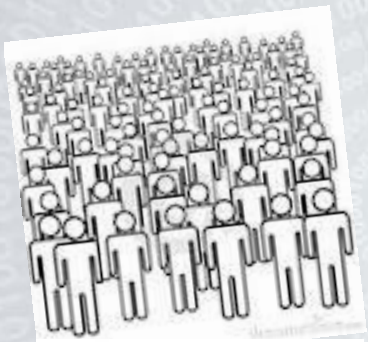
به نام خدا

کاربرد کامپیوتر
در
مهندسی پوشاک

مدرس:
دکتر پدرام پیوندی

تهیه جداول سایز بندی

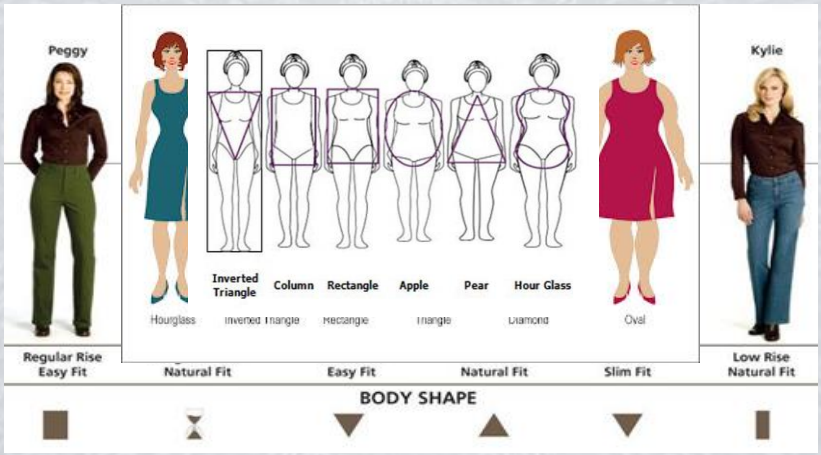
مقدمه:



www.pedram-payvandy.com

تهیه جداول سایز بندی

مشکلات سایزبندی های موجود:



www.pedram-payvandy.com

تهیه جداول سایز بندی

انواع مختلف شکل بدن

ساعت شنی:
این شکل به عنوان بهترین اندام برای زنان انتخاب شده است. در این شکل دورسینه و باسن تقریبا یک اندازه هستند و یک محیط دارند. اگر چه سینه می‌تواند 3سانتیمتر از باسن بزرگ‌تر باشد، ولی کمر خیلی کوچک‌تر خواهد بود. در این نوع شکل تفاوت بین کمر و سینه و کمر و باسن حداکثر 22سانتیمتر است.

۲- قاشقی یا گلابی شکل:
در این نوع اندام اختلاف بین اندازه های دورباسن و دورسینه زیاد است در حالی که اختلاف بین کمر و سینه بسیار کم است.در این شکل باسن 5سانتیمتر یا بیشتر، از دورسینه بزرگ‌تر است و کمر، کمتر از 66 سانتیمتر از دورسینه کوچک‌تر است.

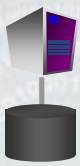
۳- مربع مستطیل (فط‌کش):
در این نوع اندام، بین اندازه‌های کمر، سینه و باسن تفاوت کمی وجود دارد. کمر، کمتر از 22سانتیمتر از سینه کوچک‌تر است.

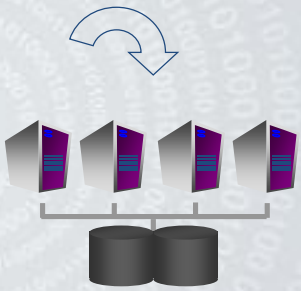
۴- مخروطی (مثلث):
معمولاً در این نوع شکل تفاوت بین سینه و کمر، کم است اما باسن 22سانتیمتر از قفسه‌ی سینه کوچک‌تر است. دورکمر 9سانتیمتر از باسن بزرگ‌تر است.

www.pedram-payvandy.com

آشنایی با داده کاوی و کاربرد آن در صنعت پوشاک

داده کاوی





امروزه در اکثر سازمانها، داده‌ها به سرعت در حال جمع آوري و ذخيره شدن مي باشند.

کم‌کم سازمانها علاقه‌مند شدند تا از این منبع بزرگ و ارزشمند براي رشد و توسعه خود استفاده کنند.

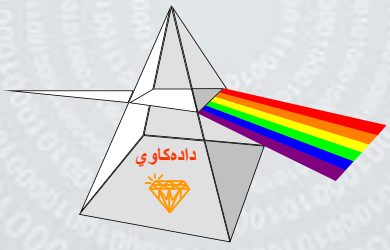
http://www.pedram-payvandy.com

داده کاوی

اما این داده‌های خام و اعداد و ارقام مربوطه به تنهایی هیچ کمکی نمی کنند، وحتی می توان ادعا کرد که علیرغم این حجم انبوه‌داده‌ها، امروزه سازمانها با فقر دانش روبرو هستند.

داده کاوی

علم داده کاوی عبارت است از: "استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده‌های بسیار بزرگ"



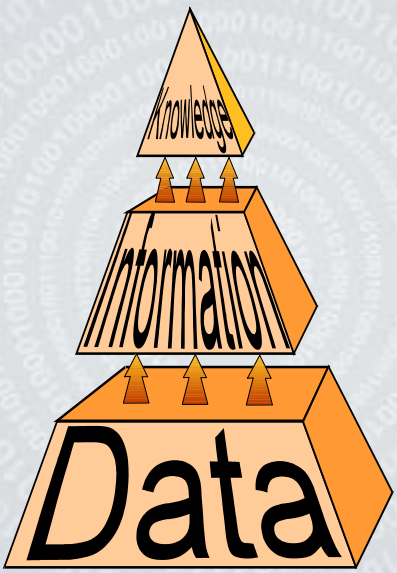
داده ،اطلاعات،دانش

داده‌ها به تنهايي هيچ گونه اطلاعاتي به ما نمي دهند.

با پردازشي بسيار جزيي داده‌ها به اطلاعات تبديل مي شوند.

دانش سطح بالاتري نسبت به داده و اطلاع مي باشد.
دانش يک مفهوم ذهني در مورد يک مسئله مي باشد.

http://www.pedram-payvandy.com



Operational Processes

http://www.pedram-payvandy.com

داده کاوی

چرا، کي، کجا

داده‌هاي مرتبط وجود داشته باشند.

داده‌ها در يکجا مجتمع شده باشند و بتوان به آنها دسترسي داشت.

توانايي کامپيوترها اين امکان را برهر.

نرم افزارهاي کامپيوتري مناسب داده‌کاوي در دسترس باشند.

http://www.pedram-payvandy.com

تفاوت آمار و داده کاوی

داده کاوی:

- به فرضیه احتیاجی ندارد.
- ابزارهای داده کاوی از انواع مختلف داده، نه تنها عددی می‌توانند استفاده کنند.
- الگوریتمهای داده کاوی به طور اتوماتیک روابط را ایجاد می‌کنند.
- داده کاوی به داده‌های صحیح و درست نیاز دارد.
- نتایج داده کاوی نسبتا پیچیده می‌باشد و نیاز به متخصصانی جهت بیان آنها به مدیران دارد.

آنالیز آماری:

- آمار شناسان همیشه با یک فرضیه شروع به کار می‌کنند.
- آنها از داده‌های عددی استفاده می‌کنند.
- آمارشناسان باید رابطه‌هایی را ایجاد کنند که به فرضیه آنها مربوط است.
- آنها می‌توانند داده‌های نابجا و نادرست را در طول آنالیز مشخص کنند.
- آنها می‌توانند نتایج کار خود را تفسیر و برای مدیران بیان کنند.

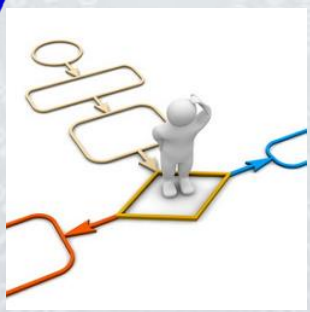
http://www.pedram-payvandy.com

نقش داده کاوی

ذخیره ساز داده ها براي شرکت مثل يک حافظه عمل مي کند ولي حافظه بدون هوش وآگاهی فایده چندانی ندارد

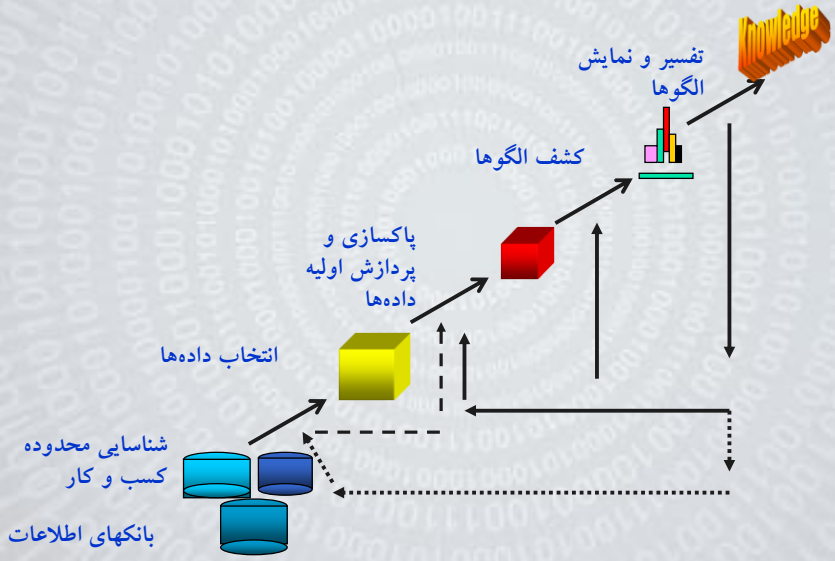
آگاهي به ما اجازه مي دهد:

- در حافظه هاي قبلي خود جستجو کنيم
- به الگوهاي خاصي توجه نمائيم
- قوانين را ايجاد کنيم



http://www.pedram-payvandy.com

فرآیند داده کاوی



http://www.pedram-payvandy.com

روشهای داده کاوی

روشهای داده کاوی بطور کلی به دو دسته زیر تقسیم می شوند:

- الگوریتمهای یادگیری با نظارت
- الگوریتمهای یادگیری بدون نظارت

در الگوریتمهای یادگیری با نظارت هدف از داده کاوی مشخص است و می دانیم که به دنبال چه نوع دانشی می گردیم. مانند دسته بندی.

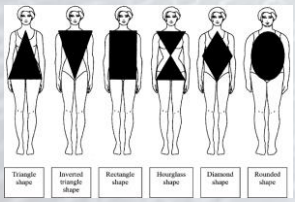
در روشهای یادگیری بدون نظارت، هدف کاملا تعریف شده نیست. مانند خوشه بندی.

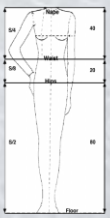
عملیات و تکنیک های داده کاوی

عملیات	دسته بندي پيش بيني خوشه بندي تخمين تعيين ارتباط بين متغير ها
روش ها	شبكه هاي عصبي درخت تصميم گيري الگوريتمهاي نزديكترين همسايه تحليل خوشه استنتاج قانون الگوريتم ژنتيك

کاربرد داده کاوی در صنعت پوشاک

- سایزبندی
- خوشه بندی فرم بدن،تصور ازبدن،مشکلات قسمتهای مختلف بدن
- پیش بینی بازار پوشاک
- برنامه ریزی تولید
- پیش بینی روند مد
- خوشه بندی منابع داده اینترنتی لباس
- ....





http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

Using decision tree-based data mining to establish a sizing for the manufacture of garments system
Chih-Hung Hsu · Mao-Jiun J.Wang
Int J Adv Manuf Technol (2005) 26: 669–674

An application of data mining technique in developing sizing system for army soldiers in Taiwan
Hai Fen Lin,Chih-Hung Hsu · Mao-Jiun J.Wang
WSEAS Transactions on Computers , Volume 7 Issue 4, April 2008
DEVELOPING FEMALE SIZE CHARTS FOR FACILITATING GARMENT PRODUCTIONBY USING DATA MINING
Chih-Hung Hsu
Journal of the Chinese Institute of Industrial Engineers, Vol. 24, No. 3, pp. 245-251 (2007)

http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

DEVELOPMENT OF A NEW SIZING SYSTEM BASED ON DATA MINING APPROACHES
Jamal SHAHRABI, Maryam Salehi ESFANDARANI
7th International Conference - TEXSCI 2010 September 6-8, Liberec, Czech Republic

Data Mining for Prediction of Clothing Insulation
M.Martin Jeyasingh, Kumaravel Appavoo,P.Sakthivel
International Journal of Modern Engineering Research (IJMER) Vol.2, Issue.2, Mar-Apr 2012 pp-001-005

Application of a data mining method on textile sales to identify the consumer preferences
F. Boussu, J.J. Denimal
Computing Engineering in Systems Applications (CESA 2003)

http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

هشتمین کنفرانس ملی مهندسی نساجی ایران- آبان ۱۳۹۰-دانشگاه یزد، دانشکده مهندسی نساجی
8th National Conference on Textile Engineering- October ۲۰۱۱- Yazd, Iran

استفاده از تکنیک های داده کاوی در سایزبندی مانتو مدارس یزد

ندا دهقان چناری[۱] ،آرزو زارع[۲] ،پدرام پیوندی[۳] ،محسن هادی زاده[۴]
[۱،۲]دانشجوی ارشد تکنولوژی مهندسی نساجی [۳،۴]استادیار
دانشکده مهندسی نساجی،دانشگاه یزد
peivandi@aut.ac.ir

هشتمین کنفرانس ملی مهندسی نساجی ایران- آبان ۱۳۹۰-دانشگاه یزد، دانشکده مهندسی نساجی
8th National Conference on Textile Engineering- October 2011- Yazd, Iran

بررسی عوامل تأثیر گذار بر تصمیم خرید تی شرت مردانه با استفاده از
روش خوشه بندی k-mean

نگین ایرانمنش[۱]، دکتر پدرام پیوندی[۲*]
[۱]دانشجوی کارشناسی مهندسی پوشاک،[۲] استادیار
دانشکده مهندسی نساجی،دانشگاه یزد
peivandi@yazduni.ac.ir

http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

هشتمین کنفرانس ملی مهندسی نساجی ایران- آبان ۱۳۹۰-دانشگاه یزد، دانشکده مهندسی نساجی
8th National Conference on Textile Engineering- October ۲۰۱۱- Yazd, Iran

استفاده از روش خوشه بندی k-mean جهت گروه بندی فرم بدن
(مطالعه موردی دانش آموزان دبیرستانی دختر شهر یزد)

آرزو زارع[۱]، ندا دهقان چناری[۲] ، پدرام پیوندی[۳]، محسن هادی زاده[۴]
[۱،۲]دانشجوی ارشد تکنولوژی مهندسی نساجی [۳،۴]استادیار

دانشکده نساجی،دانشگاه یزد
peivandi@aut.ac.ir

هشتمین کنفرانس ملی مهندسی نساجی ایران- آبان ۱۳۹۰-دانشگاه یزد، دانشکده مهندسی نساجی
8th National Conference on Textile Engineering- October 2011- Yazd, Iran

بررسی تصور از بدن ومشکلات فرم قسمتهای بدن(مطالعه موردی
دانشجویان دختر وپسر شهر یزد)

زهره زارع نژاد[۱]، طاهره محمدی[۲]،ثریا نقدی[۳] ،دکتر پدرام پیوندی[۴*]
[۱]دانشجوی ارشد تکنولوژی مهندسی نساجی، [۲،۳]دانشجوی کارشناسی مهندسی پوشاک، [۴] استادیار
دانشکده مهندسی نساجی،دانشگاه یزد
peivandi@yazduni.ac.ir

http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

کاربرد شبکه‌ی عصبی خودسازمان دهنده مبتنی بر داده کاوی به منظور ایجاد یک جدول اندازه‌ی لباس

Applying self-organizing neural network based on data mining to establish a clothing size chart

فاطمه ماکویی، فرزانه میرجلیلی، محسن هادی زاده*و پدرام پیوندی

یزد، دانشگاه یزد، مجتمع فنی مهندسی، دانشکده مهندسی نساجی، صندوق پستی۷۴۱-۸۹۱۹۵

علوم و فناوری نساجی
مجله علوم و فناوری نساجی
سال اول، شماره ۳، شماره پیاپی ۳
صفحه .....، ۱۳۹۱
ISSN: ۲۱۵۱-۷۱۶۲

http://www.pedram-payvandy.com

کاربرد داده کاوی در صنعت پوشاک

علوم و فناوری نساجی
سال دوم، شماره ۴، شماره پیاپی ۵
صفحه ۲۱۵-۲۰۹، ۱۳۹۱
ISSN: ۲۱۵۱-۷۱۶۲

استخراج و طبقه‌بندی تصاویر طرح سنگ‌شور پوشاک جین با استفاده از الگوریتم خوشه‌بندی K-means

Extracting and Clustering Stone-Wash Design in Jeans Images Using K-means Algorithm

زینب مزدک، پدرام پیوندی*، علی اصغر علمدار یزدی

یزد، دانشگاه یزد، مجتمع فنی و مهندسی، دانشکده مهندسی نساجی، صندوق پستی ۸۹۱۹۵-۷۴۱

تاریخ دریافت: ۹۱/۰۹/۰۲ تاریخ پذیرش: ۹۱/۱۲/۲۰

کاربرد داده کاوی در صنعت پوشاک

علوم و فناوری نساجی
سال سوم، شماره ۱، شماره پیاپی ۶
صفحه ۱۰-۳، ۱۳۹۲
ISSN: ۲۱۵۱-۷۱۶۲

خوشه‌بندی تصاویر پوشاک با استفاده از پردازش تصویر و الگوریتم K-means

Garment Images Clustering Based on Image Processing Techniques and K-Means Algorithm

زهره زارع‌نژاد، پدرام پیوندی*

یزد، دانشگاه یزد، مجتمع فنی مهندسی، دانشکده مهندسی نساجی، صندوق پستی ۸۹۱۹۵-۷۴۱

تاریخ دریافت: ۹۱/۰۸/۲۴ تاریخ پذیرش: ۹۲/۰۳/۲۳

کاربرد داده کاوی در صنعت پوشاک

The sixth Iran Data Mining Conference / IDMC 2012
Dec. 18,19 / 2012 , Tehran, Iran

ششمین کنفرانس داده کاوی ایران، ۲۸ و ۲۹آذر ۱۳۹۱، تهران، ایران
کد مقاله : ۱۰۱

کلیه حقوق مادی و معنوی این مقاله متعلق به دبیرخانه دائمی کنفرانس داده کاوی ایران است و درج آن فقط در سایت رسمی کنفرانس داده کاوی ایران مجاز بوده و هرگونه انتشار غیر از آن، خلاف عرف، اخلاق و قانون و مستوجب پیگیری است.

کاربرد داده کاوی در بازاریابی پوشاک(مطالعه موردی چادر)

زهره زارع‌نژاد[۱]؛ زهرا دهقان بهابادی[۲]؛ پدرام پیوندی[۳]

چکیده

چادر بعنوان یکی از پوشش های جامع و کامل در کشورهای اسلامی خصوصاً ایران مطرح می‌باشد. با گذشت زمان و پیشرفت جوامع نیاز به تغییر در فرم چادر به منظور ایجاد راحتی بیشتر زنان احساس می کردد.امروزه طرحهای جدید همچون چادر دانشجویی، ملی، شالدار ،...در بین دختران و بانوان ایرانی رواج پیدا کرده است و طراحان در پی ایجاد طرحهای جدید برای ایجاد راحتی و پوشش مناسب برای بانوان ایرانی هستند.چادر جلابیب ،صدفی ،...از جمله طرحهای شاخص در این زمینه می باشد.

در این مقاله به بررسی و تحلیل بازاریابی انواع مدلهای چادردرسطح دانشگاه با استفاده از داده کاوی پرداخته شده است.جهت تجزیه و تحلیل و کاوش داده‌های مربوط برای بررسی بازاریابی انواع مدلهای چادر درموقعیت‌های مختلف ازروش خوشه بندی K-mean و تجزیه وتحلیل آماری استفاده شده است. برای بدست‌آوردن دسته‌بندی، خوشه‌های مختلف و همچنین پارامترهای اندازهگیری شده، مورد بررسی قرارگرفت. همچنین تجزیه وتحلیل با استفاده از آزمون دیویس بولدین انجام شده است. این مطالعه روی تعداد ۱۰۰۰ نفر از دانشجویان دختر دانشگاه یزد انجام شده و نتایج حاکی از آن است که چادر دانشجویی در بین دیگر چادرها بیشترین تقاضا را داشته است و چادر ساده علی رغم کمتر بودن راحتی نسبت به سایر چادرها، هنوز تمایل زیادی برای پوشیدن آن وجود دارد که به دلیل زیبایی آن نسبت به سایر چادر هاست، پس نکته قابل توجه در بازاریابی چادر این است که طراحان علاوه بر توجه به راحتی، به زیبایی چادر نیز توجه کنند

کاربرد داده کاوی در صنعت پوشاک

The sixth Iran Data Mining Conference / IDMC 2012
Dec. 18.19 / 2012 . Tehran. Iran

ششمین کنفرانس داده کاوی ایران، ۲۸ و ۲۹آذر ۱۳۹۱، تهران، ایران
کد مقاله : ۲۸۱

کلیه حقوق مادی و معنوی این مقاله متعلق به دبیرخانه دائمی کنفرانس داده کاوی ایران است و درج آن فقط در سایت رسمی کنفرانس داده کاوی ایران مجاز بوده و هرگونه انتشار غیر از آن، خلاف عرف، اخلاق و قانون و مستوجب پیگیری است.

استفاده از روش ترکیبی الگوریتم ژنتیک و کی–مینز در گروه‌بندی تصور افراد از بدن خود

معصومه هوشمند[۱] ؛ فاطمه معدنی[۲] ؛پرستو نقدی[۳]؛پدرام پیوندی[۴]

چکیده

امروزه با رشد روزافزون فناوری، حجم وسیعی از داده‌ها با سرعت بالایی جمع‌آوری و انبارش می‌شوند. لذا اهمیت دست‌یابی و پردازش اطلاعات در راستای تصمیم گیری سازمان با دو پارامترحداقل زمان و عدم دخالت انسان، احساس می‌شودکه نمی توان به صورت یکپارچه از این حجم داده ها استفاده نمود، داده‌کاوی یکی از مهمترین روش‌هایی است که به مسئله استخراج اطلاعات از پایگاه داده‌ها می پردازدهدف از این مقاله استفاده از روش های داده‌کاوی به عنوان یک تکنیک به منظور گروه‌بندی تصور افراد از بدن خود با بررسی پارامترهایی اجتماعی و فردی راحتی مثل معیارهای اجتماعی، محل زندگی، علاقه و ارزش فرد می باشد. بدین منظور از طریق مطالعه میدانی ۴۰۱ پرسش نامه مؤثر دربین دانشجویان خوابگاه دختران دانشگاه یزد و گرفتن سایز(۱۸ اندازه از بدن افراد) و همچنین با بکارگیری روش داده کاوی ترکیبی کی–مینز و الگوریتم ژنتیک روابط بین پارامترهای اجتماعی و فردی راحتی و تصور افراد از بدن خود بررسی شد و با استفاده از الگوریتم ژنتیک یک مدلی برای توصیف داده ها نسبت به روش K-Means بکار گرفته شده است .

کاربرد داده کاوی در صنعت پوشاک

ششمین کنفرانس داده کاوی ایران، ۲۸ و ۲۹آذر ۱۳۹۱، تهران، ایران
کد مقاله : ۴۱۳
The sixth Iran Data Mining Conference / IDMC 2012
Dec. 18,19 / 2012 , Tehran, Iran

کلیه حقوق مادی و معنوی این مقاله متعلق به دبیرخانه دائمی کنفرانس داده کاوی ایران است و درج آن فقط در سایت رسمی کنفرانس داده کاوی ایران مجاز بوده و هرگونه انتشار غیر از آن، خلاف عرف، اخلاق و قانون و مستوجب پیگیری است.

استفاده از الگوریتم k-means در دسته بندی طرح لباس
(مطالعه موردی مانتو)

زهره زارع‌نژاد[۱]، پدرام پیوندی[۲]

چکیده

در عصر حاضر، هنر طراحی لباس به دلیل کاربردهای مختلف آن در زوایای مختلف زندگی انسان از اهمیت ویژه ای برخوردار است. انسان امروزی که در صحنه رقابت برای برتری یافتن و ارتقاء سطح دانش و بینش خود همواره از ابزارها و شیوه‌های گوناگونی بهره می‌جوید، لباس را به عنوان یکی از موثرترین ابزارها مورد توجه قرار داده است. در این راستا نرم افزارهای طراحی لباس زیادی جهت طراحی مدل لباس ابداع شده است که بر اساس بانک اطلاعاتی موجود طراحی لباس را انجام می‌دهند و چون کاربر باید از بین این بانک اطلاعاتی گسترده طرح مورد نظرش را پیدا کند بسیار زمانگیر و پر هزینه است. در این مقاله با استفاده از الگوریتم k_means تعداد ۱۰۰ طرح مانتو بر اساس میزان شباهت و نزدیکی آنها گروهبندی شده است بنابر این با کمک نرم افزارهای طراحی لباس ، این پروسه هزینه بر و وقت گیر در زمانی کوتاه‌تر و با هزینه کم‌تر قابل اجرا خواهد بود و باعث تسهیل امر طراحی و ارتقاء سطح آن و کمک به طراحان شود.

کاربرد داده کاوی در صنعت پوشاک

اولین کنفرانس بازشناسی الگو و تحلیل تصویر ایران (۱۶ تا ۱۸ اسفند ۱۳۹۱)
The First Iranian Conference on Pattern Recognition and Image Analysis
March 6-8 ,2013

استفاده از روش خوشه‌بندی K-means در استخراج و طبقه‌بندی طرح سنگشور پوشاک جین

زینب مزدک[۱]؛ پدرام پیوندی[۲]؛ علی اصغر علمدار یزدی[۳]

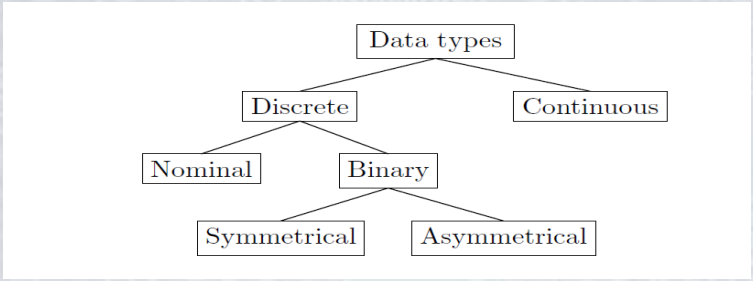
چکیده– استفاده از لباس ها و خصوصاً شلوارهای جین در بین نسل جوان روز به روز افزایش می‌یابد. با توجه به استقبال گسترده ای که از این لباس‌ها در بازارهای جهانی می‌شود، طراحی و کنترل کیفیت این لباس‌ها از اهمیت ویژه ای برخوردار است. یکی از راه‌های ایجاد تغییر و تنوع در این لباس‌ها، ارائه طرح‌های مختلف بر روی آن‌ها با استفاده از روش‌های مختلف سنگشور می‌باشد. به عبارت دیگر سنگشور خود روش ایجاد طرحی جدید بر روی شلوار جین می‌باشد. با توجه به اهمیت این موضوع در صنعت پوشاک، هنوز طبقه‌بندی خاصی برای طرح‌های سنگشور مورد استفاده در کارخانجات وجود ندارد. در این مقاله از ۳۰۶ شلوار جین تحت شرایط یکسان نورپردازی، تصویربرداری شد. سپس با استفاده از الگوریتم خوشه‌بندی k-mean طرح سنگشور استخراج شد و در نهایت با استفاده از خوشه‌بندی k-mean نسبت به طبقه‌بندی طرح‌ها در ۳ گروه اقدام شد.

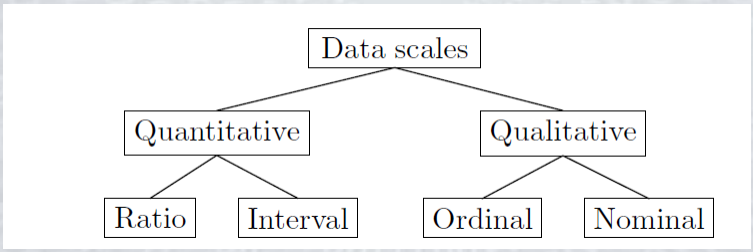
در تحقیقی که در سال ۲۰۰۷ انجام شد، کاربرد الگوریتم خوشه‌بندی K-means استاندارد و ژنتیک را در بخش‌بندی تصویر ارئه نمود و آن‌ها را با هم مقایسه کرد. نتایج نشان داد که الگوریتم ژنتیک K-means در پردازش تصاویر مفیدتر بوده و نتایج بخش‌بندی بهتری را می‌دهد[1]. در سال ۲۰۱۰ روش جدید بخش‌بندی تصویر ارائه شد که بر اساس ویژگی‌های رنگ ارائه شده مراحل کار به دو بخش تقسیم می‌شود:در مرحله اول افزایش تفکیک رنگ تصویر با استفاده از Decorrelation Stretching صورت می‌گیرد و سپس با استفاده از الگوریتم خوشه‌بندی K-means نواحی به ۵ خوشه، گروه‌بندی می شوند. از آنجا که رنگ استفاده شده برای بخش‌بندی تصویر

داده کاوی

http://www.pedram-payvandy.com

داده





http://www.pedram-payvandy.com

فقدان داده

- بی اهمیت بودن
- عدم ثبت
- نامربوط بودن
- عدم پاسخگویی

http://www.pedram-payvandy.com

پیش پردازش فقدان داده

- حذف کل اطلات با داده ناقص
- جایگذاری داده ناقص با میانگین داده ها
- جایگذاری داده ناقص با بیشترین تکرار در داده ها
- جایگذاری با تمام مقادیر موجود

http://www.pedram-payvandy.com

پیش پردازش داده

min-max normalization
z-score normalization
normalization by decimal scaling

کمی سازی

استاندارد سازی

مقیاس بندی

$$s_f = \frac{1}{n}(|x_{1f} - m_f| + |x_{2f} - m_f| + ... + |x_{nf} - m_f|)$$

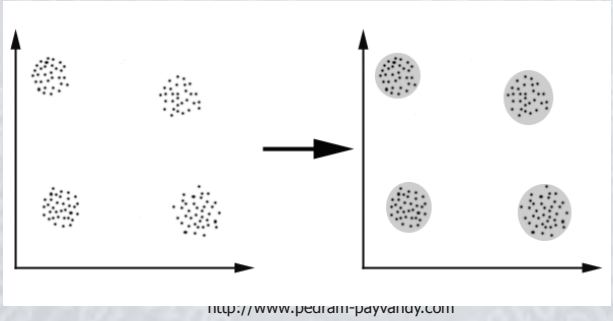
$$m_f = \frac{1}{n}(x_{1f} + x_{2f} + ... + x_{nf})$$

$$z_{if} = \frac{x_{if} - m_f}{s_f}$$

http://www.pedram-payvandy.com

خوشه بندی

روش در یادگیري بدون نظارت در نظر گرفت. خوشه‌بندي را مي‌توان به عنوان مهمترين خوشه‌بندي با يافتن يک ساختار درون يک مجموعه از داده‌هاي بدون برچسب درگير است. خوشه به مجموعه‌اي از داده‌ها گفته مي‌شود که به هم شباهت داشته باشند. در خوشه‌بندي سعي مي‌شود تا داده‌ها به خوشه‌هايي تقسيم شوند که شباهت بين داده‌هاي درون هر خوشه حداکثر و شباهت بين داده‌هاي درون خوشه‌هاي متفاوت حداقل شود.



http://www.pedram-payvandy.com

خوشه بندی

خوشه‌بندي(Clustering)

در مقابل

طبقه‌بندي(Classification)

در طبقه‌بندي دسته‌ها(کلاس) از پيش مشخص شده‌اند و هر داده به يک طبقه (کلاس) از پيشين مشخص شده تخصيص مي‌يابد. ولي در خوشه‌بندي هيچ اطلاعي از کلاسهاي موجود درون داده‌ها وجود ندارد و به عبارتي خود خوشه‌ها نيز از داده‌ها استخراج مي‌شوند.

روشهای خوشه بندی

خوشه‌بندي انحصاري (Exclusive or Hard Clustering)

و

خوشه‌بندي با هم‌پوشي (Overlapping or Soft Clustering)

خوشه‌بندي سلسله مراتبي (Hierarchical)

و

خوشه‌بندي مسطح(Flat)

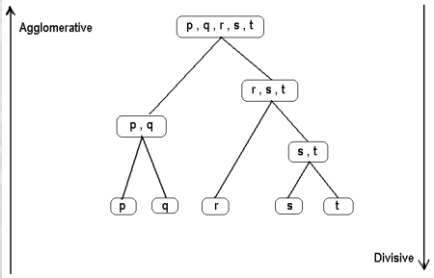
روشهای خوشه بندی

روشهاي خوشه‌بندي سلسله مراتبي

روش سلسله مراتبي خود داراي دو نوع تراكمي و تقسيم شونده

بالا به پايين (Top-Down) يا تقسيم كننده (Divisive)

پايين به بالا (Bottom-Up) يا متراكم شونده (Agglomerative)



http://www.pedram-payvandy.com

روش تراكمی

در روش تراكمي، هر شي يا مشاهده به عنوان يک خوشه در نظر گرفته مي شود.

- در گام هاي بعدي، نزديکترين دو خوشه (مشاهده)، به صورت يک خوشه تراكمي جديد با هم تركيب مي شوند، بنابراين تعداد خوشه در هر گام کم مي شود.
- در برخي موارد، سومين مشاهده انفرادي، با اولين دو مشاهده در يک خوشه تلفيق مي شود.

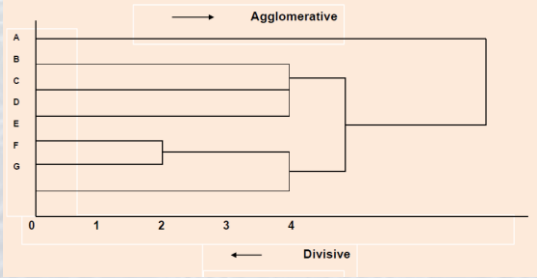
- در ساير موارد دو گروه از مشاهدات انفرادي كه در مرحله قبلي تشكيل شده‌اند يک خوشه جديد را تشكيل مي دهند و نهايتا همه مشاهدات انفرادي دريکخوشه بزرگ با هم گروه بندي مي شوند، به همين دليل برخي موارد رويه هاي تراكمي روش هاي **Build up** ناميده می شوند

http://www.pedram-payvandy.com

روش تراکمی

-یک مشخصه مهم از رویه هاي سلسله مراتبي این است که نتایج مرحله قبلي در نتایج مرحله بعدي جاي مي گیرد و شکلي شبیه به یک درخت را ایجاد مي کند.

- به علت اینکه خوشه ها تنها از طریق تلفیق خوشه های موجودتشکیل مي شوند، هر عضو از یک خوشه مي تواند عضویتش را دریک مسیر پیوسته تا نقطه شروع به عنوان یک مشاهده تکي دنبال کند.



http://www.pedram-payvandy.com

روش خوشه بندی k-means

-در ابتدا **K** نقطه به عنوان به نقاط مراکز خوشه‌ها انتخاب مي‌شوند.
-هر نمونه داده به خوشه‌اي که مرکز آن خوشه کمترین فاصله تا آن داده را داراست، نسبت داده مي‌شود.

-پس از تعلق تمام داده‌ها به یکي از خوشه‌ها براي هر خوشه یک نقطه جدید به عنوان مرکز محاسبه مي‌شود. (میانگین نقاط متعلق به هر خوشه)

-مراحل قبل به ترتیب تکرار مي‌شوند تا زماني که دیگر هیچ تغییري در مراکز خوشه‌ها حاصل نشود.

http://www.pedram-payvandy.com

مشکلات روش خوشه بندی k-means

علي‌رغم اينكه خاتمه‌پذيري الگوريتم بالا تضمين شده است ولي جواب نهايي آن واحد نبوده و همواره جوابي بهينه نمي‌باشد. به طور كلي روش ساده بالا داراي مشكلات زير است.

جواب نهايي به انتخاب خوشه‌هاي اوليه وابستگي دارد.

روالي مشخص براي محاسبة اولية مراكز خوشه‌ها وجود ندارد.

اگر در تكراري از الگوريتم تعداد داده‌هاي متعلق به خوشه‌اي صفر شد راهي براي تغيير و بهبود ادامة روش وجود ندارد.

در اين روش فرض شده است كه تعداد خوشه‌ها از ابتدا مشخص است. اما معمولا در كاربردهاي زيادي تعداد خوشه‌ها مشخص نمي‌باشد.

محاسبه فاصله بین دو خوشه

Minkoskiفاصله

یکی از معمولترین متریک های فاصله می باشد

$$Si=lp(x, \mu i)= \left(\sum_{j=1}^{n} |\mu_{ij} - x_{j|}{}^{p} \right)^{1/p}$$

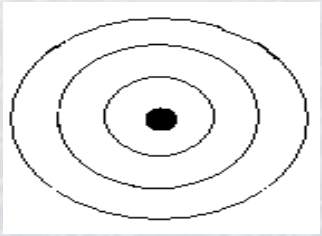
وزن داده شده به اختلاف در هر بعد را کنترل می‌کند. چنانچه این پارامتر برابر **P**در رابطه بالا، 1 باشد معیار فاصله منهاتان، چنانچه برابر 2 باشد معیار فاصله اقلیدسی وچنانچه به سمت بینهایت میل کندمعیار چبیشف را نتیجه می‌دهد

محاسبه فاصله بین دو خوشه

فاصله اقلیدسی:

فاصله اقلیدسی برابر طول پاره خط واصل دونقطه (کوتاه ترین فاصله دو نقطه) می‌باشد بعدی باشند فاصله n دو بردار μ_i و X واز رابطه فیثاغورث به دست می‌آید. اگر اقلیدسی آنها برابر است با :

$$d(x,\mu_j)=L_p=_2(X,\mu_j)=\sqrt{\sum_{j=1}^{n}(\mu_{ij}-x_j)^2}$$



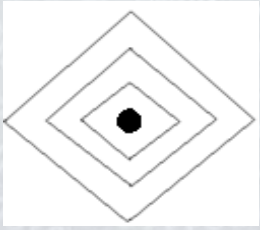
محاسبه فاصله بین دو خوشه

معیار فاصله منهاتان:

در معیار P=1 این معیار از نظر بار محاسباتی،ساده ترین معیار فاصله استو به ازای به دست می‌آید.معیار فاصله منهاتان مطابق رابطه زیر تعریف می‌شود.اگر Minkoski بردارها باینری باشند این معیار فاصله همینگ نامیده می‌شود.

$$d(x,\mu_j)=\sum_{j=1}^{n}|\mu_{ij}-x_j|$$ manhatan

Hamming



بررسی روشهای ارزیابی خوشه بندی

تراکم (Compactness)

جدایی (Separation)

فاصلهٔ بین نزدیک‌ترین داده‌ها از دو خوشه
فاصلهٔ بین دورترین داده‌ها از دو خوشه
فاصلهٔ بین مراکزخوشه‌ها

روش‌های ارزیابی خوشه‌های حاصل از خوشه‌بندی را به صورت سه دسته تقسیم می‌کنند که عبارتند از:

معیارهای خروجی (External Criteria)
معیارهای درونی (Internal Criteria)
معیارهای نسبی (Relative Criteria)

http://www.pedram-payvandy.com

علامت	مفهوم
n_c	تعداد خوشه‌ها
d	تعداد ابعاد
$d(x,y)$	فاصلهٔ بین دو داده
$\overline{X_j}$	امید ریاضی اُمین بعد داده‌ها
$\|X\|$	$\sqrt{X^T X}$ یک بردار ستونی است.
n_{ij}	تعداد داده‌های درون اُمین بعد از خوشهٔ iام
n_j	تعداد عناصر اُمین بعد از تمام داده‌ها
v_i	نقطهٔ مرکز خوشهٔ iام
c_i	iامین خوشه
$\|c_i\|$	تعداد داده‌های درون iامین خوشه

http://www.pedram-payvandy.com

شاخص Dunn Index

$$D = \min_{i=1\ldots n_c}\left\{\min_{j=i+1\ldots n_c}\left(\frac{d(c_i,c_j)}{\max_{k=1\ldots n_c}(diam(c_k))}\right)\right\}, \text{where}$$

$$d(c_i,c_j) = \min_{x\in c_i, y\in c_j}\{d(x,y)\} \text{ and } diam(c_i) = \max_{x,y\in c_i}\{d(x,y)\}$$

اگر مجموعة داده‌اي، داراي خوشه‌هايي جداپذير باشد، انتظار مي‌رود فاصلة بين خوشه‌ها زياد و قطر خوشه‌هاي (Diameter) آن كوچك باشد. در نتيجه مقداري بزرگ‌تر براي رابطة اين معيار مقداري مطلوب‌تر است. معايب اين معيار عبارتند از:

محاسبة زمانبر

حساسيت به نويز (قطر خوشه‌ها در صورت وجود يک دادة نويزي مي‌تواند بسيار تغيير كند.)

شاخص ديويس بولدين
Davies Bouldin Index))

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

$$d_{ij} = d(v_i, v_j), \quad s_i = \frac{1}{\|c_i\|}\sum_{x\in c_i} d(x, v_i)$$

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} R_i, \quad \text{where}$$

$$R_i = \max_{j=1\ldots nc, i\neq j}(R_{ij}), \quad i = 1\ldots n_c$$

اين شاخص در واقع ميانگين شباهت بين هر خوشه با شبيه‌ترين خوشة به آن را محاسبه مي‌كند. مي‌توان دريافت كه هرچه مقدار اين شاخص كمتر باشد، خوشه‌هاي بهتري توليد شده است.

http://www.pedram-payvandy.com