بسم الله الرحمن الرحیم

تحلیل داده‌های کلان در شبکه

گردآوری و تدوین:
معاونت پژوهش و تولید علم

دانشگاه اطلاعات و امنیت ملی
معاونت پژوهش و تولید علم
مؤسسه چاپ و انتشارات
۱۳۹۶

عنوان و نام پدیدآور	:	تحلیل داده‌های کلان در شبکه/گردآوری و تدوین معاونت پژوهش و تولید علم.
مشخصات نشر	:	تهران: دانشگاه اطلاعات و امنیت ملی، موسسه چاپ و انتشارات، ۱۳۹۶.
مشخصات ظاهری	:	۶۲۲ص. ش ۳۴۰
شابک	:	۹۷۸-۶۰۰-۸۴۱۱-۷۹-۶
وضعیت فهرست نویسی	:	فیپا
موضوع	:	داده‌های کلان -- مدیریت
موضوع	:	Big data -- Management
موضوع	:	داده‌های کلان -- تدابیر ایمنی
موضوع	:	Big data -- Security measures
موضوع	:	داده‌پردازی -- مدیریت
موضوع	:	Electronic data processing -- Management
موضوع	:	شبکه‌های کامپیوتری -- تدابیر ایمنی
موضوع	:	Computer networks -- Security measures
موضوع	:	اینترنت -- تدابیر ایمنی
موضوع	:	Internet -- Security measures
موضوع	:	حفاظت داده‌ها
موضوع	:	Data protection
شناسه افزوده	:	دانشگاه اطلاعات و امنیت ملی. موسسه چاپ و انتشارات
شناسه افزوده	:	دانشگاه اطلاعات و امنیت ملی. معاونت پژوهش و تولید علم
رده‌بندی کنگره	:	۱۳۹۶ ت۲۵د/ QA۷۶
رده بندی دیویی	:	۷۴/۰۰۵
شماره کتابشناسی ملی	:	۴۶۶۲۵۰۹

مشخصات:

عنوان: تحلیل داده‌های کلان در شبکه

گردآوری و تدوین: معاونت پژوهش و تولید علم

ناشر: مؤسسه چاپ و انتشارات دانشگاه اطلاعات و امنیت ملی

شمارگان: ۱۰۰۰ نسخه

تاریخ انتشار: ۱۳۹۶

نوبت چاپ: اول

قیمت: ۲۸۰۰۰۰ ریال

شابک: ۶- ۷۹-۸۴۱۱-۶۰۰-۹۷۸

تلفن: ۲۲۴۶۹۷۳۹

تهران: صندوق پستی ۱۴۳۸ـ ۱۶۷۶۵

مقدمهٔ ناشر

بی‌تردید یکی از مؤلفه‌های مهم تحولات عصر اطلاعات، در حوزهٔ تولید داده خودنمایی می‌کند. تغییر در الگوی زیستی بشر از منظر وابستگی اعتیادگونه به فن ابزارهای نوین و ثبت و ضبط اطلاعات مربوط به تمام کنش‌های سایبری این ابزارها، حجم انبوهی از داده‌ها را روی دست بشر گذاشته است. شرکت آی‌بی‌ام در گزارشی اعلام کرده است که نود درصد کل داده‌هایی که بشر از آغاز حیاتش بر روی کرهٔ خاکی از خود به‌جا گذاشته، در طی دو سال اخیر تولید شده است. این داده تقاً یعنی انقلابی ویرانگر و ساختارشکن در روند تولید داده توسط بشر. نکتهٔ مهم این است که این رشد سرسام‌آور به‌صورت نمایی ادامه دارد و اغراق نیست اگر تصور شود با توسعهٔ اینترنت اشیا و فراگیر شدن آن تا سال ۲۰۲۰ حجم دادهٔ تولیدی به همان نسبت افزایش یابد.

اما آیا انقلاب داده صرفاً ناظر بر حجم داده‌های تولیدی است؟ به‌طور قطع این تصویر کاملی از انقلاب داده نیست، وجه مهم‌تر این تحول در ارزش افزوده‌ای است که این داده‌ها تحلیل و فراوری می‌شوند. حوزهٔ اطلاعاتی و امنیتی نیز از پیشروترین زمینه‌ها در استفاده از ابرداده‌ها برای تأمین اهداف امنیتی است. توسعهٔ مهارت‌های تحلیلی و تسلط بر ساختارهای فنی و ابزاری شبکه‌ها ضرورتی است که بدون نائل شدن به آن، امکان مدیریت تولید، ذخیره، بازیابی، و تحلیل داده‌های بزرگ دور از دسترس است.

در همین راستا، معاونت پژوهش و تولید علم بر آن است با تولید محتوای غنی علمی در حوزه‌های متنوع مرتبط با داده‌ها و شبکه‌های داده‌ای، زمینهٔ رشد و بالندگی علمی و

حرفه‌ای دانشجویان و افسران اطلاعاتی را فراهم آورد که کتاب حاضر حاصل بخشی از این تلاش‌هاست. این کتاب در دو بخش و هفده فصل تدوین شده و مجموعه‌ای از ترجمهٔ مقاله‌های برتر در حوزهٔ داده‌های بزرگ و شبکه‌های داده‌ای است. بخش اول شامل یازده فصل با موضوعات برتر در حوزهٔ داده‌های بزرگ و شبکه‌های داده‌ای است. بخش اول شاکل یازده فصل با موضوعات مرتبط با ابرداده‌ها، داده‌های باز، زیرساخت‌های داده و مباحث حرفه‌ای، حقوقی، حاکمیتی و اخلاقی حوزهٔ داده‌هاست. بخش دوم در شش فصل به تشریح مباحث تحلیل داده‌ها با رویکرد امنیت شبکه‌ای خواهد پرداخت.

در فصل اول با عنوان «مفهوم‌پردازی داده» مباحث مفهومی و فلسفی در مورد داده تشریح شده و به‌جای نگاه فنی و فنی به فرایند چگونگی تولید و تحلیل داده، خود داده از منظر هستی‌شناسی مورد مطالعه قرار گرفته است. نویسنده با تأکید بر چهارچوب‌های هویتی فنی، اقتصادی، اخلاقی، فضایی و فلسفی داده، درک داده و ماهیت صحیح انقلاب داده‌ای را مستلزم تحلیلی متنوع‌تر از چیزی می‌داند که تبیین داده بزرگ و باز در حال حاضر ارائه می‌دهد.

در فصل دوم با عنوان «داده‌های کوچک، زیرساخت‌های داده‌ای و واسطه‌های داده‌ها» با اشاره به اینکه قبل از ظهور اصطلاح ابرداده یا داده‌های بزرگ، عبارت داده‌های کوچک موضوعیت نداشته و تمام گونه‌های داده یکسان تلقی می‌شدند، اشعار می‌دارد با ورود به عصر داده‌های بزرگ نیاز به تبیین داده‌های کوچک، زیرساخت‌های داده‌ای مربوط به انتقال و استفادهٔ مجدد و اقتصادی از آن ضروری است.

«داده‌های باز و مرتبط» عنوان فصل سوم کتاب است که نویسنده با اشاره به هزینه و منابع لازم برای تولید مجموعه داده‌ها و ازسویی ظهور بازارهای پُرسود استخراج ارزش افزوده از داده‌ها، اذعان می‌دارد که دسترسی به داده تا حدودی محدود شده است. حتی در مورد داده‌های نسبتاً باز نیز نیازمندی به تجهیزات و سخت‌افزارهای ویژه یا مهارت‌های تحلیلی پیچیده برای بهره‌برداری از داده‌ها، امکان استفادهٔ عمومی از این مزیت را محدود ساخته است. جنبش دادهٔ باز براساس سه اصلِ باز بودن، مشارکت و همکاری در صدد رفع

این انحصارها و تمهید زمینه‌های استفادهٔ عمومی از داده‌هاست.

فصل چهارم با عنوان «داده‌های بزرگ» به فرازوفرود عمیق در زمینهٔ انتظارها و توقع‌ها نسبت به کاربرد و مزایای آن اشاره می‌کند. علی‌رغم اغراق و بزرگ‌نمایی ویژگی‌ها و کاربردهای داده‌های بزرگ در ابتدا، در سال‌های بعد نوعی سرخوردگی به‌وجود آمد و در سال‌های اخیر با ارتقاء زیرساخت‌های سخت‌افزاری و توسعهٔ مهارت‌های تحلیلی، یک بازسازی واقعی و عملیاتی از کارایی و سودمندی را شاهد بوده است. نویسنده با تأکید بر سه خصوصیت حجم، سرعت و تنوع داده‌های بزرگ، تلاش دارد فهم صحیحی از مفهوم و کاربرد داده‌های بزرگ ارائه دهد.

در فصل پنجم با عنوان «تهیه‌کنندگان و منابع داده‌های بزرگ» تولیدکنندگان و منابع داده‌های بزرگ به تفصیل مورد مطالعه قرار گرفته‌اند. توسعهٔ فناوری‌ها و ادغام نرم‌افزارها با سخت‌افزارها، توسعهٔ دستگاه‌های سیار به اینترنت و ادغام تمام شئون زندگی روزمره از ترافیک و تجارت گرفته تا مسافرت و انرژی‌های تولید و توزیع کالا با اینترنت و تولید داده و ازسویی مقرون‌به‌صرفه شدن ذخیره‌سازی داده‌های تولیدشده از این فرایندها موجب توسعهٔ روزافزون منابع و تولیدکنندگان و حجم داده‌ها شده است.

در فصل ششم با عنوان «تجزیه و تحلیل داده» به این نکته اشاره شده که داده‌ها به‌تنهایی مفید نبوده و کاربرد آنها منوط به استخراج مفهوم و ارزش نهفتهٔ آنهاست. بر این اساس، چالش‌های این فراوری، ازجمله موضوع وفور و پراکندگی داده‌ها، سیر زمانی و پویایی داده، ابهام و درهم‌ریختگی، ارتباط درونی، ماهیت نیمه‌ساختارمند یا فاقد ساختار داده‌ها و اینکه حجم اعظم داده بدون وجود هرگونه پرسش یا هدف اولیه تولید می‌شوند، مقدمات مهمی هستند که باید مورد توجه قرار گیرند.

«منطق حرفه‌ای و دولتی برای داده‌های بزرگ» عنوان فصل هفتم این کتاب است که هدف آن تأکید بر این اصل است که انقلاب داده‌ای به شکل منفعل و فارغ از مبانی فکری هدایت‌کنندهٔ آن شکل نمی‌گیرد. رویکردهای هرچند متنوع نسبت به شیوه‌های جدید

دانستن و عمل کردن در عصر اطلاعات و منافع حاصل از این تحقق این تحول «اصول گفتمانی» انقلاب داده‌ای را شکل می‌دهند.

در فصل هشتم با عنوان «مفصل‌بندی پژوهش‌های علوم انسانی، علوم اجتماعی و علوم پایه» مباحث معرفت‌شناسی مبتنی بر زمینه‌ها و چالش‌های نوظهور رشته‌های علمی مختلف، مورد توجه قرار گرفته است. براساس این گزاره که «تحولات در علوم اغلب با تحول در سنجش آغاز می‌شود»[1] این سؤال مطرح می‌گردد که آیا تغییرات در نحوهٔ تولید مطالب خام اطلاعات و دانش و نیز نحوهٔ مدیریت و تحلیل آنها تغییر فاحشی در اصول اساسی نحوهٔ درک و کاربرد علوم انسانی و علوم اجتماعی به‌وجود خواهد آورد و آیا نیازی به تولید رشته‌های درسی جدید و حوزه‌های تحقیقی مستقل وجود دارد یا خیر؟

در فصل نهم با عنوان «مسائل فنی و سازمانی» چالش‌های به‌وجودآمده تحت تأثیر توسعهٔ زیرساخت‌های داده‌ای، داده‌ها، بازنمود داده‌ها نظیر حیطه‌بندی مجموعهٔ داده‌ها، دسترسی به داده، کیفیت داده، یکپارچه‌سازی و قابلیت کاربرد متقابل داده، مباحث تحلیلی و محیطی و مهارت‌ها، توانمندی‌ها و ظرفیت‌های سازمانی مورد بحث و بررسی قرار گرفته است.

در فصل دهم با عنوان «نگرانی‌های اخلاقی، سیاسی، اجتماعی و حقوقی» به تناقض نماهایی پرداخته شده که در فرایند تولید و بهره‌برداری از داده‌های بزرگ رخ می‌نمایند. در حوزهٔ حاکمیتی، مردم در قبال از دست رفتن حریم خصوصی خود کالای امنیت را دریافت می‌کنند. اصول اخلاقی، حقوقی و اجتماعی که بتواند بین این دو کفهٔ تعادل مطلوبی برقرار کند در این فصل مورد بحث قرار گرفته است. اصولی برای هدایت تجارت داده‌ای و سایه‌های داده‌ها، حریم خصوصی، امنیت داده‌ای، پروفایل‌بندی، طبقه‌بندی اجتماعی، کنترل خاموش، حاکمیت مشارکتی و فن‌سالارانه، مالکیت حقوق معنوی سایبری و... .

«مفهوم انقلاب داده‌ها» آخرین فصل بخش نخست است که به شکاف‌های موجود در تفکر مفهومی و دانش، عطف توجه نموده و گفته شده که چنین فضایی باید به دو شیوه

---

۱. سینال آرال

مدنظر باشد: اول، از طریق بازتاب فلسفی و تحلیل انتقادی مفهومی و مختصر و دوم، از طریق روش تجربی جامع در مورد پیدایش، تشکیل، عملکرد و تکامل مجموعهٔ داده‌ها.

شروع بخش دوم با فصل دوازدهم تحت عنوان «استنباطی برای گراف و شبکه: تطبیق ابزارهای کلاسیک با داده‌های مدرن» تلاش دارد کاربست نظریهٔ گراف در تحلیل و فهم شبکه و کار روی داده‌های رابطه‌ای براساس استفاده از روش‌های تحلیل آماری و ریاضیات را تشریح کند. نویسنده ضمن طرح مثال‌های زنده، ایده‌های جدیدی در خصوص روش‌های استنباط الگو و ساختار شبکه‌های عظیم با تأکید بر نقش آمار و ریاضیات ارائه می‌دهد.

فصل سیزدهم با عنوان «تشخیص سریع حملات در شبکه‌های رایانه‌ای با روش‌های تشخیص سریع ترین نقطهٔ تغییر» ضمن مفهوم «سریع‌ترین نقطهٔ تغییر» براساس ایجاد تعادل بین دو مؤلفهٔ تعداد هشدار اشتباه و سرعت تشخیص ناهنجاری، یک سیستم تشخیص نفوذ ناهنجاری پیشنهاد داده و الگوریتم‌های محاسباتی آن را تشریح و کارایی آن در تشخیص حملات را نیز به آزمون گذاشته است.

فصل چهاردهم با عنوان «تشخیص آماری نفوذ در شبکه‌های رایانه‌ای با استفاده از آمار اسکن» به معرفی یک روش قابل ارتقاء محاسباتی برای تشخیص زیرگراف‌های ناهنجار در گراف‌های عظیم وابسته به زمان می‌پردازد. هدف اصلی نویسنده رفع چالش موارد نفوذ به شبکه‌های رایانه‌ای با حدود پانصد میلیون رویداد ارتباطی روزانه است. (شبکه‌های در سطح شرکت‌ها و سازمان‌های بزرگ) که در یک کاربست واقعی، اعتبارسنجی این روش نیز انجام و ارائه شده است.

«تعیین مشخصات رفتار پویای گروهی در شبکه‌های اجتماعی برای علم سایبرنتیک» عنوان فصل پانزدهم این کتاب است که با توسعهٔ مفهوم «جوامع ویژه» در شبکه که منبعث از احساسات، علایق و گرایش‌های مشترک کاربران شبکه است، ضمن معرفی روش‌های محاسباتی برای برچسب‌زنی و کشف این علایق مشترک در شرایط پویا و متحرک، جوامع ویژه در شبکه را شناسایی می‌کند. ادعای نویسندگان این فصل، کاربرد تحقیق در شناسایی

گروه‌ها و جوامع ویژه‌ای است که ممکن است مقاصد خرابکارانه‌ای در سر داشته باشند.

فصل شانزدهم با «عنوان چند رویکرد برای تشخیص ناهنجاری‌های موجود در ترافیک داده‌های شبکه» ابتدا تشریح می‌کند که برخی از انواع حملات سایبری شامل ایجاد اختلال در ترافیک داده‌های شبکه است. سپس دو رویکرد عمده دربارهٔ روش‌های تشخیص به‌هنگام ناهنجاری‌های ترافیک شبکه‌های بزرگ مقیاس ارائه داده و در مثالی واقعی، حملات صورت‌گرفته به شبکهٔ یک آی‌ای‌پی فرانسوی و میزان کارایی این روش‌ها در تشخیص به موقع این حملات را مورد بررسی قرار داده است.

فصل پایانی نیز با عنوان «پایش یک دستگاه در شبکهٔ ارتباطات» به رفتار ناهنجار و خرابکارانه در شبکه با سطح تحلیل کاملاً خرد و در حد یک دستگاه منفرد متصل به شبکه تأکید دارد. الگوهای مدل‌سازی مبتنی بر یال و گره که با ساده‌سازی‌های متعدد و تکنیک‌های داده گاهی برای کاهش حجم محاسبات در خلال تشخیص اولیهٔ ساختارهای ناهنجار محلی در ترکیب با ابزارها و روش‌های پیچیده‌تر منجر به شناسایی و پایش دستگاه‌های مخرب در شبکه می‌شوند، هستهٔ اصلی مباحث تحقیق را شامل می‌گردد.

مطالعهٔ کتاب پیش رو به خواننده درک جدید و عمیق‌تری از داده‌های بزرگ و ماهیت و ظرفیت‌های بالقوهٔ آن داده و موجب ارتقاء مهارت‌های تحلیلی و آشنایی با کاربرد نظریهٔ گراف و دیگر نظریات و روش‌های محاسباتی پُرکاربرد در حوزهٔ مطالعات شبکه و تحلیل داده خواهد شد.

معاونت پژوهش و تولید علم

فهرست مطالب