روی CD اشاره نمود.

در حالی‌که نظریه اطلاعات توجه به روش‌های کدگذاری خروجی یک متغیر تصادفی با استفاده بهینه از توزیع احتمال آن متغیر به‌منظور کاهش افزونگی دارد، در نظریه کدگذاری کانال هدف اضافه نمودن افزونگی به‌منظور ایجاد توان‌مندی لازم جهت خنثی‌نمودن توان تخریبی پارازیت روی داده‌ها است. این دو در اولی تا حد امکان افزونگی کاهش یافته و در دومی به‌قدر کافی افزونگی به داده‌ها افزوده می‌شود. در هر دو مرحله هدف ایجاد بستر مناسب جهت ارسال سریع، درست و اقرون داده با ا رسنده به گیرنده است.

کتاب حاضر حاوی مفاهیم و روش‌های اساسی کدگذاری منبع (خروجی یک متغیر تصادفی) و کانال است. با توجه به ساختار ارائه مطالب، می‌توان از این کتاب به‌عنوان منبع اصلی ارائه یک درس کدگذاری (منبع و کانال) در مقطع کارشناسی برای رشته‌های ریاضی کاربردی، کامپیوتر، و مخابرات استفاده نمود.

تلاش شده است که واژه‌ها و نمادها کاملاً به فارسی بیان شوند. با این حال به‌جهت وجود محدودیت‌ها دو مورد استثنا رخ داده است. براساس ساختار دستوری جمله از هر دو کلمه 'کدگشایی' و ' دکد کردن' استفاده شده است. به‌منظور یکسان نمودن بازگویی اعداد موجود در یک رابطه یا جدول نماد '/' به‌عنوان علامت تقسیم به‌کار رفته و از این‌رو بر خلاف متون فارسی برای ممیز از '.' استفاده شده است؛ بنابراین به‌عنوان نمونه اعداد ۰.۹ و ۵.۳ به‌ترتیب چهار تقسیم بر نه و پنج و سه‌دهم خوانده می‌شوند.

مرتضی اسماعیلی
عضو هیئت علمی دانشگاه صنعتی اصفهان
بهمن ۱۳۸۲

پیش‌گفتار مؤلف

این پیش‌گفتار در زمانی در حال نوشته شدن است که قرن بیستم در حال پایان است. تاریخ نویسان احتمالاً از این قرن به‌عنوان قرن اطلاعات یاد خواهند کرد، درست مانند قرن قبل که به مرحله صنعتی شدن نسبت داده می‌شود. پیشرفت‌های فنی متوالی مانند تلفن، رادیو، تلویزیون، کامپیوتر و اینترنت تأثیرات عمیقی روی روش زندگی ما داشته‌اند. می‌توانیم تصویرهایی از سطح مریخ با شکل ابتدایی جهان را ببینیم. محتوای کتاب‌های یک قفسه از کتابخانه قابل فشرده شدن روی یک مطعه پلاستیکی تقریباً بی‌وزن است. میلیاردها نفر می‌توانند یک بازی فوتبال را هم‌زمان دنبال کنند، یا موفق به برقراری تماس فوری با دوستان خود در سراسر دنیا بدون ترک خانه بسند. به‌طور خلاصه، در حال حاضر می‌توان انبوهی از اطلاعات را با سرعت، دقت و کارایی اعجاب‌انگیزی ذخیره، ارسال و پردازش نمود.

واضح است که این پیشرفت‌ها بدون مبانی نظری اتفاق نمی‌افتد و همچون در بسیاری موارد دیگر بخش عمده‌ای از این پیشرفت مرهون ریاضیات است. بسیاری از پیشرفت‌های اولیه ریاضی در این زمینه در اواسط قرن بیستم به‌وسیله مهندسین حاصل شد و این اغلب با اتکا بر درک شهودی و تجربی بوده تا این‌که یک دانش نظری عمیق آنها را به کشف‌هایشان هدایت کرده باشد. ریاضی‌دانان، که از دیدن کاربردهای جدید موضوع کاری خود خوشحال بودند خیلی زود وارد صحنه شدند و مثال‌های عملی مهندسین را به نظریه‌های وسیعی همراه با تعاریف، قضایا و اثبات‌ها توسعه دادند. شاخه‌های جدیدی در ریاضیات خلق شدند و چندین شاخه قدیمی‌تر تحت تأثیر کاربردهای غیرمنتظره تقویت شدند: چه کسی می‌توانست پیش‌بینی کند که کدهای تصحیح کننده خطا می‌توانند بر مبنای خم‌های جبری روی میدان‌های متناهی باشند یا این‌که سیستم‌های رمزنگاری وابسته به اعداد اول باشند؟

نظریه اطلاعات و نظریه کدگذاری دو جنبه مرتبط با هم از مسئله چگونگی ارسال سریع و دقیق اطلاعات از یک منبع از طریق یک کانال به یک گیرنده است. این در برگیرنده مسئله چگونگی ذخیره کردن اطلاعات نیز می‌شود که در آن گیرنده می‌تواند همان منبع ولی در زمان بعد باشد. به‌عنوان مثال، کشف فضا منجر به ایجاد تقاضا برای ارسال دقیق یک سیگنال خیلی ضعیف از طریق یک کانال فوق‌العاده شلوغ شده است؛ دلیلی برای فرستادن یک فضاپیما به مریخ وجود ندارد اگر نتوان پیام‌هایی را که او می‌فرستد شنید و کدگشایی کرد.

این نظریه در ساده‌ترین فرم خود از روش‌های مقدماتی نظریه احتمال و جبر خطی استفاده می‌کند، اگرچه پیشرفت‌های بعدی بر مبنای موضوعاتی چون ترکیبیات و هندسه جبری بوده است.

یک مسئله مهم چگونگی فشرده‌سازی اطلاعات به‌منظور ارسال سریع و یا ذخیره اقتصادی آن می‌باشد. این‌را می‌توان با کاهش افزونگی انجام داد: یک مثال آشنا استفاده از اختصاراتی چون 'UK' ، 'IBM' و 'radar' به‌جای نام کامل می‌باشد که بسیاری از سمبل‌های آن‌ها از نقطه‌نظر محتوای اطلاعاتی زاید هستند. مشابهاً، ما اغلب نام نزدیک‌ترین دوستان و بستگان خود را کوتاه به‌کار می‌بریم، به‌طوری‌که William تبدیل به Will یا Bill می‌شود.

یک مسئله مهم دیگر چگونگی کشف و اصلاح خطا در اطلاعات است. نمی‌توان همیشه به انسان و ماشین از جهت عدم ارتکاب اشتباه اطمینان کرد، و اگر این اشتباهات تصحیح نشوند پیامدهای نامطلوبی را می‌توانند در پی داشته باشند. در این‌جا شیوه حل مسئله افزایش افزونگی است که با اضافه کردن سمبل‌هایی که باعث تقویت و حفاظت پیام می‌شود صورت می‌گیرد. این‌را الفبای NATO متشکل از Alpha، Bravo، Charlie، ...، که در ارتباطات گفتاری توسط نیروهای مسلح، خطوط هوایی و سرویس‌های اضطراری به‌کار می‌رود حروف $A, B, C, \ldots$ را با کلماتی جایگزین می‌کند که به‌نحوی انتخاب شده‌اند که تا حد امکان به لحاظ شنیداری متفاوت باشند؛ به‌عنوان نمونه حروف B و V اغلب از نظر صدا غیرقابل تمیز هستند (اساساً این دو حرف بعضی زبان‌ها یکی هستند)، اما ممکن نیست که Victor به‌عنوان Bravo درنظر گرفته شود حتی اگر به‌صورت Bictor شنیده شود.

نظریه اطلاعات که بخش عمده آن ریشه در مقاله مهم سال ۱۹۴۸ شانون [Sh 48] دارد، از توزیع احتمال برای اندازه‌گیری اطلاعات (از طریق تابع آنتروپی)، و ارتباط دادن آن با میانگین طول کلمه در کدگذاری‌های آن اطلاعات استفاده می‌کند. بالاخص، قضیه اساسی شانون وجود کدهای خوب تصحیح کننده خطا را تضمین کرده، و هدف نظریه کدگذاری استفاده از روش‌های ریاضی برای ساخت یک چنین کدهایی به‌همراه الگوریتم‌های کارا برای استفاده از آنها می‌باشد. نظریه کدگذاری، علی‌رغم نام آن، درگیر در مطالعه کدهای مخفی نیست: این موضوع که رمزنگاری می‌باشد ارتباط نزدیکی با نظریه اطلاعات از طریق مفاهیم آنتروپی و افزونگی دارد، ولی ما به جهت این‌که فنون ریاضی مربوطه نسبتاً متفاوت هستند در این‌جا به آن نمی‌پردازیم.

این کتاب بر مبنای یک درس سال سوم دوره کارشناسی که در اوایل دهه ۱۹۸۰ در دانشگاه سوتامتن[1] ارائه شد تنظیم شده و هدف آن تلاشی برای توصیف ایده‌های اصلی نظریه اطلاعات و نظریه کدگذاری می‌باشد. پیش‌نیازهای اصلی آن نظریه مقدماتی احتمال و جبر خطی و مقدار کمی حسابان است. اکثر کتاب‌های درسی در این زمینه تا حد زیادی و یا به‌طور کامل تنها روی یکی از دو موضوع نظریه اطلاعات و نظریه کدگذاری تمرکز دارند. با این حال، این دو موضوع از طریق قضیه شانون تا حد زیادی به یکدیگر مرتبط بوده، و ما

[1] Southampton University

احساس می‌کنیم که دلایل قوی برای فراگیری توأم آنها، حداقل در مراحل اولیه، وجود دارد.

فصل‌های ۱—۵ (حدود ۶۰٪ کتاب) اساساً راجع به نظریه اطلاعات است. فصل اول، که پیش‌نیاز خیلی کمی دارد، نشان می‌دهد که چگونه اطلاعات را باید کدگذاری کرد به‌قسمی‌که عمل کدگشایی غیرمبهم و لحظه‌ای باشد. نتایج اصلی این فصل قضیه سردناس — پترسن[1] و نامساوی‌های کرفت و مک‌میلان[2] می‌باشند که در رابطه با وجود یک چنین کدهایی هستند. فصل ۲ به معرفی کدهای هافمن می‌پردازد که، نسبتاً مانند کد مرس[3]، میانگین طول کلمه را با نسبت دادن کدکلمه‌های کوتاه‌تر به سمبل‌های محتمل‌تر می‌نیمم می‌کند؛ در این‌جا (مانند فصل‌های ۳—۵) از نظریه مقدماتی احتمال، و در واقع توزیع احتمال متقارن استفاده می‌کنیم. در فصل ۳، از تابع آنتروپی که بر پایه یک توزیع احتمال و لگاریتم آنهاست در جهت اندازه‌گیری اطلاعات و مرتبط نمودن آن، به‌وسیله قضیه‌ای از شانون، با میانگین طول کلمات کدگذاری‌ها استفاده می‌کنیم. فصل ۴ به مطالعه چگونگی ارسال اطلاعات از طریق یک کانال، که ممکن است به دلیل وجود پارازیت خطاهایی رخ دهد، اختصاص دارد. احتمال‌های شرطی امکان تعریف یک دستگاه آنتروپی را فراهم می‌سازد که بر مبنای آن می‌توان اطلاعات را از چند نقطه‌نظر، مثلاً از دیدگاه فرستنده و گیرنده، اندازه‌گیری کرد. این منجر به مفهوم ظرفیت کانال می‌شود که عبارت است از بیشترین مقدار اطلاعات که یک کانال می‌تواند ارسال کند. در فصل ۵ قضیه اساسی شانون را ملاحظه می‌کنیم که می‌گوید تا زمانی که وجود پارازیت، می‌توان اطلاعات را با دقت دلخواه و با نرخ به‌قدر دلخواه نزدیک به ظرفیت کانال ارسال نمود. رئوس اثبات این قضیه را برای کانال ساده ولی مهم دوتایی متقارن ارائه داده و سپس اثبات کامل قضیه را برای این کانال در ضمیمه C می‌آوریم؛ این اثبات متکی بر تنها نتیجه پیشرفته مورد نیاز ما از نظریه احتمال، یعنی قانون اعداد بزرگ، است که در ضمیمه B ارائه شده است.

ایده اساسی قضیه شانون این است که می‌توان اطلاعات را با صحت بالائی ارسال کرد و این کار با استفاده از کدکلماتی که به‌قدر کافی با هم تفاوت دارند و از این‌رو حتی اگر بعضی از سمبل‌های آنها نادرست دریافت شوند گیرنده آنها را با هم اشتباه نخواهد کرد، صورت می‌گیرد (Bravo و Victor را به یاد آورید). متأسفانه قضیه و اثبات آن چگونگی پیدا کردن مثال‌های مشخصی از این کدها را بیان نمی‌کنند، و این کار هدف نظریه کدگذاری است که در فصل‌های ۶ و ۷ به آن می‌پردازیم. در این فصل‌ها که نسبتاً از فصل‌های قبلی خود طولانی‌تر هستند چند مثال نسبتاً ساده از کدهای تصحیح کننده خطا را معرفی می‌کنیم. برای این منظور در فصل ۶ از روش‌های مستقیم و مقدماتی استفاده می‌کنیم؛ نتیجه اصلی این فصل کران کره‌چینی همینگ[4] است که با به‌کار بردن یک ایده ساده هندسی یک کران بالا روی تعداد کدکلماتی که بتوانند تعداد خطای مفروضی را تصحیح کنند ارائه می‌دهد.

در فصل ۷ مثال‌های نسبتاً پیشرفته‌تری از کدهای تصحیح کننده خطا را می‌سازیم؛ این کار

---

[1] Sardinas - Patterson theorem
[2] Kraft and Mcmillan inequalities
[3] Morse code
[4] Hamming's sphere - packing bound

با استفاده از جبر خطی و نظریه ماتریس‌ها، و مشخصاً مفاهیمی چون فضاهای برداری، زیر فضاها، مبنا و بعد، رتبه ماتریس و عملیات سطری و ستونی صورت می‌گیرد. همچنین به‌صورت مختصر چگونگی ارتباط ایده‌هایی از ترکیبیات و هندسه، مانند طرح‌های قالبی و هندسه‌های تصویری، را با کدها نشان می‌دهیم.

محدودیت‌های معمول فضا و زمان ما را وادار به حذف موضوعات جالبی چون رابطه با رمزنگاری که در بالا به آن اشاره شد، و اشاره مختصر به چند مطلب دیگر نموده است. به‌عنوان نمونه، در نظریه اطلاعات منابع مارکف (آنهایی که حافظه‌ای از پیشامدهای قبل دارند) تنها به‌عنوان یک تمرین ظاهر شده، و مشابهاً در نظریه کدگذاری به کدهای دوری و ارتباط آنها با حلقه چندجمله‌ای‌ها اشاره نشده است. در عوض در انتهای کتاب پیشنهادهایی را برای مطالعه بیشتر آورده‌ایم.

یک درس نظریه اطلاعات را می‌توان براساس فصل‌های ۱–۵، و احتمالاً با مطالب بیشتری راجع به منابع مارکف و یا رابطه با رمزنگاری تنظیم نمود. یک درس نظریه کدگذاری می‌تواند بر پایه فصل‌های ٦ و ۷ و مقداری پیش‌نیاز از فصل ۵ به همراه مطالب اضافی دیگری همچون کدهای دوری یا شمارش وزن استوار باشد.

تلاش کرده‌ایم تا حد امکان کلیه قضایا را به بنیان‌گزاران ایده‌های اصلی معرفی شده در کتاب داده، و از منابع اولیه نتایج، مثال‌ها، و تمرینات قدردانی کنیم. بدون شک از این دیدگاه بدون ضعف نبوده ولی ضعف‌های احتمالی بدون هیچ قصد و عمدی هستند.

فهرست مندرجات