

کاوش دادگان انبوه

Mining
of
Massive
Datasets

نویسنده‌گان:

Jure Leskovec
Anand Rajaraman
Jeffrey D. Ullman

مترجم:

دکتر مهدی اسماعیلی

طراحی جلد، حروفچینی و صفحه‌آرایی: همتا بیداریان
مدیر تولید و ناظر چاپ: حسین رعدشندی

کاوش دادگان انبوه

مترجم: دکتر مهدی اسماعیلی
ناشر: انتشارات آتی‌نگر
ناشر همکار: انتشارات وینا
چاپ اول، ۱۳۹۶
شمارگان: ۵۰۰ نسخه
قیمت: ۳۷۵,۰۰۰ ریال
شابک: ۹۷۸-۶۱-۷۶۳۱-۶۰۰

ISBN: 978-600-7631-61-4

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

نشانی دفتر فروش: خیابان جمالزاده جنوبی، رو به روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱
نامبر: ۶۶۵۶۵۳۳۶-۸
تلفن: ۰۶۵۶۵۳۳۷
www.ati-negar.com * info@ati-negar.com



ISBN: 978-600-7631-61-4

سرشناسه: لسکووتس، یوری Jure Leskovec
کاوش دادگان انبوه / نویسندهان [بوری لسکووتس، آناند راجارامان، جفری دی اولمن] مترجم مهدی اسماعیلی.
تهران: آتی‌نگر، وینا ۱۳۹۶
صف: ۴۷۲ ص: مصور، جدول، نمودار.

HF5415/۱۲۵ ک۲ ۱۳۹۶
۰۰۶/۳۱۲
۴۶۶۵۳۲۲

یادداشت: عنوان اصلی: Mining of Massive Datasets, 2 ed, 2015

موضوع: داده‌کاوی (Data mining)، داده‌های کلان (Big Data)

شناسه افزوده: راجارامان، آناند Anand Rajaraman

شناسه افزوده: اولمن، جفری دی.. Jeffrey D. Ullman

شناسه افزوده: اسماعیلی، مهدی، مترجم ۱۳۵۰ -

ردیبندی کنگره

ردیبندی دیجیتی

شماره کتابشناسی ملی

فهرست مطالب

۹

فصل اول: داده‌کاوی

۹	۱-۱ داده‌کاوی چیست؟
۱۳	۲-۱ حدود آمار برای فرآیند داده‌کاوی
۱۵	۳-۱ اشاره‌ای به چند مطلب مفید
۲۳	۴-۱ سازماندهی کتاب
۲۵	۵-۱ خلاصه فصل
۲۶	۶-۱ مراجع

۲۷

فصل دوم: متدولوژی

۲۸	۱-۲ سیستم فایل توزیع شده
۳۱	۲-۲ چارچوب MapReduce
۳۶	۳-۲ اجرای چند الگوریتم با MapReduce
۴۷	۴-۲ تعمیم‌هایی برای MapReduce
۵۲	۵-۲ مدلی برای هزینه ارتباطی
۵۹	۶-۲ نظریه پیچیدگی برای MapReduce
۷۲	۷-۲ خلاصه فصل
۷۵	۸-۲ مراجع

۷۷

فصل سوم: یافتن آیتم‌های مشابه

۷۸	۱-۳ کاربردهایی برای جستجوی آیتم‌های مشابه
۸۲	۲-۳ آماده‌سازی اسناد متنی
۸۵	۳-۳ تلخیص مجموعه‌ها همراه با حفظ تشابه
۹۱	۴-۳ درهم‌سازی حساس به محل
۹۶	۵-۳ سنجه‌های فاصله
۱۰۱	۶-۳ نظریه مربوط به توابع حساس به محل
۱۰۶	۷-۳ خانواده‌های LSH برای دیگر سنجه‌های فاصله
۱۱۲	۸-۳ کاربردهایی از توابع درهم‌ساز حساس به محل

۱۱۹.....	۹-۳ روش‌هایی برای درجه بالایی از شباهت
۱۲۷.....	۱۰-۳ خلاصه فصل
۱۳۰.....	۱۱-۳ مراجع

۱۳۱

فصل چهارم: کاوش داده‌های جریانی

۱۳۲.....	۱-۴ داده‌های جریانی
۱۳۶.....	۲-۴ نمونه‌گیری
۱۳۹.....	۳-۴ پالایش داده‌های جریانی
۱۴۶.....	۵-۴ تخمین گشتاورها
۱۵۱.....	۶-۴ شمارش تعداد یک‌های موجود در یک پنجره
۱۵۸.....	۷-۴ پنجره‌های وزندار
۱۶۰.....	۸-۴ خلاصه فصل
۱۶۲.....	۹-۴ مراجع

۱۶۵

فصل پنجم: تحلیل لینک

۱۶۵.....	۱-۵ الگوریتم PageRank
۱۷۸.....	۲-۵ محاسبه مؤثر PageRank
۱۸۴.....	۳-۵ و تأکید بر روی عنوان‌های خاص PageRank
۱۸۸.....	۴-۵ هرزلینک (افزایش مصنوعی PageRank)
۱۹۳.....	۵-۵ مراکز و مراجع
۱۹۸.....	۶-۵ خلاصه فصل
۲۰۱.....	۷-۵ مراجع

۲۰۳

فصل ششم: مجموعه اقلام مکرر

۲۰۴.....	۱-۶ مدل سبد خرید
۲۰۹.....	۲-۶ سبدهای خرید و الگوریتم Apriori
۲۱۸.....	۳-۶ مدیریت مجموعه داده‌های بزرگ‌تر در حافظه اصلی
۲۲۵.....	۴-۶ الگوریتم‌هایی با تعداد گذرهای محدود
۲۳۱.....	۵-۶ شمارش اقلام مکرر در داده‌های جریانی
۲۳۵.....	۶-۶ خلاصه فصل
۲۳۸.....	۷-۶ مراجع

فصل هفتم: خوشه‌بندی

۲۳۹

۲۳۹.	۱-۷ مقدمه‌ای بر تکنیک‌های خوشه‌بندی
۲۴۴.	۲-۷ خوشه‌بندی سلسله‌مراتبی
۲۵۳.	۳-۷ الگوریتم‌های k-means
۲۶۱.	۴-۷ الگوریتم CURE
۲۶۴.	۵-۷ خوشه‌بندی در فضاهای غیر‌اقلیدسی
۲۶۹.	۶-۷ خوشه‌بندی داده‌های جریانی
۲۷۴.	۷-۷ خلاصه فصل
۲۷۸.	۸-۷ مراجع

فصل هشتم: تبلیغ بر روی وب

۲۷۹

۲۷۹.	۱-۸ تبلیغ آنلاین
۲۸۲.	۲-۸ الگوریتم‌های آنلاین
۲۸۵.	۳-۸ مسئله تطبیق
۲۸۸.	۴-۸ Adwords
۲۹۸.	۵-۸ پیاده‌سازی Adwords
۳۰۲.	۶-۸ خلاصه فصل
۳۰۵.	۷-۸ مراجع

فصل نهم: سیستم‌های پیشنهاددهنده

۳۰۷

۳۰۸.	۱-۹ مدلی برای سیستم‌های پیشنهاددهنده
۳۱۲.	۲-۹ پیشنهادات مبتنی بر محتوا
۳۲۱.	۳-۹ فیلترینگ مشارکتی
۳۲۷.	۴-۹ کاهش ابعاد
۳۳۶.	۵-۹ چالش نیتفلیکس
۳۳۷.	۶-۹ خلاصه فصل
۳۳۹.	۷-۹ مراجع

فصل دهم: کاوش گراف شبکه‌های اجتماعی

۳۴۱

۳۴۲.	۱-۱۰ شبکه‌های اجتماعی همچون گراف‌ها
۳۴۷.	۲-۱۰ خوشه‌بندی گراف شبکه‌های اجتماعی

۳۵۵.	۳-۱۰ شناسایی مستقیم اجتماعات.....
۳۶۰.	۴-۱۰ افزار گرافها.....
۳۶۷.	۵-۱۰ یافتن اجتماعات همپوشانی.....
۳۷۳.	۶-۱۰ تکنیک simrank.....
۳۷۷.	۷-۱۰ شمارش مثلثها.....
۳۸۳.	۸-۱۰ خصوصیات همسایگی گرافها.....
۳۹۳.	۹-۱۰ خلاصه فصل.....
۳۹۷.	۱۰-۱۰ مراجع.....

۳۹۹

فصل یازدهم: کاهش ابعاد

۴۰۰.	۱-۱۱ مقادیر و بردارهای ویژه ماتریس‌های متقارن.....
۴۰۵.	۲-۱۱ تحلیل مؤلفه اصلی.....
۴۱۱.	۳-۱۱ تجزیه مقدار منفرد.....
۴۲۰.	۴-۱۱ تجزیه CUR.....
۴۲۶.	۵-۱۱ خلاصه فصل.....
۴۲۸.	۶-۱۱ مراجع.....

۴۲۹

فصل دوازدهم: یادگیری ماشین

۴۳۰.	۱-۱۲ مدل یادگیری ماشین.....
۴۳۷.	۲-۱۲ پرسپترون‌ها.....
۴۴۰.	۳-۱۲ ماشین‌های بردار پشتیبان.....
۴۴۲.	۴-۱۲ یادگیری از نزدیکترین همسایه‌ها.....
۴۴۹.	۵-۱۲ مقایسه روش‌های یادگیری.....
۴۵۰.	۶-۱۲ خلاصه فصل.....
۴۷۲.	۷-۱۲ مراجع.....