



## استفاده از روش خوشه بندی k-mean جهت گروه بندی فرم بدن

ندا دهقان، آرزو زارع، پدram پیوندی\*، محسن هادی زاده

دانشکده مهندسی نساجی، دانشگاه یزد

\*peivandi@yazduni.ac.ir

### چکیده:

افراد تمایل به پوشیدن لباسی دارند که متناسب با بدن خاص آنها باشد. با دانستن اندازه یک سری از پارامترهای طولی و عرضی بدن افراد می توان فرم بدن را ارزیابی کرد و با توجه به آن سایزبندی را انجام داد. اشکال، فرم بدن افراد ثابت و بدن هر کس متفاوت و منحصر به فرد است. پس شناخت این تفاوت ها و درصد افرادی که دارای این شکل بدن هستند برای طراحان و تولیدکنندگان پوشاک حائز اهمیت می باشد. در تحقیق پیش رو اندازه هفت نقطه بدن از ۱۰۲۰ دختران یزدی در محدوده سنی ۱۵الی۱۷ سال به منظور تعیین فرم بدن آنها گرفته شد. این کار با استفاده از روش خوشه بندی k-mean صورت گرفت و نهایتا فرم بدن افراد در سه گروه تقسیم بندی شد.

**واژه های کلیدی:** اشکال بدن، روش خوشه بندی k-mean

## K-mean clustering method for grouping the body form

A, Zare. N, Dehghan. \* P, Payvandy. M, Hadizade.  
Yazd ,Yazd University ,Department of Textile Engineering

peivandi@yazduni.ac.ir\*

### Abstract:

People tend to wear clothing that is appropriate to their particular body. By knowing the size of a series of longitudinal and transverse parameters of the body can be evaluated form the body and because it did sizing. Forms, the form of bodies are fixed and everyone is unique and different. So recognizing these differences and the percentage of people with this body shape is important for designers and clothing manufacturers. Seven leading in body size of 1020 yazdi girls in 15 to 17 years in order to determine the form of their bodies were taken. This work was carried out using k-mean clustering method and form of the body was divided into three groups.

**Key words:** body forms, k-mean clustering method

## ۱- مقدمه:

دسته بندی شکل های مختلف اندام انسان با کمک آنالیز داده های مربوط به قسمت های مختلف بدن انجام می گیرد. بکارگیری تکنیک های داده کاوی روی این داده ها و استخراج قواعد معنی دار آنها می تواند جهت دسته بندی این شکل های مختلف و در نهایت تعیین سیستم ساینبدی پوشاک مفید واقع گردد [۱]. شناخت تفاوت اشکال بدن در طراحی ویژه برای پوشاک حائز اهمیت می باشد. هر کس می خواهد لباسی به پیروی از شکل بدن بپوشد که متناسب با بدن خاص خود باشد. درک شکل بدن ضروری است. بدن به مرور سال ها در اشکال و اندازه و ویژگی ها تغییر می کند. اشکال بدن زنان بسیار متنوع تر از اشکال بدن مردان می باشد. در برخی از تحقیقات انجام شده، ارتباط شکل بدن با موارد بسیار زیادی بررسی شده است. در سال ۲۰۰۶ شکل بدن و اندازه زنان امریکایی با محاسبه نسبت کمر به باسن ترسیم شده است. بررسی رابطه بین این نسبت و جذابیت زنان نشان داد که افزایش نسبت در صورت کاهش دور باسن باعث افزایش جذابیت و در صورت افزایش هر دو پارامتر (دور باسن و دور کمر) باعث کاهش جذابیت افراد می شود [۲]. در سال ۲۰۱۰ با اندازه گیری دو پارامتر FA (نوسانات غیر متقارن) WHR (نسبت دور کمر به دور باسن) رتبه بندی جذابیت زنان انجام شده است. اگر WHR کمتر از ۰,۷ باشد بهینه جذابیت است [۳]. هم چنین در تحقیق دیگری در سال ۲۰۱۱ شکل بدن زنان امریکایی قبل و بعد از رژیم غذایی و ورزش، با عکس برداری از زنان در طول این مدت بررسی شده است [۴]. در این تحقیق فرم بدن دختران یزدی بررسی و با استفاده از روش خوشه بندی k-mean تعداد گروه شکل بدن مشخص شد.

## ۲- روش تحقیق:

داده کاوی به بررسی و تجزیه تحلیل مقدار عظیمی از داده ها به منظور کشف الگوها و قوانین معنی دار اطلاق میگردد داده کاوی در دو نوع ظاهر میشود، داده کاوی هدایت شده و داده کاوی هدایت نشده. در داده کاوی هدایت شده هدف دسته بندی اطلاعات بر اساس برخی پارامترها مشخص می باشد اما در داده کاوی هدایت نشده هدف یافتن الگوها یا تشابهات بین گروههایی از اطلاعات بدون استفاده از هیچگونه پیش زمینه ای در مورد اطلاعات می باشد. از نمونه روش های داده کاوی هدایت نشده و هدایت شده میتوان به خوشه بندی و دسته بندی اشاره نمود. خوشه بندی به عمل تقسیم جمعیت نا همگن به تعدادی از زیر مجموعه ها یا گروه های همگن گفته میشود. در دسته بندی هر داده به دسته ای از پیش تعیین شده بر اساس دانش قبلی اختصاص می یابد اما در خوشه بندی هیچ دسته از پیش تعیین شده ای وجود ندارد. در واقع خوشه بندی راهی برای یافتن ساختار داده های پیچیده فراهم میکند. به همین دلیل در این تحقیق از روش خوشه بندی جهت گروه بندی افراد با توجه به فرم بدن آنها استفاده شده است. الگوریتم k-mean یکی از پرکاربردترین الگوریتم های خوشه بندی می باشد. حرف k که در اسم این الگوریتم وجود دارد به این واقعیت اشاره دارد که هدف این الگوریتم پیدا کردن تعداد ثابتی از خوشه ها براساس نزدیکی نقاط داده ها به هم می باشد. الگوریتم k-mean به شرح زیر می باشد:

۱. انتخاب K داده به عنوان مرکز خوشه

۲. تعیین فواصل بقیه داده ها با مرکز خوشه ها

۳. قرار گیری داده هایی که به مرکز هر خوشه نزدیکترند در آن خوشه

۴. محاسبه میانگین هر خوشه به عنوان مرکز جدید خوشه

۵. تکرار مرحله دوم تا چهارم تا رسیدن با عدم تغییر در خوشه ها

روش خوشه بندی k-mean بستگی به عواملی چون تعداد خوشه و روش تعیین فاصله بین خوشه ها دارد. یکی از مهمترین مسایل در خوشه بندی انتخاب تعداد خوشه های مناسب می باشد. تعداد خوشه ای مناسب می باشد که (۱) تراکم: نمونه های موجود در یک خوشه تا حد امکان شبیه به یکدیگر باشند معیار رایج برای تعیین میزان تراکم داده ها واریانس داده ها است و (۲) جدایی: نمونه های متعلق به خوشه های متفاوت تا حد امکان از یکدیگر جدا باشند. عبارات فوق را بدین صورت نیز بیان می کنند که خوشه ها باید دارای ماکزیمم فشردگی<sup>۱</sup> باشند و تا حد امکان جدایی<sup>۲</sup> آنها نیز زیاد باشد. اگر تنها معیار فشردگی مورد استفاده قرار گیرد در آن صورت هر داده می تواند به صورت یک خوشه در نظر گرفته شود چرا که هیچ خوشه ای فشرده تر از خوشه ای با یک داده نمی باشد. اگر تنها معیار جدایی در نظر گرفته شود در آن صورت بهترین خوشه بندی این می باشد که کل داده ها یک خوشه در نظر گرفته شود. با این فرض که فاصله هر خوشه از خودش صفر است. بنابراین باید از ترکیب دو معیار فوق استفاده شود. برای سنجش میزان جدایی خوشه ها از توابع فاصله ای استفاده می شود از جمله توابع فاصله: تابع اقلیدسی، تابع فاصله منهایان است. تابع اقلیدسی زمانی بکار می رود که پارامترهای مورد بررسی دارای وزن باشند و تاثیر فاصله در آن بیشتر مشخص میشود. در این تحقیق با توجه به این که تمام پارامترها دارای ارزش برابر می باشند از تابع فاصله منهایان استفاده شده است. ارزیابی خوشه بندی:

یکنواختی توزیع داده ها در اطراف مرکز خوشه

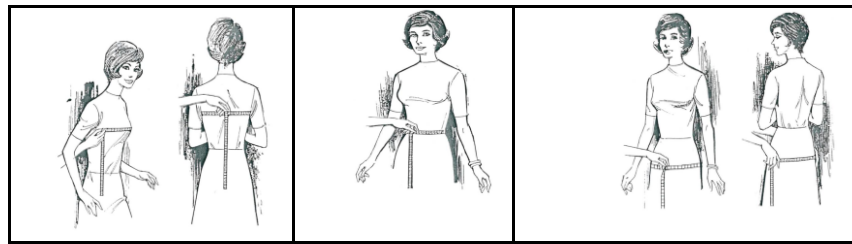
بالا بودن چگالی خوشه (نسبت تعداد الگوهای داخل خوشه به حجم آن)

بالا بودن پوشش (نسبت تعداد الگوهای داخل خوشه به تعداد کل الگوها)

بالا بودن همگنی، بررسی زیرخوشه ها برای خلوت نبودن

پایین بودن همپوشانی خوشه ها

در این تحقیق هفت پارامتر به روش مولر که به طور مثال سه پارامتر دور سینه، دور کمر و دور باسن در شکل ۱ نشان داده شده است اندازه گیری شد. داده ها از ۱۰۲۰ دختر یزدی در محدوده سنی ۱۵الی۱۷سال (مقاطع اول تا سوم دبیرستان) از مدارس مناطق مختلف شهر یزد به منظور تعیین فرم بدن آنها گردآوری شده است.



شکل (۱) روش اندازه گیری مولر (دور باسن ، دور کمر، دور سینه) [۵]

به منظور گروه بندی فرم بدن افراد با استفاده از روش خوشه بندی k-mean، تعداد نقاط (تعداد kهای) مختلف مورد بررسی قرار گرفتند. برای ارزیابی تعداد خوشه ها شاخص ها متفاوتی وجود دارد. شاخص های اعتبارسنجی برای سنجش میزان صحت نتایج خوشه بندی به منظور مقایسه بین روشهای خوشه بندی مختلف یا مقایسه نتایج حاصل از یک روش با پارامترهای مختلف مورد استفاده قرار می گیرند که یکی از این شاخص ها ، شاخص دیویس بولدین<sup>۳</sup> این معیار از شباهت بین دو خوشه ( $R_{ij}$ ) استفاده می کند که بر اساس پراکندگی یک خوشه ( $S_i$ ) و عدم شباهت بین دو خوشه ( $D_{ij}$ ) تعریف می شود. شباهت بین دو خوشه را می توان به صورتهای مختلفی تعریف کرد ولی بایستی شرایط زیر را دارا باشد.

<sup>۱</sup> Compactness

<sup>۲</sup> Separation

<sup>۳</sup> Davies Bouldin index

$$R_{ij} \geq 0$$

$$R_{ij} = R_{ji}$$

۱. اگر  $S_i$  و  $S_j$  هر دو برابر صفر باشند آنگاه  $R_{ij}$  نیز برابر صفر باشد.

۲. اگر  $S_j > S_k$  و  $d_{ij} = d_{ik}$  آنگاه  $R_{ij} > R_{ik}$

۳. اگر  $S_j = S_k$  و  $d_{ij} < d_{ik}$  آنگاه  $R_{ij} > R_{ik}$

معمولا شباهت بین دو خوشه به صورت زیر تعریف می شود:

$$R_{ij} = (s_i + s_j) / d_{ij} \quad (1)$$

که در آن  $d_{ij}$  و  $s_i$  با روابط زیر محاسبه می شوند.

$$d_{ij} = d(\varphi_i, \varphi_j) \quad (2)$$

$$s_i = \sum d(x, \varphi_i) / \|c_i\| \quad (3)$$

با توجه به مطالب بیان شده و تعریف شباهت بین دو خوشه شاخص دیویس بولدین به صورت زیر تعریف می شود.

$$DB = \sum_{i=1}^{nc} R_i / n_c \quad (4)$$

که  $R_i$  در آن به صورت زیر محاسبه می شود.

$$R_i = \max_{j=1 \dots nc, i \neq j} (R_{ij}), i=1 \dots nc \quad (5)$$

این شاخص در واقع میانگین شباهت بین هر خوشه با شبیه ترین خوشه به آن را محاسبه می کند. می توان دریافت که هرچه مقدار این شاخص کمتر باشد، خوشه های بهتری تولید شده است.

### ۳- نتایج و بحث :

در تحقیق پیشرو تابع k-mean برای خوشه بندی و تابع DB برای ارزیابی تعداد خوشه ها اجرا شد. در جدول زیر برای تعداد خوشه های مختلف مقدار DB محاسبه و تعداد نفرات در هر خوشه مشخص شدند.

جدول (۱) بررسی تعداد خوشه برای بدست آوردن بهترین DB

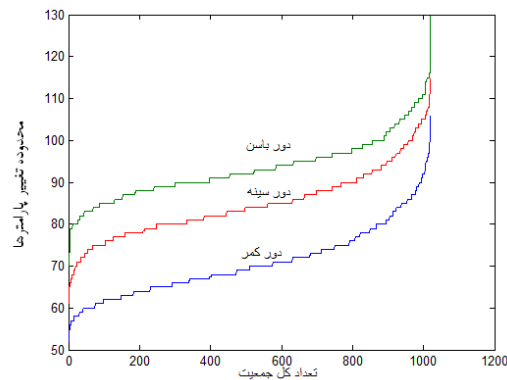
تعداد نفرات در هر خوشه	بهترین مقدار DB بعد از ۱۰۰۰ بار خوشه بندی با انتخاب مقدار اولیه تصادفی	K
۳۶۹ ۵۳۶	DB= ./۸۳۰۱	۲
۱۴۹ ۳۳۴ ۴۲۲	DB= ./۷۷۳۸	۳
۱۹۴ ۱۲۴ ۲۷۷ ۳۱۰	DB= ./۸۶۹۵	۴
۵۷ ۱۷۷ ۲۷۱ ۲۸۶ ۱۱۴	DB= ./۹۴۸۴	۵
۵۵ ۱۷۲ ۱۸۵ ۱۰۴ ۲۶۲ ۱۲۷	DB= ./۹۳۵۶	۶

در این تحقیق علاوه بر شاخص DB که معیاری برای میزان صحت نتایج خوشه بندی است، تعداد نفرات در هر خوشه نیز دارای اهمیت می باشد. از آنجا که هدف از گروه بندی فرم بدن علاوه بر تشخیص شکل بدن، ارائه سامانه ساینز بندی براساس شکل بدن افراد نیز است. در این جدول بهترین تعداد خوشه با کمترین مقدار DB مشخص می شود یعنی در ابتدا  $k=3$  و بعد  $k=2$  اما در این دو نوع خوشه بندی تعداد نفرات هر خوشه برای گروه بندی جمعیت مناسب نمی باشد. بنابراین بهترین تعداد خوشه برای گروه بندی فرم بدن افراد چهار می باشد که مقدار DB آن از  $k=5$ ،  $k=6$  کمتر و تقسیم بندی تعداد نفرات در هر خوشه نیز مناسب می باشد. حال با در نظر گرفتن  $k=4$  به عنوان بهترین تعداد خوشه و با توجه به مراکز خوشه ها، WHR (نسبت دور کمر به دور باسن) نسبت جذابیت جمعیت هدف در جدول زیر آورده شده است.

جدول (۲) مشخصات بهترین تعداد خوشه

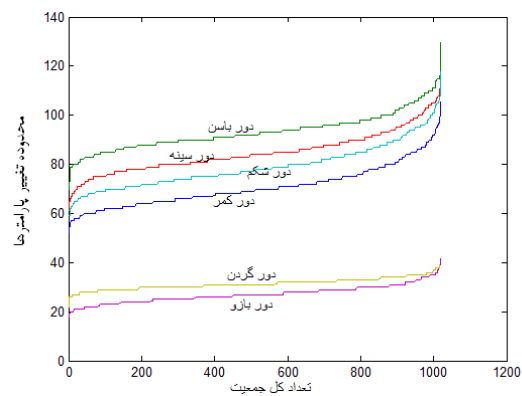
k	DB	مراکز خوشه ها			تعداد نفرات در هر خوشه	WHR
		دور باسن	دور کمر	دور سینه		
۴	.۸۶۹۵	۷۶/۱۶	۶۲/۲	۸۵/۲۷	۱۹۴	.۷۳۲
		۸۱/۴	۶۶/۸۵	۹۰/۷۵	۳۱۰	.۷۳۲
		۸۶/۵	۷۲/۳۵	۹۴/۵	۲۷۷	.۷۷۴
		۹۴/۲۵	۷۹/۸	۱۰۱/۶	۱۲۴	.۷۸۸

به طور معمول تعیین فرم بدن زنان با توجه به سه پارامتر دور سینه، دور باسن و دور کمر انجام میشود [۶]. منحنی تغییرات این سه پارامتر در شکل ۲ نشان داده شده است.



شکل (۲) منحنی تغییرات (دور کمر، دور باسن، دور سینه)

همان طور که در شکل مشخص شده است روند تغییرات پارامترها مشابه است و با توجه به مراکز خوشه که در جدول ۲ آمده است. تمام افراد دارای فرم بدن ساعت شنی (دور سینه و باسن تقریباً یک اندازه و یک محیط اما کمر خیلی کوچکتر) می باشند و نسبت WHR تمام افراد تقریباً یکسان می باشد. گروه بندی فرم بدن جمعیت با استفاده از شش پارامتر دور سینه، دور کمر، دور باسن، دور شکم، دور بازو و دور گردن نیز انجام شد.



شکل (۳) منحنی تغییرات

شکل ۳ منحنی تغییرات شش پارامتر را نشان می دهد. جدول ۳ نیز برای بدست آوردن بهترین تعداد خوشه جهت گروه بندی براساس این شش پارامتر را نشان می دهد.

جدول (۳) بررسی تعداد خوشه برای بدست آوردن بهترین DB

بهترین مقدار DB بعد از ۱۰۰۰ بار خوشه بندی با انتخاب مقدار اولیه تصادفی	K
DB=۰/۷۴۲۳	۲
DB=۰/۶۷۵۱	۳
DB=۰/۷۰۵۴	۴
DB=۰/۶۹۹۳	۵
DB=۰/۷۸۹۳	۶

جمعیت هدف با توجه به کمترین مقدار DB به سه گروه تقسیم بندی می شوند.

جدول (۴) مشخصات بهترین تعداد خوشه

k	DB	مراکز خوشه ها					
		دور باسن	دور کمر	دور سینه	دور شکم	دور بازو	دور گردن
۳	۰/۶۷۵۱	۸۸	۶۴/۲۲	۷۵/۵	۷۲/۰۵	۲۵	۳۰/۲
		۹۴/۸	۷۲/۸۷	۸۶/۷۸	۸۱/۳۶	۲۷/۹	۳۱/۸
		۱۰۶/۰۵	۸۶/۳	۹۹/۶	۹۵/۴	۳۲	۳۴/۲

در جدول ۴ مراکز خوشه ها روند افزایشی را نشان می دهد همان طور که سایزبندی های موجود دارای روند افزایشی بین سایزهای مختلف هستند. با در نظر گرفتن این موضوع می توان فرم بدن افراد را نیز معیاری برای سایزبندی دانست.

#### ۴- نتیجه گیری:

نتایج نشان داد که تمام دختران یزدی دارای فرم بدن مشابه "ساعت شنی" هستند. بنابراین با افزودن سه پارامتر دیگر تقسیم بندی فرم بدن در سه گروه انجام شد و مشخص شد که در گروه ها شکل و سایز بدن افراد مطابق با روند طبیعی افزایش می یابد.

#### مراجع:

- [۱] باقرزاده، ر. شهرابی، ج. لطیفی، م. شیخ زاده نجار، س. ارزیابی کاربرد تکنیک های داده کاوی در ایجاد سامانه های سایزبندی پوشاک، دومین کنفرانس داده کاوی ایران، ۱۳۸۷
- [2] Nana A. Dawson-Andoh, James J. Gray, José A. Soto, Scott Parker, Body shape and size depictions of African American women in JET magazine Body Image, Volume 8, Issue 1, Pages 86-89, 2006
- [3] Helen K. Perilloux, Gregory D. Webster and Steven J. C. Gaulin, Signals of Genetic Quality and Maternal Investment Capacity: The Dynamic Effects of Fluctuating Asymmetry and Waist-to-Hip Ratio on Men's Ratings of Women's Attractiveness, Social Psychological and Personality Science 2010
- [4] Heather Thompson-Brenner, Christina L. Boisseau, Michelle S. St. Paul, Representation of ideal figure size in Ebony magazine: A content analysis Original Research Article Body Image, Volume 8, Issue 4, Pages 373-378, 2011
- [۵] ها کوپیان، س. عزتی، م. طراحی و برش لباس زنانه به روش مولر، ۱۳۸۶
- [6] Devendra Singh and Robert K. Young, Body Weight, Waist-to-Hip Ratio, Breasts, and Hips: Role in Judgments of Female Attractiveness and Desirability for Relationships, Austin, 1995