

بنام خدا

داده کاوی

مفاهیم و تکنیک‌ها

(ویراست سوم)

Data Mining

Concepts and Techniques

Jiawei Han
Micheline Kamber
Jian Pei

دکتر محمد علی بالافر

(عضو هیئت علمی دانشگاه تبریز)

سرشناسه	هان، جیاوی Han, Jiawei
عنوان و نام پدیدآور	داده‌کاوی: مفاهیم و تکنیک‌ها/ مولفین هان- کامبر-جین پی؛ [مترجم] محمدعلی بالاقر
وضعیت ویراست	ویراست ۳.
مشخصات نشر	تهران: ادبستان، ۱۳۹۳.
مشخصات ظاهری	۵۸۴ ص.: مصور، جدول، نمودار.
شابک	978-600-196-057-4
وضعیت فهرست نویسی	فیا
یادداشت	عنوان اصلی: Data mining: concepts and techniques, 3rd ed., c2011.
یادداشت	واژه‌نامه.
یادداشت	کتابنامه.
موضوع	داده‌کاوی
شماره افزوده	کمبر، میشلین
شماره افزوده	Kamber, Micheline
شماره افزوده	پی، جیان
شماره افزوده	(Pel, Jian, (Computer scientist
شماره افزوده	بالاقر، محمدعلی، ۱۳۵۴ - مترجم
رده کتاب‌نگره	۱۳۹۳ ۲۵۲۵/۷۶/۹۵۸
رده بندی دیسی	۰۰۶/۲۱۲
شماره کتابشناختی ملی	۲۵۳۲۲۲۵

نام کتاب	داده‌کاوی: مفاهیم و تکنیک‌ها (ویراست سوم)
مؤلفین	هان، جیاوی - کامبر، میشلین - جین پی
مترجم	دکتر محمدعلی بالاقر
ناشر	ادبستان
لیتوگرافی	فرانقش
چاپ	دیبا - نقشینه
نوبت چاپ	اول ۱۳۹۳
تیراژ	۵۰۰
قیمت	۲۶۰۰۰ تومان

دفتر مرکزی کتاب آیلار

انقلاب - خیابان شهید منیری جوادید (اردیبهشت) - خیابان شهدای ژاندارمری - رده ۱۴۶ (ساختمان آیلار) - تلفن: ۶۶۴۰۱۲۵۵ - دورنگار: ۶۶۴۹۴۴۳۱

فروشگاه شماره ۱ (کتاب آیلار)

انقلاب - روبروی دبیرخانه دانشگاه تهران - بازارچه کتاب. تلفن: ۶۶۹۵۶۴۳۳ - ۶۶۴۱۱۸۶۵

فروشگاه شماره ۲ (کتاب آیلار)

کریمخان زند - مابین ایرانشهر و خردمند جنوبی - شماره ۱۳۲، تلفن: ۸۸۳۱۹۷۴۰ - ۱

فروشگاه شماره ۳ (کتاب آیلار)

انقلاب - روبروی دانشگاه تهران - پاساژ فروزنده - شماره ۳۲۱، تلفن: ۶۶۹۶۳۵۹۶ - ۹۷

ISBN: 978-600-196-057-4

شابک: ۹۷۸-۶۰۰-۱۹۶-۰۵۷-۴

فهرست عناوین

فصل ۱: مقدمه‌ای بر داده کاوی	۱۷
فصل ۲: شناخت داده	۵۳
فصل ۳: پیش‌پردازش داده	۹۷
فصل ۴: انبار داده و پردازش تحلیل‌های برخط	۱۳۹
فصل ۵: تکنولوژی مکعب داده	۲۰۱
فصل ۶: روش الگوهای پرتکرار، انجمن‌ها، و همبستگی‌ها: مفاهیم اولیه و متدها	۲۵۷
فصل ۷: کاوش پیشرفته الگو	۲۹۳
فصل ۸: کلاس‌بندی: مفاهیم اولیه	۳۴۱
فصل ۹: کلاس‌بندی مبتنی بر شجره	۴۰۳
فصل ۱۰: تحلیل خوشه‌ها: مفاهیم اولیه و متدها	۴۴۱
فصل ۱۱: تحلیل‌های پیشرفته خوشه‌ها	۴۹۵
فصل ۱۲: تشخیص پرت‌ها	۵۴۱

فهرست مطالب

فصل ۱: مقدمه‌ای بر داده کاوی

۱۷	۱-۱ چرا داده کاوی؟
۱۷	۱-۱-۱ حرکت به سمت عصر اطلاعات
۱۹	۱-۱-۲ داده کاوی روندی در تکنولوژی اطلاعات
۲۲	۱-۲ داده کاوی چیست؟
۲۴	۱-۳ چرا نوع داده‌ای را می‌توان استخراج کرد؟
۲۵	۱-۳-۱ داده کاوی در پایگاه داده
۲۶	۱-۳-۲ انبارهای داده
۲۹	۱-۳-۳ داده تراکم‌یافته
۳۰	۱-۳-۴ دیگر نوع‌های داده
۳۱	۱-۴ الگوهای قابل استخراج
۳۲	۱-۴-۱ بیان کلاس/مفهوم: توصیف/تجزیه و تحلیل
۳۳	۱-۴-۲ استخراج الگوهای پرتکرار، انباشته و پیوستگی‌ها
۳۴	۱-۴-۳ کلاس‌بندی و رگرسیون در تحلیل‌های پیشگویی
۳۶	۱-۴-۴ تحلیل‌های کلاستر
۳۷	۱-۴-۵ تحلیل‌های پرت
۳۷	۱-۴-۶ آیا تمام الگوها مد نظر هستند؟
۴۰	۱-۵ تکنولوژی‌های بکار گرفته شده
۴۰	۱-۵-۱ آمار
۴۱	۱-۵-۲ یادگیری ماشین
۴۳	۱-۵-۳ سیستم‌های پایگاه داده و انبارهای داده
۴۴	۱-۵-۴ بازیابی اطلاعات
۴۵	۱-۶ کدام کاربردها مد نظر هستند
۴۵	۱-۶-۱ تجارت هوشمند
۴۵	۱-۶-۲ موتورهای جستجوی وب
۴۶	۱-۷ مباحث اصلی در داده کاوی

۴۶	۱-۷-۱ متدولوژی کاوش
۴۸	۱-۷-۲ تعامل کاربر
۴۸	۱-۷-۳ کارایی و مقیاس پذیری
۴۹	۱-۷-۴ تنوع پایگاه های داده
۴۹	۱-۷-۵ داده کاوی و جامعه
۵۰	۱-۸ خلاصه فصل
۵۱	۹- تمرینات

فصل ۲: شناخت داده

۵۴	۲-۱ اشیا، داده ها و اشیاء صفت
۵۴	۲-۱-۱ صفت های کیفی
۵۵	۲-۱-۲ صفات اشیاء
۵۶	۲-۱-۳ صفات باینری
۵۷	۲-۱-۴ صفات ترتیبی
۵۸	۲-۱-۵ صفات عددی
۵۹	۲-۱-۶ صفات گسسته و پیوسته
۵۹	۲-۲ توصیف آماری داده
۵۹	۲-۲-۱ سنجش تمایل مرکزی: میانگین، میانه و مد
۶۲	۲-۲-۲ سنجش پراکندگی داده: دامنه، چارک، واریانس، انحراف معیار، و دامنه میان چالاک
۶۶	۲-۲-۳ نمایش گرافیکی داده های آماری
۷۱	۲-۳ تجسم داده
۷۱	۲-۳-۱ تکنیک های تجسم pixel-oriented
۷۳	۲-۳-۲ تکنیک های تجسم geometric projection
۷۶	۲-۳-۳ تکنیک های تجسم icon-based
۷۷	۲-۳-۴ تکنیک های تجسم hierarchical
۷۹	۲-۳-۵ تجسم داده های پیچیده و روابط
۸۰	۲-۴ سنجش شباهت و عدم شباهت
۸۱	۲-۴-۱ Data matrix در برابر Dissimilarity matrix
۸۲	۲-۴-۲ سنجش های مجاورتی برای صفات اسمی

۸۴	۲-۴-۳	سنجش‌های مجاورتی برای صفات باینری
۸۵	۲-۴-۴	Minkowski distance: عدم شباهت داده عددی
۸۷	۲-۴-۵	سنجش‌های مجاورتی برای صفات ترتیبی
۸۹	۲-۴-۶	عدم شباهت برای صفاتی از نوع ترکیبی
۹۰	۲-۴-۷	شباهت کسینوسی
۹۳	۲-۵	خلاصه فصل
۹۳	۲-۶	تمرینات

فصل ۳: پیش‌پردازش داده

۹۸	۳-۱	پیش‌پردازش داده (مقدمه)
۹۸	۳-۱-۱	کیفیت داده: (پیش‌پردازش داده؟)
۹۹	۳-۱-۲	وظایف اصلی پیش‌پردازش داده
۱۰۲	۳-۲	پاک‌سازی داده
۱۰۳	۳-۲-۱	مقادیر مفقوده
۱۰۴	۳-۲-۲	داده مزاحم
۱۰۶	۳-۲-۳	پاک‌سازی داده بعنوان یک فرآیند
۱۰۹	۳-۳	یکپارچه‌سازی داده
۱۰۹	۳-۳-۱	مشکل شناسایی موجودیت
۱۱۰	۳-۳-۲	تحلیل‌های افزونگی و همبستگی
۱۱۴	۳-۳-۳	تکثر تاپل
۱۱۴	۳-۳-۴	تشخیص تداخل مقدار داده و رفع آن
۱۱۵	۳-۴	تقلیل داده
۱۱۵	۳-۴-۱	نگاهی به راهبردهای تقلیل داده
۱۱۶	۳-۴-۲	Wavelet Transforms
۱۱۷	۳-۴-۳	Principal Components Analysis
۱۱۸	۳-۴-۴	Attribute Subset Selection
۱۲۰	۳-۴-۵	مدل‌های رگرسیون و log-linear: تقلیل داده پارامتری
۱۲۱	۳-۴-۶	هیستوگرام‌ها
۱۲۳	۳-۴-۷	خوشه‌بندی

۱۲۳	۳-۴-۸ نمونه برداری
۱۲۵	۳-۴-۹ مکعب تجميع داده
۱۲۷	۳-۵ تغییر شکل دادن و گسسته کردن داده
۱۲۷	۳-۵-۱ مروری بر راهبردهای تغییر شکل دادن داده
۱۲۸	۳-۵-۲ تغییر شکل داده از طریق نرمال سازی
۱۳۰	۳-۵-۳ گسسته سازی از طریق binning
۱۳۰	۳-۵-۴ گسسته سازی از طریق تحلیل های هیستوگرام
۱۳۱	۳-۵-۵ گسسته سازی توسط خوشه، درخت تصمیم گیری، و تحلیل های همبستگی
۱۳۲	۳-۵-۶ تولید سلسله مراتب مفهوم برای داده اسمی
۱۳۴	۳-۶ خلاصه نویسی
۱۳۵	۳-۷ تمرینات

فصل ۴: انبار داده و پردازش تحلیل های برخط

۱۴۰	۴-۱ انبار داده: مفاهیم مقدماتی
۱۴۰	۴-۱-۱ انبار داده چیست؟
۱۴۲	۴-۱-۲ تفاوت بین سیستم های پایگاه داده عملیاتی
۱۴۴	۴-۱-۳ چرا انبار داده مجزا؟
۱۴۵	۴-۱-۴ انبار داده: یک معماری چند لایه
۱۴۷	۴-۱-۵ مدل های انبار داده: انبار داده سازمانی، داده باقی مانده و انبار مجازی
۱۴۹	۴-۱-۶ استخراج، تغییر شکل و بار شدن
۱۴۹	۴-۱-۷ مخزن متادیتا
۱۵۱	۴-۲ مدل سازی انبار داده: مکعب داده و OLAP
۱۵۱	۴-۲-۱ مکعب داده: مدل داده چند بعدی
۱۵۴	۴-۲-۲ ستاره ای، دانه برفی، و منظومه ای: طرح های برای مدل های داده چند بعدی
۱۵۸	۴-۲-۳ ابعاد: نقش سلسله مراتب مفهومی
۱۶۰	۴-۲-۴ سنجش: رده بندی و محاسبات
۱۶۱	۴-۲-۵ عملیات های نمونه OLAP
۱۶۴	۴-۲-۶ مدل پرس و جو استارنت برای پرس و جو از پایگاه داده های چند بعدی
۱۶۵	۴-۳ طراحی و کاربرد انبار داده

۱۶۵	۴-۳-۱ چارچوب تحلیل تجاری در طراحی انبار داده
۱۶۷	۴-۳-۲ فرآیند طراحی انبار داده
۱۶۸	۴-۳-۳ کاربرد انبار داده در پردازش اطلاعات
۱۷۰	۴-۳-۴ از پردازش تحلیل برخط به داده کاوی چند بعدی
۱۷۱	۴-۴ پیاده‌سازی انبار داده
۱۷۱	۴-۴-۱ محاسبات کارآمد مکعب داده: مقدمه
۱۷۵	۴-۴-۲ شاخص‌گذاری داده OLAP: شاخص Bitmap و Join
۱۷۸	۴-۴-۳ پردازش کارآمد پرس‌وجوهای OLAP
۱۷۹	۴-۴-۴ معماری سرور OLAP: ROLAP در برابر MOLAP در برابر HOLAP
۱۸۱	۴-۵ تعمیم داده: ظرفیت استقراء صفت-گرا
۱۸۳	۴-۵-۱ استقراء صفت-گرا برای خصوصیات داده
۱۸۹	۴-۵-۲ پیاده‌سازی استقراء صفت-گرا
۱۹۱	۴-۵-۳ استقراء صفت-گرا برای دسته‌بندی کلاس
۱۹۵	۴-۶ خلاصه فصل
۱۹۶	۴-۷ تمرینات

فصل ۵: تکنولوژی مکعب داده

۲۰۲	۵-۱ محاسبه مکعب داده: مفاهیم مقدماتی
۲۰۲	۵-۱-۱ تجسم مکعب: مکعب‌های Full, Iceberg, Closed و Star
۲۰۷	۵-۱-۲ راهبردهای عمومی برای محاسبه مکعب داده
۲۰۹	۵-۲ متدهای محاسبه مکعب داده
۲۱۰	۵-۲-۱ متد Multiway array aggregation برای محاسبه مکعب کامل
۲۱۵	۵-۲-۲ BUC: محاسبه مکعب‌های Iceberg از مستطیل Apex به سمت پایین
۲۱۵	۵-۲-۳ الگوریتم Star-Cubing: محاسبه مکعب‌های iceberg با استفاده از ساختار دینامیکی
۲۱۹	Star-tree
۲۲۶	۵-۲-۴ پیش محاسبه Shell fragments برای OLAP سریع با ابعاد بالا
۲۳۴	۵-۳ پردازش پرس‌وجوهای پیشرفته بواسطه تکنولوژی مکعب
۲۳۵	۵-۳-۱ مکعب‌های نمونه: کاوش برپایه OLAP بر روی داده نمونه
۲۴۱	۵-۳-۲ مکعب‌های رتبه‌بندی: محاسبات کارآمد از پرس‌وجوهای top-k

۲۴۴	۵-۴ تحلیل داده چند بعدی در فضای مکعب
۲۴۴	۵-۴-۱ مکعب‌های پیشگو: پیشگویی در فضای مکعب
۲۴۶	۵-۴-۲ مکعب‌های multifeature: تجمیع پیچیده در دانه‌بندی‌های متعدد
۲۴۸	۵-۴-۳ کاوش فضای مکعب به شیوه مبتنی بر استثنا، discovery-driven
۲۵۲	۵-۵ خلاصه فصل
۲۵۳	۵-۶ تمرینات

فصل ۶: کاوش الگوهای پرتکرار، انجمن‌ها، و همبستگی‌ها: مفاهیم اولیه و متدها

۲۵۸	۶-۱ مفاهیم اولیه
۲۵۸	۶-۱-۱ تحلیل سبب و اثر
۲۶۰	۶-۱-۲ مجموعه ایتیم‌ها، پرتکرار، محصور و قوانین انجمنی
۲۶۳	۶-۲ متدهای کاوش مجموعه ایتیم‌ها
۲۶۴	۶-۲-۱ Apriori: یافتن مجموعه‌های ایتیم تکرار از طریق محدود کردن تولید کاندیدا
۲۶۹	۶-۲-۲ تولید قوانین انجمنی از مجموعه ایتیم پرتکرار
۲۷۰	۶-۲-۳ بهبود کارایی Apriori
۲۷۲	۶-۲-۴ شیوه الگوی رشد برای کاوش مجموعه‌های ایتیم پرتکرار
۲۷۶	۶-۲-۵ کاوش مجموعه‌های ایتیم پرتکرار با استفاده از روش‌های عمودی
۲۷۸	۶-۲-۶ کاوش محصور و الگوهای max
۲۸۰	۶-۳ کدام الگوها جالب‌تر هستند؟ متدهای ارزیابی الگو
۲۸۰	۶-۳-۱ قوانین قوی ضرورتاً جالب توجه نیستند
۲۸۱	۶-۳-۲ از تحلیل‌های انجمنی به تحلیل‌های همبستگی
۲۸۳	۶-۳-۳ مقایسه معیارهای ارزیابی الگو
۲۸۸	۶-۴ خلاصه فصل
۲۸۹	۶-۵ تمرینات

فصل ۷: کاوش پیشرفته الگو

۲۹۳	۷-۱ کاوش الگو: نقشه راه
-----	-------------------------

۲۹۷	۷-۲ کاوش الگو در فضای چند سطحی، چند بعدی
۲۹۸	۷-۲-۱ کاوش انجمن‌های چند سطحی
۳۰۲	۷-۲-۲ کاوش انجمن‌های چند بعدی
۳۰۳	۷-۲-۳ کاوش قوانین انجمنی کمی
۳۰۶	۷-۲-۴ کاوش الگوهای نادر و الگوهای منفی
۳۰۹	۷-۳ کاوش الگوی پرتکرار مبتنی بر محدودیت
۳۱۰	۷-۳-۱ کاوش قوانین انجمنی با Metarule-guided
۳۱۱	۷-۳-۲ تولید الگو مبتنی بر محدودیت: هرس فضای الگو و هرس فضای داده
۳۱۷	۷-۴ کاوش داده با ابعاد بالا و الگوهای بسیار بزرگ
۳۱۸	۷-۴-۱ کاوش الگوهای بسیار بزرگ با Pattern-Fusion
۳۲۳	۷-۵ کاوش فشرده‌سازی الگوهای تقریبی
۳۲۴	۷-۵-۱ کاوش الگوهای فشرده‌سازی، به شرط خوشه‌بندی الگو
۳۲۷	۷-۵-۲ استخراج k الگو برتر از مجموعه داده از افزونه
۳۲۹	۷-۶ کاوش الگو و کاربرد
۳۲۹	۷-۶-۱ یادداشت معنایی از الگوهای تکرار
۳۳۴	۷-۶-۲ کاربردهای کاوش الگو
۳۳۷	۷-۷ خلاصه فصل
۳۳۸	۷-۸ تمرینات

فصل ۸: کلاس‌بندی: مفاهیم اولیه

۳۴۱	۸-۱ مفاهیم اولیه
۳۴۱	۸-۱-۱ کلاس‌بندی چیست؟
۳۴۲	۸-۱-۲ شیوه رایج در کلاس‌بندی
۳۴۴	۸-۲ استنتاج درخت تصمیم
۳۴۶	۸-۲-۱ استنتاج درخت تصمیم
۳۵۰	۸-۲-۲ معیارهای انتخاب صفت
۳۵۷	۸-۲-۳ هرس کردن درخت
۳۶۰	۸-۲-۴ مقیاس‌پذیری و استنتاج درخت تصمیم
۳۶۲	۸-۲-۵ کاوش ویژگی‌ها برای استنتاج درخت تصمیم

۳۶۳	۸-۳ متدهای کلاس‌بندی Bayesian
۳۶۴	۸-۳-۱ قضیه Bayes
۳۶۵	۸-۳-۲ کلاس‌بندی Naïve Bayesian
۳۶۹	۸-۴ کلاس‌بندی مبتنی بر قانون
۳۶۹	۸-۴-۱ استفاده از قوانین IF-THEN برای کلاس‌بندی
۳۷۲	۸-۴-۲ استخراج رل از درخت تصمیم
۳۷۳	۸-۴-۳ استنتاج رل با استفاده از الگوریتم پوشش ترتیبی
۳۷۸	۸-۵ ارزیابی و انتخاب مدل
۳۷۹	۸-۵-۱ معیارهای ارزیابی کارایی کلاس‌بندی کننده
۳۸۵	۸-۵-۲ Random Subsampling و Bootstrap
۳۸۵	۸-۵-۳ Cross-Validation
۳۸۶	۸-۵-۴ Bootstrapping
۳۸۷	۸-۵-۵ انتخاب مدل با استفاده از تست‌های آماری
۳۸۹	۸-۵-۶ مقایسه Classifier با هزینه فایده و منحنی‌های ROC
۳۹۳	۸-۶ تکنیک‌های برای بهبود مدل کلاس‌بندی
۳۹۳	۸-۶-۱ معرفی متدهای Ensemble
۳۹۴	۸-۶-۲ Bagging
۳۹۶	۸-۶-۳ Boosting و AdaBoost
۳۹۸	۸-۶-۴ Random Forests
۳۹۸	۸-۷ خلاصه فصل
۴۰۰	۸-۸ تمرینات

فصل ۹: کلاس‌بندی: متدهای پیشرفته

۴۰۳	۹-۱ Bayesian Belief Networks
۴۰۴	۹-۱-۱ مفاهیم و مکانیزم‌ها
۴۰۶	۹-۱-۲ آموزش شبکه‌های Bayesian Belief
۴۰۸	۹-۲ کلاس‌بندی با Backpropagation
۴۰۹	۹-۲-۱ شبکه عصبی چند لایه feed-forward
۴۱۰	۹-۲-۲ تعریف توپولوژی شبکه

۴۱۱	Backpropagation ۹-۲-۳
۴۱۷	درون جعبه سیاه: پس انتشار و تفسیر ۹-۲-۴
۴۱۸	Support Vector Machines ۹-۳
۴۱۹	۹-۳-۱ حالتی که داده بفرم خطی جدانشدنی هستند
۴۲۴	۹-۳-۲ حالتی که داده بفرم خطی جدانشدنی هستند
۴۲۵	۹-۴ کلاس بندی با الگوهای پرتکرار
۴۲۶	۹-۴-۱ کلاس بندی انجمنی
۴۲۹	۹-۴-۲ کلاس بندی افتراقی مبتنی بر الگوهای پرتکرار
۴۳۳	۹-۵ یادگیری تنبیل (یا یادگیری از همسایه ها)
۴۳۳	k-nearest-neighbor classifier ۹-۵-۱
۴۳۶	Case-Based Reasoning ۹-۵-۲
۴۳۶	۹-۶ خلاصه فصل
۴۳۸	۹-۷ تمرینات

فصل ۱۰: تحلیل خوشه: مفاهیم و مدها

۴۴۲	۱۰-۱ تحلیل خوشه
۴۴۲	۱۰-۱-۱ تحلیل خوشه چیست؟
۴۴۴	۱۰-۱-۲ الزامات تحلیل خوشه
۴۴۷	۱۰-۱-۳ نگاهی بر متدهای اولیه خوشه بندی
۴۵۰	۱۰-۲ متدهای پارتیشن بندی
۴۵۰	۱۰-۲-۱ k-means تکنیک مبتنی بر مرکز ثقل
۴۵۳	۱۰-۲-۲ k-medoids تکنیک نمایندگی مبتنی بر شی
۴۵۶	۱۰-۳ متدهای سلسله مراتبی
۴۵۸	۱۰-۳-۱ خوشه بندی توده شده در برابر خوشه بندی تقسیمی
۴۶۰	۱۰-۳-۲ معیارهای فاصله در متدهای الگوریتمی
۴۶۱	۱۰-۳-۳ BIRCH: خوشه بندی سلسله مراتبی چندفازی با استفاده از درختان CF
۴۶۴	۱۰-۳-۴ Chameleon: خوشه بندی سلسله مراتبی چند فازی با استفاده از مدل سازی پویا
۴۶۶	۱۰-۳-۵ خوشه بندی سلسله مراتبی احتمالی
۴۷۰	۱۰-۴ متدهای مبتنی بر چگالی

۴۷۰	۱۰-۴-۱: DBSCAN خوشه‌بندی مبتنی بر چگالی برپایه نواحی متصله با چگالی بالا
۴۷۳	۱۰-۴-۲: OPTICS: مرتب‌سازی نقاط برای شناسایی ساختار خوشه
۴۷۶	۱۰-۴-۳: DENCLUE: خوشه‌بندی مبتنی بر توابع توزیع چگالی
۴۷۸	۱۰-۵: متدهای مبتنی بر گرید
۴۷۹	۱۰-۵-۱: STING: STatistical InforMation Grid
۴۸۱	۱۰-۵-۲: CLIQUE: متد خوشه‌بندی زیرفضا همانند apriori
۴۸۳	۱۰-۶: ارزیابی خوشه‌بندی
۴۸۴	۱۰-۶-۱: ارزیابی گرایش خوشه‌بندی
۴۸۶	۱۰-۶-۲: تعیین تعداد خوشه‌ها
۴۸۷	۱۰-۶-۳: اثبات کیفیت خوشه‌بندی
۴۹۰	۱۰-۷: خلاصه فصل
۴۹۲	۱۰-۸: تمرینات

فصل ۱۱: تحلیل‌های پیشرفته خوشه‌بندی

۴۹۵	۱۱-۱: خوشه‌بندی مبتنی بر مدل احتمالی
۴۹۷	۱۱-۱-۱: خوشه‌های فازی
۴۹۹	۱۱-۱-۲: خوشه‌های مبتنی بر مدل احتمالی
۵۰۳	۱۱-۱-۳: الگوریتم Expectation-Maximization
۵۰۶	۱۱-۲: خوشه‌بندی داده با ابعاد بالا
۵۰۷	۱۱-۲-۱: خوشه‌بندی داده‌هایی با ابعاد بالا: مشکلات، چالش‌ها و راهکارهای رایج
۵۰۹	۱۱-۲-۲: متدهای خوشه‌بندی زیرفضا
۵۱۰	۱۱-۲-۳: Biclustering
۵۱۸	۱۱-۲-۴: متدهای کاهش ابعاد و خوشه‌بندی طیفی
۵۲۰	۱۱-۳: خوشه‌بندی گراف و داده شبکه
۵۲۱	۱۱-۳-۱: کاربردها و چالش‌ها
۵۲۳	۱۱-۳-۲: معیارهای شباهت
۵۲۷	۱۱-۳-۳: متدهای خوشه‌بندی گراف
۵۳۱	۱۱-۴: خوشه‌بندی با قیدها
۵۳۲	۱۱-۴-۱: رده‌بندی قیدها

۵۳۴	۱۱-۴-۲ متدهای مربوط به خوشه‌بندی با قیدها
۵۳۷	۱۱-۵ خلاصه فصل
۵۳۹	۱۱-۶ تمرینات

فصل ۱۲: تشخیص پرت‌ها

۵۴۲	۱۲-۱ پرت‌ها و تحلیل پرت‌ها
۵۴۲	۱-۱ پرت‌ها چیستند؟
۵۴۳	۱-۲ انواع پرت‌ها
۵۴۶	۱-۳ چالش‌ها در تشخیص پرت
۵۴۷	۱۲-۲ متدهای تشخیص پرت
۵۴۷	۱۲-۲-۱ متدهای طارقی و نیمه نظارتی و بدون نظارت
۵۴۹	۱۲-۲-۲ متدهای آماری، مبتنی بر مجاورت و متدهای مبتنی بر خوشه‌بندی
۵۵۱	۱۲-۳ شیوه‌های آماری
۵۵۲	۱۲-۳-۱ متدهای پارامتریک
۵۵۶	۱۲-۳-۲ متدهای غیرپارامتریک
۵۵۸	۱۲-۴ شیوه‌های مبتنی بر مجاورت
۵۵۹	۱۲-۴-۱ تشخیص پرت مبتنی بر فاصله و متدهای مبتنی بر رتبه
۵۶۰	۱۲-۴-۲ متد مبتنی بر گرید
۵۶۲	۱۲-۴-۳ تشخیص پرت مبتنی بر تراکم
۵۶۶	۱۲-۵ شیوه‌های مبتنی بر خوشه‌بندی
۵۶۹	۱۲-۶ شیوه‌های مبتنی بر کلاس‌بندی
۵۷۲	۱۲-۷ کاوش پرت‌های ضمنی و جمعی
۵۷۲	۱۲-۷-۱ تبدیل تشخیص پرت ضمنی به تشخیص پرت متعارف
۵۷۳	۱۲-۷-۲ مدل‌سازی رفتار نرمال با توجه به زمینه‌ها
۵۷۴	۱۲-۷-۳ کاوش پرت‌های جمعی
۵۷۵	۱۲-۸ تشخیص پرت در داده با ابعاد بالا
۵۷۶	۱۲-۸-۱ بسط روش متعارف تشخیص پرت
۵۷۷	۱۲-۸-۲ یافتن پرت‌ها در زیرفضاها

۵۷۸.....	۱۲-۸-۳ مدل سازی پرت ها با ابعاد بالا
۵۸۰.....	۱۲-۹ خلاصه فصل
۵۸۱.....	۱۲-۱۰ تمرینات

www.ketab.ir