

شناسنامه کتاب

عنوان کتاب : دسته بندی به کمک تکنیک های خوشه بندی سلسله مراتبی

عنوان کتاب به انگلیسی : Classification by hierarchical clustering

نام پدیدآورنده : حسین هشیارمنش

ویراستاران : سید علی موسویان جندقی، علی هشیارمنش

نام ناشر : نشر شار

شابک : 978-600-6165-07-3

وضعیت فهرست نویسی : بر اساس اطلاعات فیبا

رده بندی کنگره : QA۲۷۸۵۵د۵۱۷۳۹۲

رده بندی دیویی : ۵۹۱/۵۳

شماره کتابشناسی ملی : ۳۱۲۵۰۶۸

سال انتشار : چاپ اول اردیبهشت ماه ۱۳۹۲

مشخصات ظاهری : ۸۰ ص: مصور، جدول، نمودار.

شمارگان : ۱۰۰ نسخه

قیمت : ۵۰۰۰۰ ریال

انتشارات شار (ناشر کتاب های دانشگاهی سری شار)



آدرس : تهران، خ انقلاب، خ دانشگاه، خ شهدای ژاندارمری، پ ۵۸ (ساختمان ایرانیان)، واحد ۱۰

تلفن انتشارات : ۶۶۴۸۴۲۹۳ - ۴ فکس : ۶۶۴۰۵۲۰۸

آدرس سایت اینترنتی انتشارات : <http://www.shar.ir/?Publications>

آدرس پست الکترونیکی : Info@shar.ir

تمام حقوق این اثر برای انتشارات شار محفوظ است

پیش‌گفتار

امروزه با توجه به گسترش علوم مختلف برخی از تکنیک‌ها و روش‌های یک علم به دلیل توانایی در برطرف کردن نیازهای علوم مختلف، کاربرد وسیع و گسترده‌ای پیدا کرده‌اند. یکی از این تکنیک‌ها، خوشه‌بندی است. خوشه‌بندی یکی از تکنیک‌های آماری برای تلخیص اطلاعات است که امروزه به دلیل توانایی و قابلیت‌های بالایی که در تلخیص اطلاعات و دسته‌بندی آن‌ها دارد مورد توجه محققان و پژوهشگران علوم دیگر قرار گرفته است. وسعت و گسترش این تکنیک را می‌توان در نرم افزارهای غیر آماری دید که یکی از گزینه‌های آن، گزینه خوشه‌بندی است. برخی از کاربردهای تکنیک خوشه‌بندی در علوم دیگر می‌توان موارد زیر را اشاره کرد: در دوربین عکاسی می‌توان در کنار حفظ کیفیت عکس حجم آن را کاهش داد، دلیل و یا دلایل مرگ ناشی از بیماری خاص را شناسایی کرد، کارایی هر دستگاه صنعتی را بیشتر کرد. با تمام این مزیت‌ها چنانکه بدون توجه به ماهیت هر یک از روش‌های خوشه‌بندی از آن‌ها استفاده کنیم می‌تواند خسارات جبران ناپذیری بر جای گذارد.

در این کتاب به دلیل فراگیر شدن استفاده از تکنیک‌های خوشه‌بندی در علوم مختلف و برای جلوگیری از بروز خسارت‌ها در استفاده از روش‌های خوشه‌بندی سلسله مراتبی سعی شده است هر یک از روش‌های خوشه‌بندی سلسله مراتبی را که در علوم مختلف مورد استفاده محققان و پژوهشگران قرار می‌گیرد را به زبانی ساده بیان کرده و نحوه اجرای هر یک به صورت شهودی شرح داده شده است. در این کتاب سعی شده است خلاء وجود کتابی که به زبان ساده بتواند برای تمامی محققان و پژوهشگران علوم مختلف، روش‌های خوشه‌بندی سلسله مراتبی را شرح دهد، برطرف شود. امید است این کتاب بتواند محققان و پژوهشگران علوم مختلف را در استفاده مناسب از هر یک از روش‌های خوشه‌بندی سلسله مراتبی، برای پیشبرد علم، یاری رساند. در

اینجا بر خود واجب می‌دانم از آموزش‌ها، راهنمایی‌ها، و حمایت‌های بی‌دریغ جناب آقای دکتر محمدپور، که سبب نوشتن و چاپ این کتاب شده است، صمیمانه تشکر و قدردانی کنم. از آنجایی که این کتاب خالی از اشکال نیست لذا از تمامی خوانندگان محترم تقاضا دارم اطلاعات تکمیلی خود و کاستی‌ها و نارسایی‌های این کتاب را با آدرس پست الکترونیکی h.clustering@gmail.com در میان بگذارند تا در چاپ بعدی لحاظ شود.

حسین هشیارمنش

بهار ۱۳۹۲

فهرست مطالب

۱	روش نزدیکترین فاصله	۱
۳	۱.۱ مثال اول	
۸	۲.۱ مثال دوم	
۱۱	روش دورترین فاصله	۲
۱۲	۱.۲ مثال اول	
۱۶	۲.۲ مثال دوم	
۱۹	روش پیوند میانگین	۳
۲۱	۱.۳ مثال اول	
۲۲	۲.۳ مثال دوم	
۲۷	روش پیوند مبانه	۴
۲۸	۱.۴ مثال اول	
۳۱	۲.۴ مثال دوم	
۳۳	روش پیوند به مرکز	۵
۳۴	۱.۵ مثال اول	
۳۷	۲.۵ مثال دوم	
۳۹	دسته‌بندی به کمک خوشه‌بندی	۶
۴۰	۱.۶ تعیین تعداد خوشه بهینه با استفاده از توزیع فاصله بین خوشه‌ها	
۴۱	۱.۱.۶ معرفی روش	

۴۳	۷ روش برنامه نویسی خوشه‌بندی سلسله مراتبی در R
۴۴	۱.۷ تعاریف لازم
۵۰	۲.۷ خوشه‌بندی سلسله مراتبی در R
۵۲	۳.۷ تعیین تعداد خوشه
۵۳	۱.۳.۷ دستورهای تعیین تعداد خوشه در R
۵۴	۲.۳.۷ تعیین تعداد خوشه با استفاده از توزیع فاصله بین خوشه‌ها
۵۶	واژه نامه فارسی به انگلیسی

www.ketab.ir

فهرست شکل‌ها

۵	۱۰۱	نمایش نمرات درس ریاضی کلاس الف بر روی صفحه جهت بررسی فاصله بین مشاهدات
۷	۲۰۱	نمایش درختی خوشه‌بندی نمرات درس ریاضی کلاس الف به روش نزدیک‌ترین فاصله. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد.
۹	۳۰۱	نمایش نمرات دروس ریاضی و آمار کلاس الف بر روی صفحه جهت بررسی فاصله بین مشاهدات.
۱۰	۴۰۱	نمایش درختی خوشه‌بندی افراد کلاس الف بر اساس نمرات درس ریاضی و آمار به روش نزدیک‌ترین فاصله. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد.
۱۵	۱۰۲	نمایش درختی خوشه‌بندی نمرات درس ریاضی کلاس الف به روش دورترین فاصله. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد.
۱۷	۲۰۲	نمایش درختی خوشه‌بندی افراد کلاس الف بر اساس نمرات درس ریاضی و آمار به روش دورترین فاصله. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد.
۲۴	۱۰۳	نمایش درختی خوشه‌بندی نمرات درس ریاضی کلاس الف به روش پیوند میانگین. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد.

- ۲۰۳ نمایش درختی خوشه‌بندی افراد کلاس الف بر اساس نمرات درس ریاضی و آمار به روش پیوند میانگین. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد. ۲۵
- ۱۰۴ نمایش درختی خوشه‌بندی نمرات درس ریاضی کلاس الف به روش پیوند میانه. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد. ۳۰
- ۲۰۴ نمایش درختی خوشه‌بندی افراد کلاس الف بر اساس نمرات درس ریاضی و آمار به روش پیوند میانه. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد. ۳۲
- ۱۰۵ نمایش درختی خوشه‌بندی نمرات درس ریاضی کلاس الف به روش پیوند به مرکز. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد. ۳۶
- ۲۰۵ نمایش درختی خوشه‌بندی افراد کلاس الف بر اساس نمرات درس ریاضی و آمار به روش پیوند به مرکز. در این نمودار می‌توان مراحل خوشه‌بندی، نحوه ادغام‌ها و فاصله بین خوشه‌ها را ملاحظه کرد. ۳۸

فهرست جدول‌ها

۴	۱.۱	نمره دانش آموزان کلاس الف از درس ریاضی
	۲.۱	ماتریس فاصله. محاسبه فاصله دو به دوی نمرات درس ریاضی کلاس الف با استفاده از متر اقلیدسی.
۵	۳.۱	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۹ و ۲۰، به روش نزدیک‌ترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۶	۴.۱	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۳ و ۱۴، به روش نزدیک‌ترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۶	۵.۱	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۶ و خوشه {۱۳، ۱۴} به روش نزدیک‌ترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۶	۶.۱	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو خوشه {۱۹، ۲۰} و {۱۳، ۱۴، ۱۶} به روش نزدیک‌ترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۶	۷.۱	مراحل خوشه‌بندی، ادغام مشاهدات و خوشه‌ها، فاصله بین خوشه‌ها بر اساس خوشه‌بندی به روش نزدیک‌ترین فاصله بر روی نمرات درس ریاضی کلاس الف.
۷	۸.۱	نمره دانش آموزان کلاس الف از درس ریاضی و آمار.
۹	۹.۱	ماتریس فاصله. محاسبه فاصله دو به دوی نمرات درس ریاضی و آمار کلاس الف با استفاده از متر اقلیدسی.
۹	۱.۲	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۹ و ۲۰، به روش دورترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۱۳	۲.۲	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۳ و ۱۴، به روش دورترین فاصله، تشکیل یک خوشه جدید داده‌اند.
۱۴	۳.۲	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۶ و خوشه {۱۳، ۱۴} به روش دورترین فاصله، تشکیل یک خوشه جدید داده‌اند.

۴.۲	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۰ و خوشه {۱۳، ۱۴، ۱۶} به روش دورترین فاصله، تشکیل یک خوشه جدید داده‌اند.	۱۴
۵.۲	مراحل خوشه‌بندی، ادغام مشاهدات و خوشه‌ها و فاصله بین خوشه‌ها بر اساس خوشه‌بندی به روش دورترین فاصله بر روی نمرات درس ریاضی کلاس الف	۱۴
۱.۳	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۹ و ۲۰، به روش پیوند میانگین، تشکیل یک خوشه جدید داده‌اند.	۲۲
۲.۳	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۳ و ۱۴، به روش پیوند میانگین، تشکیل یک خوشه جدید داده‌اند.	۲۳
۳.۳	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۶ و خوشه {۱۳، ۱۴} به روش پیوند میانگین، تشکیل یک خوشه جدید داده‌اند.	۲۳
۴.۳	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۰ و خوشه {۱۳، ۱۴، ۱۶} به روش پیوند میانگین، تشکیل یک خوشه جدید داده‌اند.	۲۳
۵.۳	مراحل خوشه‌بندی، ادغام مشاهدات و خوشه‌ها و فاصله بین خوشه‌ها بر اساس خوشه‌بندی به روش پیوند میانگین بر روی نمرات درس ریاضی کلاس الف.	۲۳
۱.۴	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۹ و ۲۰، به روش پیوند میانه، تشکیل یک خوشه جدید داده‌اند.	۲۹
۲.۴	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۳ و ۱۴، به روش پیوند میانه، تشکیل یک خوشه جدید داده‌اند.	۲۹
۳.۴	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۶ و خوشه {۱۳، ۱۴} به روش پیوند میانه، تشکیل یک خوشه جدید داده‌اند.	۲۹
۴.۴	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که خوشه {۱۹، ۲۰} و خوشه {۱۳، ۱۴، ۱۶} به روش پیوند میانه، تشکیل یک خوشه جدید داده‌اند.	۳۰
۵.۴	مراحل خوشه‌بندی، ادغام مشاهدات و خوشه‌ها و فاصله بین خوشه‌ها بر اساس خوشه‌بندی به روش پیوند میانه بر روی نمرات درس ریاضی کلاس الف.	۳۱
۱.۵	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۹ و ۲۰، به روش پیوند به مرکز، تشکیل یک خوشه جدید داده‌اند.	۳۵
۲.۵	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که دو مشاهده ۱۳ و ۱۴، به روش پیوند به مرکز، تشکیل یک خوشه جدید داده‌اند.	۳۵
۳.۵	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۶ و خوشه {۱۳، ۱۴} به روش پیوند به مرکز، تشکیل یک خوشه جدید داده‌اند.	۳۶
۴.۵	محاسبه ماتریس فاصله نمرات درس ریاضی کلاس الف زمانی که مشاهده ۱۰ و خوشه {۱۳، ۱۴، ۱۶} به روش پیوند به مرکز، تشکیل یک خوشه جدید داده‌اند.	۳۶

- ۵.۵ مراحل خوشه‌بندی، ادغام مشاهدات و خوشه‌ها و فاصله بین خوشه‌ها بر اساس خوشه‌بندی به روش پیوند به مرکز بر روی نمرات درس ریاضی کلاس الف ۳۷

www.ketab.ir

www.ketab.ir

مقدمه

در هر دانشگاه تعدادی دانشجوی در رشته‌های مختلف مشغول به تحصیل هستند. در بانک اطلاعاتی هر دانشگاه سال ورود، نمرات هر درس، کم‌ترین نمره در هر ترم، بیشترین نمره در هر ترم، معدل هر ترم، نمرات دروس دوره دبیرستان، و اطلاعات دیگری موجود می‌باشد. اگر هدف پژوهشگری تحقیق درباره دانشجویان رشته خاصی، از دانشگاهی مورد نظر باشد، ابتدا در دانشگاه هدف، دانشجویان را بر اساس رشته تحصیلی طبقه‌بندی کرده سپس دانشجویانی که در طبقه مورد نظر پژوهش قرار گرفته‌اند را مورد بررسی و تحقیق قرار می‌دهد.

حال فرض کنید با توجه به هدف پژوهش، نتوان ایده‌ای خاص برای طبقه‌بندی ارائه داد. به عنوان مثال اگر هدف از پژوهشی بررسی تأثیرات علاقه تحصیلی و اعتقادات عقلی و دینی بر پشتکار تحصیلی دانشجویان باشد، شاید نوع پژوهش، ایده‌ای را برای طبقه‌بندی و دسته‌بندی در اختیار پژوهشگر قرار ندهد. روش‌های مختلف خوشه‌بندی به پژوهشگران این امکان را می‌دهد که دسته‌بندی، طبقه‌بندی و سطح بندی مشاهدات را با کمک تکنیک‌ها و روش‌های خوشه‌بندی انجام دهد.

تکنیک‌های خوشه‌بندی قبل از زمان آدم و حوا به کار گرفته می‌شده است. واضح‌ترین مثال

آن در طبیعت نیز خوشه‌بندی یا قوت‌های انگور است. اما به صورت فرمول بندی شده از سال ۱۹۳۹ بوده است. از جمله دانشمندان تأثیرگذار در این علم می‌توان به رابرت ترايون (Robert Tryon) و موجنا اشاره کرد. در خلال این سال‌ها تکنیک‌ها و روش‌های خوشه‌بندی با توجه به نیاز پژوهشگران تنوع و گسترش یافتند. به کمک این تکنیک‌ها و روش‌ها دانش بشر در زمینه‌های مختلف بهداشت و درمان، فنی و مهندسی، علوم اجتماعی و انسانی رشد و توسعه یافته است. یکی از جالب‌ترین موارد استفاده از خوشه‌بندی و کمک آن به نسل بشر، زمانی بود که دانشمندان بزرگ پزشکی در حال بررسی تأثیر بیماری جنون گاوی بر روی انسان بودند.

خوشه‌بندی را می‌توان به دو قسمت روش‌های خوشه‌بندی، که بر اساس فلسفه‌ای خاص بنا نهاده شده‌اند، و وسایل اندازه‌گیری، که با رشد علوم ریاضی وسعت یافته‌اند، تقسیم کرد. روش‌ها و تکنیک‌های خوشه‌بندی را می‌توان به سلسله مراتبی، بخش بندی و مدل پایه تقسیم کرد.

در روش سلسله مراتبی، قابلیت اندازه‌گیری نسبت به دو روش دیگر بالاتر است ولی برای داده‌های زیاد و با بعد بالا به دلیل زمان زیادی که صرف محاسبات می‌نماید، مقرون به صرفه و گاهی نیز قابل محاسبه نیست.

در این کتاب روش‌های خوشه‌بندی مختلف سلسله مراتبی را در فصل‌های ۱ تا ۵ مورد بررسی قرار می‌دهیم. در فصل ۱ خوشه‌بندی به روش نزدیک‌ترین فاصله، در فصل ۲ خوشه‌بندی به روش دورترین فاصله، در فصل ۳ خوشه‌بندی به روش پیوند به میانگین، در فصل ۴ خوشه‌بندی به روش پیوند به میانه و در فصل ۵ خوشه‌بندی به روش پیوند به مرکز با تفسیر ساده شرح داده شده است و نحوه محاسبه فاصله بین خوشه‌ها را با توضیحات کامل بیان کرده و در هر فصل با ۲ مثال سعی می‌شود خواننده به درک شهودی و قابل قبولی از روش‌های خوشه‌بندی سلسله مراتبی برای دسترسی به مناسب‌ترین نحوه دسته‌بندی مشاهدات در پژوهش‌هایش برسد. در

فصل ۶ به معرفی روش جدیدی برای تعیین تعداد خوشه بر اساس خواص تابع توزیع می‌پردازیم و در فصل آخر به آموزش خوشه‌بندی با استفاده از نرم افزار **R** می‌پردازیم.

www.ketab.ir