مجله علوم و فناوری نساجی
سال سوم، شماره ۳، شماره پیاپی ۸
صفحه ۱۶-۱۱، ۱۳۹۲
ISSN: ۲۱۵۱-۷۱۶۲

روش‌های تحلیل عاملی و خوشه‌بندی K-Means برای استخراج متغیرهای مهم اندازه‌گیری و گروه‌بندی شکل بدن

Factorial Analysis and K-Means Clustering Methods for Extraction of Human Body Size Parameters and Shape Groupings

رضا شمسایی، پدرام پیوندی*، سعید فتاحی

دانشگاه یزد، مجتمع فنی و مهندسی، دانشکده مهندسی نساجی، یزد، ایران

تاریخ دریافت: ۹۲/۰۵/۲۴ تاریخ پذیرش: ۹۲/۰۷/۲۷

چکیده

دسته‌بندی نمونه‌های مختلف اندامی انسان با استفاده از قواعد معنی‌دار به‌دست آمده از تحلیل داده‌های مربوط به بخش‌های مختلف بدن بسیار مهم و در بسیاری از علوم استفاده می‌شود. به‌کارگیری روش‌های داده‌کاوی روی این داده‌ها و استخراج قواعد معنی‌دار آن‌ها می‌تواند برای دسته‌بندی این نمونه‌های مختلف و در نهایت تعیین سامانه اندازه‌بندی پوشاک مفید واقع شود. نکته مهم در گروه‌بندی شکل بدن، شناخت تفاوت‌های بدن افراد و استخراج متغیرهای مهم طولی و عرضی بدن است که در نهایت نتایج به‌دست آمده از گروه‌بندی شکل بدن می‌تواند در اختیار طراحان و تولیدکنندگان پوشاک قرار گیرد. در پژوهش پیش رو، اندازه ۱۰ نقطه از بدن ۲۰۰۲ نفر از جوانان ایرانی (پسران) در محدوده سنی ۱۸ تا ۳۰ سال برای تعیین شکل بدن آن‌ها مدنظر قرار گرفت. از روش تحلیل عاملی برای استخراج متغیرهای مهم و از خوشه‌بندی K-Means به منظور گروه‌بندی شکل بدن استفاده شد. در نهایت، شکل بدن افراد در چهار گروه دسته‌بندی شد که نتایج به‌دست آمده از این چهار گروه می‌تواند در اندازه‌بندی پوشاک مردانه مفید واقع شود.

کلمات کلیدی
شکل بدن،
روش تحلیل عاملی،
روش خوشه‌بندی K-Means

مقدمه

افراد تمایل به پوشیدن لباسی دارند که متناسب با بدن، ویژه آن‌ها باشد. با دانستن اندازه مجموعه‌ای از متغیرهای طولی و عرضی بدن افراد می‌توان شکل بدن را ارزیابی کرد و با توجه به آن اندازه‌بندی را انجام داد. شکل بدن افراد با یکدیگر متفاوت است، بنابراین شناخت شکل بدن افراد و تفاوت‌های آن با یکدیگر برای طراحان و تولیدکنندگان پوشاک حائز اهمیت است[۱].

در سال ۲۰۰۷[۲]،۳۶ نقطه از بدن ۷۸۰۰ دانش‌آموز تایوانی اندازه‌گیری شد. سپس، به روش تحلیل عاملی با چرخش Varimax و تحلیل خوشه‌بندی سه دسته مجزا از اندام افراد معرفی شد. در سال ۲۰۰۸[۳]، ۳۸ متغیر اندام‌سنجی (anthropometric) از ۶۱۰ سرباز تایوانی اندازه‌گیری شد، سپس با استفاده از روش داده‌کاوی بر اساس درخت تصمیم‌گیری چهار نوع اندام و ۵۳ گروه اندازه به‌دست آمد. در پژوهش دیگری در سال ۲۰۱۱[۴]، شکل بدن زنان آمریکایی پیش و پس از رژیم غذایی و ورزش، با عکس‌برداری از زنان در طول این مدت بررسی شد. در نهایت، شکل بدن زنان آمریکایی به سه گروه دسته‌بندی شد. در سال ۲۰۱۱[۱]، از روش خوشه‌بندی K-Means برای گروه‌بندی شکل بدن دانش‌آموزان یزدی استفاده شد که بر این اساس شکل بدن افراد به سه گروه دسته‌بندی شد.

هدف از این پژوهش گروه‌بندی شکل بدن جوانان ایرانی

*مسئول مکاتبات، پیام‌نگار: peivandi@yazd.ac.ir

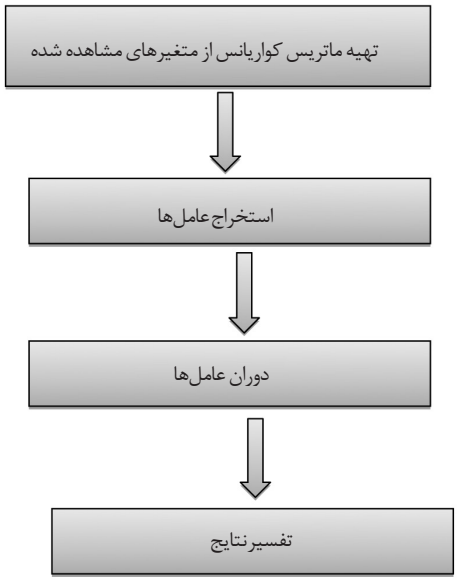
است که بدین منظور از روش تحلیل عاملی برای استخراج متغیرهای مهم اندام‌سنجی (اندازه‌گیری شده) استفاده شد، سپس عامل‌های مشخص‌شده در تحلیل عاملی به عنوان ورودی در تحلیل خوشه‌بندی استفاده شد که بر اساس این تحلیل چهار دسته مجزا از اندام افراد تشخیص داده شد.

اندازه‌گیری

مبحث اندازه‌گیری بدن انسان یا اندام‌سنجی (anthropometry) می‌تواند به عنوان علم مربوط به اندازه‌گیری انسان تعریف شود. اندام‌سنجی کلمه یونانی است که از دو واژه anthropo به معنی انسان (گونه انسان) و metry به معنی سنجش و اندازه‌گیری تشکیل شده است [۵]. اندام‌سنجی شامل داده‌های عددی درباره اندازه، شکل و سایر مشخصات فیزیکی بدن انسان است و می‌تواند در مفهوم طراحی به‌کاربرده شود.

تحلیل عاملی

تحلیل عاملی، به عنوان ابزاری در فرایند پیش‌پردازش داده‌ها بر این فرض استوار است که متغیرهای آشکار به تعداد کمتری متغیر پنهان یا مشاهده نشده بستگی دارند و هر متغیر آشکار تابع خطی از این متغیرهای پنهان است. در واقع هدف تحلیل عاملی کاهش ابعاد داده‌ها با استفاده از عامل‌های پنهان است. مراحل روش تحلیل عاملی در نمودار جریان در شکل ۱ نشان داده شده است.



شکل ۱- مراحل روش تحلیل عاملی.

تهیه ماتریس کواریانس (ضریب همبستگی) از متغیرهای مشاهده‌شده

اگر x_j ،$x_1,x_2,\ldots$ متغیرهای مدنظر بوده و هر کدام شامل n عضو (داده) باشند، در این حالت ضریب همبستگی بین دو متغیر x_i و x_j ($r_{xi,xj}$)، از معادله (۱) به‌دست می‌آید:

$$r_{x_i,x_j} = \frac{S_{xi,xj}}{\sqrt{S_{xi,xi} + S_{xj,xj}}} \qquad (۱)$$

در این معادله $S_{xi,xj}$، $S_{xi,xi}$ و $S_{xj,xj}$ به ترتیب مجموع مربعات x_i و x_j، مجموع مربعات x_i و مجموع مربعات x_j است که از معادله‌های (۲) تا (۴) محاسبه می‌شوند:

$$S_{xi,xj} = \sum_{i,j=1}^{n} x_i\ x_j - \frac{1}{n}\left(\sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j\right) \qquad (۲)$$

$$S_{xi,xi} = \sum_{i=1}^{n} x_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \qquad (۳)$$

$$S_{xj,xj} = \sum_{j=1}^{n} x_j^{\,2} - \left(\sum_{j=1}^{n} x_j\right)^2 /n \qquad (۴)$$

نظر به اینکه همبستگی یک متغیر با خودش برابر ۱ است ($r_{xi,xi}$=۱) در ماتریس کواریانس عناصر روی قطر اصلی نیز ۱ است. در نتیجه ماتریس کواریانس به شکل معادله (۵) خواهد بود [۶]:

$$\begin{bmatrix} r_{x1,x1}=1 & \cdots & r_{x1,xj} \\ \vdots & \ddots & \vdots \\ r_{xi,x1} & \cdots & r_{xj,xj}=1 \end{bmatrix} \qquad (۵)$$

استخراج عامل‌ها

روش‌های مختلفی برای استخراج عامل‌ها وجود دارد، روش مؤلفه اصلی مهم‌ترین روش برای استخراج عامل‌هاست. در این روش سعی بر آن است تا بارهای عاملی متغیرها به گونه‌ای برآورد شوند که مجموع مقادیر واریانس مشترک بیشینه شوند. با این کار عامل‌ها درصد بیشتری از واریانس متغیرهای قابل مشاهده را نشان می‌دهند. در این روش، فرایند استخراج عامل به شرح زیر است:

الف- واریانس متغیرهای قابل مشاهده محاسبه می‌شود.

از آنجا که واریانس و کوواریانس داده‌ها به مقیاس اندازه‌گیری شده بستگی دارد. بنابراین، هرگاه متغیرهای آشکار از حوزه‌های متفاوتی باشند یا با مقیاس‌های مختلفی اندازه‌گیری شده باشند، باید داده‌ها نرمال‌سازی شوند. برای نرمال‌سازی داده‌ها، از معادله (۶) استفاده می‌شود:

$$norm(x_{i,j}) = \frac{x_{i,j} - \mu_i}{\sqrt{var(x_i)}} \qquad (۶)$$

در این معادله، $x_{i,j}$ داده jام از متغیر iام، μ_i میانگین داده‌های متغیر i ام و تابع var انحراف معیار است.

با توجه به اینکه که ماتریس کوواریانس داده‌های نرمال شده با ماتریس همبستگی داده‌های اولیه یکسان است. بنابراین، به جای این مرحله می‌توان بدون انجام فرایند نرمال‌سازی داده‌ها، از ماتریس

همبستگی استفاده کرد.
ب- مقادیر ویژه و بردارهای ویژه برای ماتریس همبستگی محاسبه می‌شود.
برای محاسبه مقادیر ویژه ماتریس همبستگی از معادله (۷) استفاده شد:

$$det(A-\lambda_i I)=0 \quad i=1,2,3,\ldots,j \text{(تعداد متغیر ها)} \qquad (۷)$$

در این معادله A ماتریس همبستگی، λ_i مقادیر ویژه به‌دست آمده، () det تابع دترمینان و I ماتریس همانی هم مرتبه با ماتریس همبستگی است. سپس، برای محاسبه بردارهای ویژه متناظر با مقادیر ویژه به‌دست آمده، از معادله (۸) استفاده شد:

$$(A-\lambda_i I)z_i=o \quad i=1,2,3,\ldots,j \text{(تعداد متغیر ها)} \qquad (۸)$$

در این معادله، z_i بردارهای ویژه به‌دست آمده متناظر با مقادیر ویژه، o به معنای ماتریس صفر (یعنی ماتریسی که تمام درایه‌های آن صفر است) است و A، λ_i و I مطابق معادله (۷) تعریف می‌شود.
ج- بار عامل‌ها به ترتیب زیر محاسبه می‌شود:
اگر m عامل مدنظر باشد، به گونه‌ای که m (تعداد عامل‌ها) خیلی کوچک‌تر از j (تعداد متغیرها) باشد. در این حالت، مقادیر ویژه ($\lambda_1,\lambda_2,\ldots,\lambda_j$) و بردارهای ویژه متناظر ($z_1,z_2,\ldots,z_j$) به ترتیب نزولی از بزرگ‌ترین مقدار ویژه به کوچک‌ترین مقدار ویژه انتخاب می‌شوند. سپس، مجذور مقدار ویژه را تک‌تک در بردار ویژه متناظر با آن مطابق معادله (۹) ضرب می‌کنند که هر یک از این عبارات ستونی از ماتریس بارهای عاملی را می‌سازند.

$$\left[\sqrt{\lambda_1}\times z_1,\ldots\ldots,\sqrt{\lambda_m}\times z_m\right] \qquad (۹)$$

در این معادله، $\sqrt{\lambda_1}\times z_1$ بار عاملی متغیرها روی عامل اول و $\sqrt{\lambda_m}\times z_m$ بار عاملی متغیرها روی عامل mام نامیده می‌شود [۶].

دوران عامل‌ها

برای تسهیل فرایند تفسیر عامل‌های اولیه عمل دوران انجام می‌شود. بی‌نهایت دوران برای محورهای عامل‌ها امکان‌پذیر است. برای خودکارسازی فرایند دوران از معیارهای مختلفی استفاده می‌شود. متداول‌ترین معیارهای دوران عبارت‌اند از [۶]:
- Varimax (Promax): در این معیار سعی بر بیشینه‌کردن واریانس مربع بارهای عاملی برای عامل‌هاست.
-Covartimax: در این معیار سعی بر بیشینه‌کردن واریانس مربع بارهای عاملی برای متغیرهاست.
با توجه به دوران عامل‌ها، متغیرهای مهم استخراج می‌شوند. یعنی مشخص می‌شود که کدام متغیر با کدام عامل، بار عاملی (همبستگی) بیشتری دارد. در نهایت، متغیری که بیشترین بار عاملی را با عامل مدنظر دارد، به عنوان نماینده آن عامل معرفی می‌شود [۶].

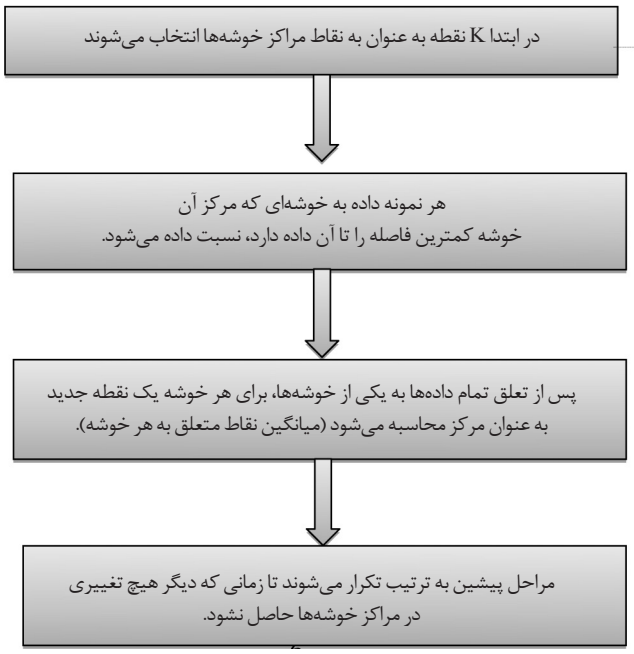
خوشه‌بندی

خوشه‌بندی یکی از مهم‌ترین ابزار کشف داده‌هاست که در کشف‌های تصادفی به‌کار گرفته می‌شود. در حال حاضر، اخذ دانش گلوگاهی عمده، در فرایند مهندسی دانش محسوب می‌شود. الگوریتم‌های یادگیری ماشین و داده‌کاوی با هدف استخراج دانش از داده‌ها، به عنوان روشی برای حل این مشکل مطرح است [۷]. خوشه‌بندی در واقع عملیاتی غیرنظارتی است. این عملیات هنگامی استفاده می‌شود که به دنبال یافتن گروه‌هایی از داده‌های مشابه باشید. بدون اینکه از قبل پیش‌بینی درباره شباهت‌های موجود داشته باشید. خوشه‌بندی معمولاً هنگامی استفاده می‌شود که به دنبال یافتن گروه‌هایی از مشتریان هستید که قبلاً شناخته نشده‌اند [۷].

الگوریتم K-Means یکی از پرکاربردترین الگوریتم‌های خوشه‌بندی است، حرف K که در اسم این الگوریتم وجود دارد، به این واقعیت اشاره دارد که هدف این الگوریتم پیداکردن تعداد ثابتی از خوشه‌ها بر اساس نزدیکی نقاط داده‌ها به هم است. الگوریتم خوشه‌بندی K-Means در شکل ۲ نشان داده شده است [۷،۸].

شاخص‌های اعتبارسنجی برای سنجش مقدار صحت نتایج خوشه‌بندی به منظور مقایسه بین روش‌های خوشه‌بندی مختلف یا مقایسه نتایج حاصل از یک روش با متغیرهای مختلف استفاده می‌شوند. طبق تعریف، خوشه‌بندی مطلوب است که در آن فاصله مراکز خوشه‌بندی از یکدیگر زیاد بوده و مقدار پراکندگی داده‌ها درون هر خوشه کم باشد. یکی از این شاخص‌ها، شاخص دیویس- بولدین (DB، Davies Bouldin index) است.

این معیار از شباهت بین دو خوشه (R_{ij}) استفاده می‌کند که بر اساس



شکل ۲- الگوریتم خوشه‌بندی K-Means.

پراکندگی یک خوشه (S_i) و عدم شباهت بین دو خوشه (d_{ij}) تعریف می‌شود. شاخص دیویس- بولدین برای خوشه‌بندی به شکل زیر تعریف می‌شود:

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} R_i \qquad (۱۰)$$

در این معادله، n_c تعداد خوشه‌هاست و R_i به شکل زیر تعریف می‌شود:

$$R_i = max\left(R_{ij}\right), i = 1 \dots n_c, j = 1 \dots n_c \qquad (۱۱)$$

در معادله (۱۱)، R_{ij} شباهت بین دو خوشه است که به شکل معادله (۱۲) تعریف می‌شود:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \qquad (۱۲)$$

d_{ij} و S_i در معادله (۱۲) به کمک معادله‌های (۱۳) و (۱۴) محاسبه می‌شوند:

$$d_{ij} = d(v_i, v_j) \qquad (۱۳)$$

در معادله (۱۳)، d تابع فاصله و V_i و V_j به ترتیب مراکز خوش iام و jام هستند، در نتیجه d_{ij} فاصله بین مراکز خوشiام و jام است.

$$S_i = \frac{1}{\|C_i\|}\sum_{x \in c_i} d(x, v_i) \qquad (۱۴)$$

در معادله (۱۴) منظور از C_i تعداد داده‌ها در خوشه iام و V_i مرکز خوشه iام است. این شاخص در واقع میانگین شباهت بین هر خوشه با شبیه‌ترین خوشه به آن را محاسبه می‌کند. می‌توان دریافت که هرچه مقدار این شاخص کمتر باشد، خوشه‌های بهتری تولید شده است [۹].

روش پژوهش

در این پژوهش ۱۰ متغیر به روش متریک اندازه‌گیری شد [۱۰] که نحوه اندازه‌گیری این متغیرها در جدول ۱ آمده است.

داده‌ها از ۱۳۰۰ نفر از دانشجویان دانشگاه یزد و ۷۱۲ نفر از سربازان پادگان شهدای جوادنیا در نزدیکی شهر قزوین به منظور تعیین شکل بدن آن‌ها گردآوری شده است. به این دلیل که ۱۰ مورد از اندازه‌گیری‌ها شامل خطا بود، از داده‌های اصلی حذف شدند. بنابراین تعداد کل اندازه‌گیری‌ها ۲۰۰۲ نفر است.

از آنجا که کارکردن با یک مجموعه ۱۰ تایی از متغیرهای اندام‌سنجی برای خوشه‌بندی بسیار دشواراست، به همین دلیل از روش تحلیل عاملی برای استخراج متغیرهای مهم استفاده شد. برای این کار نرم‌افزار آماری SAS به‌کار گرفته شد. در نهایت با طی مراحل ذکرشده در تحلیل عاملی و استفاده از روش مؤلفه اصلی، ماتریس بارهای عاملی با چرخش Promax برای متغیرها در جدول ۲ آمده است.

جدول ۱- نحوه اندازه‌گیری ۱۰ متغیر.

متغیرها	نحوه اندازه‌گیری
دورسینه	دور تا دور سینه به‌طوری که متر از برجسته‌ترین قسمت سینه بگذرد.
دور باسن	دور تا دور باسن به‌طوری که متر از پهن‌ترین بخش باسن بگذرد.
دورکمر	محیط دور کامل کمر از محل گودترین نقطه پشت کمر
دور ران	دور ران راست بلافاصله از زیر چین کپل راست
دور یقه	محیط کامل پایین دور گردن
دور زانو	دور تا دور زانو به‌طوری که متر از برجسته‌ترین بخش زانو بگذرد.
اندازه سرشانه	از خط گردن تا استخوان سرشانه
قد شلوار	طول عمودی از نقطه گودی پهلوی کمر تا کف پا بدون کفش
قد زانو	طول عمودی از نقطه گودی پهلوی کمر تا زیر استخوان زانو
قد	قد تمام از بالای سر تا پایین کف پا

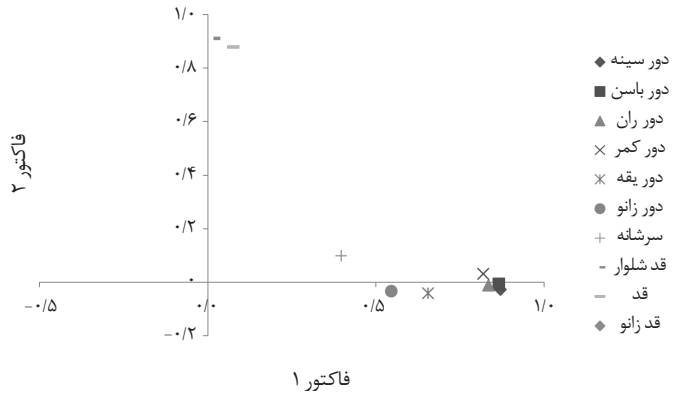
همان‌طورکه از جدول ۲ مشخص است، متغیرهایی که با بار عاملی بیشتر از ۰/۳۹ در عامل۱ قرارگرفته‌اند، شامل دور سینه، دور باسن، دور ران، دور کمر، دور یقه، دور زانو و اندازه سرشانه هستند و متغیرهایی که با بار عاملی بیشتر از ۰/۵۸ در عامل ۲ قرارگرفته‌اند، شامل قد شلوار، قد و قد زانو هستند.

بنابراین، چون متغیرهایی که در عامل ۱ قرارگرفته‌اند، بیشتر متغیرهای مربوط به دور اندام هستند، عامل ۱ عامل دور اندام نامیده می‌شود. متغیرهای دور سینه و دور باسن که بیشترین همبستگی را با عامل۱ دارند، به عنوان نماینده‌های این عامل معرفی می‌شوند و متغیرهایی که در عامل ۲ قرارگرفته‌اند، متغیرهای مربوط به طول اندام‌اند.

بنابراین عامل ۲ عامل طول اندام نامیده می‌شود و متغیر قد شلوار که بیشترین همبستگی را با عامل۲ دارد، به عنوان نماینده این عامل معرفی می‌شود. نمودار مربوط به ماتریس بارهای عاملی با چرخش Promax در شکل ۳ آمده است.

جدول ۲- ماتریس بارهای عاملی با چرخش Promax.

متغیر	عامل ۱	عامل ۲
دور سینه	۰/۸۶۹۳۸	-۰/۰۲۷۷۶
دور باسن	۰/۸۶۴۶۹	-۰/۰۰۶۵۶
دور ران	۰/۸۳۳۵۳	-۰/۰۱۱۸۹
دور کمر	۰/۸۱۸۱۴	۰/۰۳۰۹۶
دور یقه	۰/۶۵۴۱۳	-۰/۰۴۱۳۸
دور زانو	۰/۵۴۴۹۶	-۰/۰۳۳۶۳
اندازه سرشانه	۰/۳۹۵۹۹	۰/۰۹۸۷۸
قد شلوار	۰/۰۱۸۳۸	۰/۹۱۱۲۳
قد	۰/۰۷۵۲۶	۰/۸۷۹۱۹
قد زانو	-۰/۱۰۱۲۴	۰/۵۸۱۵۱



شکل ۳- نمودار مربوط به ماتریس بارهای عاملی با چرخش Promax برای متغیرها.

از شکل ۳ نیز مشخص است که دور سینه و دور باسن بیشترین همبستگی را با عامل ۱ دارند، به همین دلیل به عنوان نماینده عامل ۱ معرفی می‌شوند و قد شلوار که بیشترین همبستگی را با عامل ۲ دارد، به عنوان نماینده این عامل معرفی می‌شود، بنابراین متغیرهای مهم اندازه‌گیری دور سینه، دور باسن و قد شلوار هستند. در نتیجه متغیرهای دور سینه، دور باسن و قد شلوار به عنوان ورودی برای تحلیل خوشه‌بندی استفاده شد.

نتایج و بحث

در پژوهش پیش رو تابع K-Means برای خوشه‌بندی و تابع دیویس-بولدین (DB) برای ارزیابی تعداد خوشه‌ها (k) با استفاده از نرم‌افزار MATLAB اجرا شد. شایان ذکر است، هر خوشه‌بندی با تعداد خوشه‌های مشخص، ۱۰۰۰ مرتبه در برنامه اجرا شد و در نهایت بهترین تعداد خوشه‌بندی که دارای کمترین DB بود، گزارش شد. در جدول ۳ برای تعداد خوشه‌های مختلف مقدار DB محاسبه و تعداد نفراتی که از ۲۰۰۲ نفر اندازه‌گیری شده، در هر خوشه قرار می‌گیرند، مشخص شده است.

جدول۳- بررسی تعداد خوشه برای به‌دست آوردن بهترین DB.

تعداد خوشه (k)	بهترین مقدار DB *	تعداد نفرات در هر خوشه
۲	۱/۱۳۴۹	۸۰۷۱۱۹۵
۳	۰/۸۷۱۳	۹۸۲۱۱۱۷۲
۴	۰/۸۵۸۰	۶۰۵۵۵۸۳۴۴۴۹۵
۵	۰/۸۶۵۹	۴۹۴۵۶۵۲۹۷۲۳۵۴۱۱
۶	۰/۸۴۷۷	۵۴۷۲۳۲۳۸۹۳۴۰۲۶۸۲۲۶

* پس از ۱۰۰۰ بار خوشه‌بندی با انتخاب مقدار اولیه تصادفی.

در این پژوهش، افزون بر شاخص DB که معیاری برای مقدار صحت نتایج خوشه‌بندی است، تعداد نفرات در هر خوشه نیز حائز اهمیت است. زیرا هدف از گروه‌بندی شکل بدن افزون بر تشخیص شکل بدن، ارائه سامانه اندازه‌بندی بر اساس شکل بدن افراد نیز است. همان‌طورکه در جدول ۲ مشخص است، کمترین مقدار DB مربوط به تعداد خوشه ۶ تایی است، اما در این خوشه‌بندی تعداد نفرات هر خوشه برای گروه‌بندی جمعیت مناسب نیست.

بنابراین، بهترین تعداد خوشه برای گروه‌بندی شکل بدن افراد چهار است که مقدار DB آن از تعداد خوشه ۲، ۳ و ۵ کمتر و تقسیم‌بندی تعداد نفرات در هر خوشه نیز مناسب است. حال با درنظر گرفتن تعداد خوشه

جدول ۴- مشخصات بهترین تعداد خوشه.

تعداد نفرات در هر خوشه	قد شلوار (cm)		دور باسن (cm)		دور سینه (cm)		DB	تعداد خوشه (k)
	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین		
۴۹۵	۳/۰۴	۹۸/۰۲۰۲	۴/۴۹	۹۲/۳۴۰۴	۴/۲۹	۸۶/۸۴۰۴	۰/۸۵۸	۴
۶۰۵	۲/۸۹	۱۰۶/۳۸۵۱	۵/۶۲	۹۳/۸۸۶۰	۴/۳۱	۸۸/۹۵۷۰		
۵۵۸	۲/۴۵	۱۰۶/۵۶۷۰	۳/۹۹	۱۰۱/۰۸۶۰	۴/۱۲	۹۷/۷۰۱۶		
۳۴۴	۳/۸۱	۱۰۷/۴۴۷۷	۴/۰۱	۱۰۶/۶۹۱۹	۴/۷۷	۱۰۴/۳۱۴۰		

جدول ۵- مراکز خوشه و انحراف معیار متغیرهای اندازه‌گیری شده برای خوشه‌بندی با چهار خوشه.

تعداد نفرات در هر خوشه	قد (cm)		اندازه سرشانه (cm)		دور زانو (cm)		دور یقه (cm)		دور کمر (cm)		دور ران (cm)		دور ران (cm)	
	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین
۴۹۵	۴/۹۹	۱۶۹/۵۴	۱/۰۸	۱۶/۲۳	۲/۵۱	۴۰/۷۰	۱/۹۸	۳۶/۹۲	۵/۶۱	۸۱/۷۵	۳/۲۰	۵۳/۲۷	۳/۲۳	۵۲/۷۶
۶۰۵	۴/۶۸	۱۷۹/۶۲	۱/۱۶	۱۶/۸۱	۲/۴۲	۴۱/۴۴	۱/۸۷	۳۷/۰۸	۵/۷۲	۸۴/۶۴	۳/۵۴	۵۴/۰۳	۳/۱۰	۵۷/۰۷
۵۵۸	۴/۲۳	۱۸۰/۲۱	۱/۱۰	۱۷/۱۸	۲/۳۱	۴۳/۲۵	۱/۹۵	۳۹/۲۹	۵/۲۳	۹۳/۰۲	۳/۳۸	۵۸/۹۹	۳/۱۴	۵۷/۷۵
۳۴۴	۴/۸۱	۱۸۳/۱۲	۱/۱۲	۱۷/۷۴	۲/۲۲	۴۵/۴۷	۱/۸۸	۴۰/۲۸	۵/۳۶	۱۰۰/۲۶	۳/۲۵	۶۳/۴۱	۳/۲۳	۵۸/۳۵

چهار به عنوان بهترین تعداد خوشه، خصوصیات هر خوشه یعنی مراکز (میانگین) خوشه برای دور سینه، دور باسن و قد شلوار و نیز انحراف معیار این متغیرها در جدول ۴ آمده است.

در جدول ۴ مراکز خوشه‌ها روند افزایشی را نشان می‌دهد. همان‌طور که اندازه‌بندی‌های موجود دارای روند افزایشـــی بین اندازه‌های مختلف هستند. با درنظر گرفتن این موضوع می‌توان شکل بدن افراد را معیاری برای اندازه‌بندی دانست.

برای این چهار گروه شـــکل بدن، مراکز خوشه‌ها و انحراف معیار هفت متغیر دیگر نیز در جدول ۵ آمده اســـت. در این جدول نیز مراکز خوشه‌ها روند افزایشی را نشان می‌دهد. بنابراین، در جدول‌های ۴ و ۵ تمام مشخصات (مراکز خوشـــه و انحراف معیار برای متغیرهای اندازه‌گیری شـــده) چهار گروه شـــکل بدن آمده است که این مشـــخصات می‌تواند برای طراحان و تولیدکنندگان پوشـــاک مفید واقع شود و در واقع به اندازه‌بندی پوشاک مردانه کمک شایانی می‌کند.

نتیجه‌گیری

در ایـــن پژوهش، از روش تحلیل عاملی برای اســـتخراج متغیرهای مهم اندازه‌گیری و از روش خوشه‌بندی K-Means برای گروه‌بندی شکل بدن اســـتفاده شد. از بین ۱۰ متغیر اندازه‌گیری شده از جوانان (پسران) ایرانی سه متغیر دور سینه، دور باسن و قد شلوار به عنوان متغیرهای مهم معرفی شدند. در نهایت، شکل بدن افراد در چهار گروه دسته‌بندی و مشخص شد که در گروه‌ها شکل و اندازه بدن افراد مطابق با روند طبیعی افزایش می‌یابد. بنابراین، این گروه‌بندی شـــکل بدن را می‌توان معیاری برای اندازه‌بندی دانست که برای طراحان و تولیدکنندگان پوشاک می‌تواند مفید واقع شود.

مراجع

۱. گروه‌بندی شکل بدن با استفاده از روش خوشه‌بندی، K-mean هشتمین کنفرانس ملی نساجی. دانشگاه یزد: دانشکده مهندسی نساجی، ۱۳۹۰.

2. Chung M., Lin C., and Wang M., The development of sizing system for Taiwanese elementary-and high school students, *J. Ind. Ergonomic*., 37, 707-716, 2007.
3. Lin H., Hsu C., Wang M., and Lin Y., An application of data mining technique in developing sizing system for army soldieries in Taiwan, *J. China Text. Inst.,* 7, 1109-2750, 2008.
4. Thompson-Brenner H., Christina L., Michelle S., and Paul St., Representation of ideal figure size in Ebony magazine: A content analysis, *Body Image*, 8, 373-378, 2011.
5. Fan J., Yu W., and Hunter L., *EBook of Clothing Science and Technology: Clothing Appearance and Fit*, 2nd ed., Vol III, North America, 2004 .

۶. فرشاد فر ع.، *اصول و روش‌های آماری* چند متغیره، طاق‌بستان، ۱۳۸۴.

7. Macqueen J., Some methods for classification and analysis of multivariate observations, Procceding of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California, 1967.
8. Koikkalainen P. and Oja E., Self-organizing hierarchical feature maps, Proceding of International Joint Conference on Neural Networks, University of California, 1990.
9. Hand D., *Handbook of Data Mining: Principles of Data Mining*, 1st ed., Vol II, Taiwan, 2005.

۱۰. یونسی ن.، *الگو و برش لباس مردانه* به روش متریک، الزهرا، ۱۳۸۶.

Journal of
Textile Science
and Technology
Vol. 3, No. 3, 11-16
Autumn 2014
ISSN: 2151-7162

Factorial Analysis and K-Means Clustering Methods for Extraction of Human Body Size Parameters and Shape Groupings

R. Shamsaei, P. Payvandy*, and S. Fattahi

Department of Textile Engineering, Faculty of Engineering, Yazd University, Yazd, Iran

Received 15 August 2013; Accepted 19 October 2013

Abstract

Classification of different types of human body by adoption of meaningful rules obtained from data analysis of different parts of the body is very important as it is applied in many scientific methods. Applying data mining techniques on the obtained data and extraction of their meaningful rules could be helpful for different types of classification and ultimate determination of the garment of sizing system. It is important to note that recognition of differences in human body shape groupings and extraction of significant dimensional body variability can be used by dress designers and clothing manufacturers. In this study 10 body parts of 2002 Iranian men of 18 to 30 years old were measured for their body shape determination. In this study, factor analysis was employed to extract important variables and K-means clustering for body shape grouping. The obtained data were categorized into four groups of men's clothing sizes.

Keywords

body shape,
factor analysis method,
K-means clustering

(*) Address Correspondence to P. Payvandy, Email: peivandi@yazd.ac.ir