مجله علوم و فناوری نساجی
سال دوم، شماره ۲، شماره پیاپی ۳
صفحه ۱۰۳-۹۷، ۱۳۹۱
ISSN: ۲۱۵۱-۷۱۶۲

کاربرد شبکه عصبی خودسازمان‌دهنده براساس داده‌کاوی برای ایجاد جدول اندازه لباس

Applying Self-organizing Neural Network Based on Data Mining to Establish a Clothing Size Chart

فاطمه ماکویی، فرزانه میرجلیلی، محسن هادی‌زاده*، پدرام پیوندی

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

تاریخ دریافت: ۹۰/۹/۱۲ تاریخ پذیرش: ۹۰/۱۲/۸

چکیده

استخراج اطلاعات مناسب از میان انبوه داده‌ها و تبدیل آنها به دانش، نیازمند استفاده از روش‌های نوین است. داده‌کاوی یکی از این ابزارها و رویکردهاست که در فضای مدیریت دانش به کشف دانش از پایگاه داده‌ها کمک می‌کند. شبکه‌های عصبی خودسازمان‌دهنده بی‌شک از کارآمدترین شبکه‌های موجود برای داده‌کاوی است. پژوهش حاضر، جدول اندازه لباس را معرفی می‌کند که براساس اندازه‌های گرفته شده از ۲۰۰۰ زن یزدی است. متغیرهای ورودی مهم معرفی شده به شبکه عصبی شامل اندازه سه ناحیه دور سینه، دور کمر و دور باسن است. در این پژوهش، گروه‌های اندازه لباس متفاوت با استفاده از مدل شبکه عصبی خودسازمان‌دهنده بررسی شده است. نتایج حاصل از به‌کارگیری شبکه‌های خودسازمان‌دهنده، نمایانگر وجود چهار گروه اندازه مؤثر است که بر اساس کمترین میانگین خطا دسته‌بندی شده‌اند.

کلمات کلیدی

شبکه عصبی مصنوعی،
داده‌کاوی،
جدول اندازه لباس،
تحلیل خوشه‌ای

مقدمه

در بحث لباس، واژه سایز به معنای اندازه لباس به‌کار می‌رود. هر فرد با توجه به اینکه دارای شکل بدن متفاوتی است، اندازه لباس مخصوص به خود دارد. برای تولید پوشاک، تولیدکننده باید مجموعه‌ای از اندازه‌های کلیدی بدن مانند دور سینه، دور کمر و دور باسن را در اختیار داشته باشد. بدین منظور جدول‌های اندازه لباس تهیه شده‌اند که حاوی اندازه‌های مورد نیاز است. این جدول‌ها براساس اندازه‌های متوسط بدن انسان ایجاد می‌شوند و می‌توانند برای تهیه الگوی پوشاک استفاده شوند. همان‌طور که گفته شد، اندازه هر بدن با بدن دیگر متفاوت است، از این رو، جدول‌های اندازه لباس براساس جنسیت، گروه سنی، شکل بدن و موقعیت جغرافیایی دسته‌بندی می‌شوند. بنابراین ممکن است، هر کشوری جدول اندازه لباس ویژه‌ای داشته باشد، به‌طوری که نحوه بیان و سامانه اندازه لباس‌ها با یکدیگر متفاوت است [۱].

داشتن جدول اندازه لباس برای صنعت پوشاک و تولیدکنندگان امری کلیدی است، به‌طوری که پژوهش‌های متنوعی در این زمینه انجام شده است. Chung و همکاران در سال ۲۰۰۷، پژوهشی برای تعیین سامانه اندازه برای پوشاک دانش‌آموزان راهنمایی و دبیرستان در کشور تایوان انجام داده‌اند که با درنظر گرفتن سه معیار گروه سنی، جنسیت (مرد و زن) و نوع پوشاک (بالاتنه و پایین‌تنه)

*مسئول مکاتبات، پیام نگار: hadizadeh@yazduni.ac.ir

۱۲ســـامانه اندازه لباس تدوین شد. این پژوهش، دارای مزایای مقدار همپوشانی بیش از ۸۵ درصد و تعداد کم گروه‌های اندازه بود [۲].

Zheng و همـــکاران در ســـال ۲۰۰۷، توســـعه ســـامانه جدید اندازه‌گیری ســـینه‌بند را در کشور چین بررســـی کرده‌اند. سامانه جدید اندازه‌گیری در این پژوهش، برای زنان چینی با استفاده از دو متغیر قابل کنترل محیط زیر ســـینه و نســـبت عمق به عرض سینه توسعه داده شده است [۳].

Hsu و همکاران در سال ۲۰۰۷، پژوهشی را در زمینه استفاده از داده‌کاوی در استخراج قواعد معنی‌دار، برای دسته‌بندی اندام مردان مسن در کشور تایوان انجام داده‌اند. در نتیجه، شش قانون دسته‌بندی تیپ‌های اندامی انسان با اســـتفاده از تحلیل دو مرحله‌ای داده‌کاوی به‌دست آمد. سامانه دســـته‌بندی به‌دست آمده نقش مهمی را برای تولیدکنندگان پوشاک ایفا می‌کند[۴]. همچنین وی در سال ۲۰۰۸، تعیین شکل بدن زنان را با به‌کارگیری نسبت اندازه دور سینه به دور کمر، در کشـــور تایوان بررسی کرده اســـت. نتایج این پژوهش برای بهبود کیفیت محصولات تولیدی، راهکارهای مؤثر توسعه و استخراج جدول‌های اندازه‌گیری بدن را در اختیار تولیدکنندگان پوشاک قرار داده است[۵].

نکته مهم در تعیین ســـامانه اندازه پوشـــاک، ایجاد کمترین تعداد گروه‌های اندازه با حداکثر مقدار پوشـــانندگی جامعه اســـت. ایجاد اندازه‌های متعدد، مشکل تولید و انبار بیش از حد از یک نوع محصول و افزایش هزینه‌های مربوط به انبارداری را برای تولیدکنندگان پوشاک در پـــی دارد. همچنین، درنظر نگرفتن تفـــاوت اندام‌ها و اندازه‌های مختلف در تولید پوشاک، عدم راحتی و در نهایت نارضایتی مشتریان را در برخواهـــد داشـــت[۶]. در حال حاضر، در ایـــران جدول اندازه لباس برای هیچ‌یک از گروه‌های جنســـی و سنی (زن، مرد، کودک و نوجوان) وجود ندارد. بنابراین، از جدول‌های اندازه سایر کشورها مثل آلمان و انگلستان استفاده می‌شـــود. به‌کارگیری این جدول‌ها باعث بروز مشکلاتی در زمینه مناسب و اندازه بودن لباس می‌شود.

داده‌کاوی، فراینـــد تجزیـــه و تحلیل داده‌هـــا از زوایای مختلف و درگیرکـــردن روش‌هـــا و الگوریتم‌های مختلف بـــرای خلاصه‌کردن داده‌ها به اطلاعات مفید اســـت. پیشـــینه طرح موضوع داده‌کاوی به دهه ۱۹۸۰ برمی‌گردد. در روش‌های داده‌کاوی از بخشی از علم آمار به نام تحلیل اکتشـــافی داده‌ها اســـتفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تأکید می‌شود [۷]. افـــزون بر این، داده‌کاوی با هوش مصنوعـــی ارتباط تنگاتنگی دارد. بنابرایـــن می‌تـــوان گفت، در روش‌هـــای داده‌کاوی نظریه‌های پایگاه داده‌ها، هوش مصنوعی و علم آمار را در هم می‌آمیزند تا زمینه کاربردی فراهم شود.

هر چه حجم داده‌ها بیشـــتر و روابط میان آنها پیچیده‌تر باشـــد، دسترسی به اطلاعات نهفته در میان داده‌ها مشکل‌تر می‌شود و نقش داده‌کاوی به عنوان یکی از روش‌های کشف دانش، روشن‌تر می‌شود. روش‌هـــای داده‌کاوی مختلفـــی همچون قوانین رابطـــه‌ای، درخت تصمیم‌گیـــری (decision tree)، شـــبکه‌های عصبی مصنوعی و سری‌های زمانی در پژوهش‌های مختلف استفاده شده‌اند [۸،۹].

شبکه‌های عصبی مصنوعی خودسازمان‌دهنده کاربردهای فراوانی در زمینه‌های مختلف علوم همچون مهندســـی و پزشـــکی دارد که مهم‌تریـــن آنها اســـتخراج داده‌ها و تحلیل فضاهای پیچیده اســـت [۱۰]. از کاربردهای این نوع از شـــبکه‌ها می‌توان به خوشـــه‌بندی، تشـــخیص الگو، تحلیل تصاویر و اصوات و تشخیص خطا اشاره کرد. Min و Han در ســـال ۲۰۰۵ برای شناســـایی سرویس‌های مورد علاقه مشتریان از روش خوشه‌بندی شبکه عصبی مصنوعی استفاده کرده‌اند [۱۱].

یکی از مباحث جدید در بازاریابی براســـاس اطلاعات پایگاه داده، ارائه محصول مناسب به مشتری است. از مسائل بسیار مهم در خرید لباس، داشـــتن اطلاعات کافی درباره اندازه لباس است. اگر لباس در سامانه‌ای استاندارد دوخته شده باشد، اندازه لباس روی برچسب آن موجود است. جالب توجه آنکه برخی از افراد به‌ویژه خانم‌ها در ایران با توجه به عدم تناســـب نیم‌تنه پایینی با نیم‌تنه بالایی، نمی‌توانند از اندازه‌ای واحد برای یک لباس استفاده کنند. بنابراین مجبور هستند، انـــدازه بالاتنه را متفاوت و در اغلب اوقـــات کوچک‌تر از اندازه لباس برای پایین‌تنه تهیه کنند. بنابراین توســـعه ســـامانه اندازه لباس از اطلاعات ابعادی به روز و منطقه‌ای و همچنین تعیین قواعد معنی‌دار از اندازه‌گیری‌های مربوط به بخش‌های مرتبط با بدن انسان برای هر کشور اهمیت ویژه‌ای دارد.

در این پژوهش سعی شده اســـت، با استفاده از داده‌های به‌دست آمـــده از اندازه‌گیری‌ها و بهره‌گیـــری از روش‌های داده‌کاوی، جدول جدید اندازه لباس برای خوشه‌بندی اندام زنان یزدی ارائه شود.

روش پژوهش

شبکه‌های عصبی خودسازمان‌دهنده

شـــبکه عصبی مصنوعی را می‌توان در موارد مختلف به‌کار گرفت از جمله ذخیره‌کردن و بازبینی داده‌ها، گروه‌بندی شـــکل‌ها و داده‌ها، انجام یک نگاشـــت کلـــی از یک مجموعه ورودی بـــه یک مجموعه خروجی و خوشه‌بندی داده‌هایی که مشابه هم هستند. شبکه عصبی مصنوعی خودســـازمان‌دهنده، یک مدل یادگیری بدون ناظر اســـت. در شـــبکه عصبی مصنوعی خودسازمان‌دهنده به شبکه آموزش داده می‌شـــود تا بردارهای ورودی را طبق اینکه چگونه در فضای ورودی گروه‌بندی شده‌اند، خوشه‌بندی کند.

این مدل را ابتدا Kohonen (کوهنن) در ســـال ۱۹۸۱ توســـعه داده اســـت و گاهی به عنوان شبکه کوهنن نیز نامیده می‌شود [۱۲]. در این شبکه عصبی مصنوعی، خروجی سلول‌های عصبی با یکدیگر رقابت می‌کنند تا یکی از آنها که امتیاز بیشتری دارد، در رقابت برنده شده و در بین سایر سلول‌ها متمایز شود. این شبکه‌ها به علت وجود لایه رقابتی خود از سایرشبکه‌ها متفاوت‌اند.

الگوریتم شبکه کوهنن با n ورودی به ترتیب زیر است [۱۲]:

۱- تعیین مقادیر اولیه وزن‌های شبکه به‌طور تصادفی

$\omega_{ij}(t) \quad (0 \le i \le n-1)$

w_{ij} (t) مقدار ضرایب وزنی از ورودی i به نرون (neuron) خروجی j باشد.

۲- واردکردن ورودی‌ها به شبکه

بردارهـــای ورودی x_0 (t)، x_1 (t)، x_2 (t)، x_3 (t)، ...، x_{n-1} (t) وارد می‌شوند.

۳- محاسبه فاصله‌ها

فاصله بین بردارهای ورودی و خروجی هر نرون به کمک معادله (۱) محاسبه می‌شود:

$$d_j = \sum_{i=0}^{n=1} \left(x_i(t) - \omega_{ij}(t)\right)^2 \qquad (۱)$$

۴- انتخاب کوتاه‌ترین فاصله

نرون خروجی دارای کوتاه‌ترین فاصله مشخص می‌شود.

۵- اصلاح ضرایب وزنی

ضرایب وزنی جدید با استفاده از معادله (۲) محاسبه می‌شود:

$$\omega_{ij}(t+1) = \omega_{ij}(t) + \alpha(t) \cdot h\,(t)[x_i(t) - \omega_{ij}(t)] \qquad (۲)$$

عبارت α(t) شـــاخص بهره (gain index) و h(t) تابع همسایگی (neighborhood function) است.

۶- الگوریتم از مرحله ۲ تکرار می‌شود.

تکرار تا زمانی انجام می‌شود که شبکه آموزش دیده باشد. به عبارت دیگر، مرحلـــه آموزش تا زمانی ادامه می‌یابد کـــه بردارهای وزن به حالت پایدار برسند و دیگر تغییر نکنند.

داده‌ها و روش‌ها

داده‌کاوی از جملـــه روش‌های نوین علمی اســـت کـــه در توصیف، تشریح، پیش‌بینی و کنترل پدیده‌ها به‌کار می‌رود. در شکل ۱ فرایند کســـب دانش از پایگاه داده‌ها به‌طورکلی نشان داده شده است [۷].

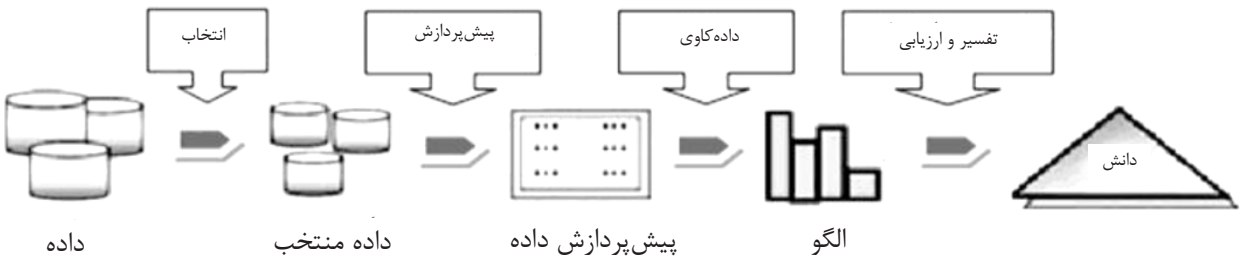
جدول ۱- شاخص‌های آماری متغیرهای ابعادی بدن.

متغیر ابعادی	حداکثر	حداقل	میانگین	میانه	مد	محدوده	انحراف معیار
دور سینه	۱۲۱	۸۳	۱۰۱/۷	۱۰۲	۱۰۰	۳۸	۸/۴۸
دور کمر	۱۰۹	۶۷	۸۳/۷	۸۳	۸۰	۴۲	۹/۷۹
دور باسن	۱۲۸	۹۰	۱۱۰/۲	۱۱۰	۱۱۰	۳۸	۸/۵۳
بلندی بالاتنه جلو	۴۵	۳۲	۳۹	۳۹	۴۰	۱۳	۳/۲۱
بلندی بالاتنه پشت	۴۳	۳۰	۳۵/۴	۳۴	۳۴	۱۳	۴/۱۳
کارور جلو	۳۹	۲۹	۳۴/۴	۳۴	۳۴	۱۰	۲/۴۴
کارور پشت	۴۶	۳۱	۳۷/۷	۳۸	۳۸	۱۵	۳/۱۴
بلندی سینه	۳۱	۱۷	۲۳/۳	۲۳	۲۲	۱۴	۳/۲۵
فاصله سینه	۲۵	۱۶	۱۹/۹	۲۰	۲۰	۹	۱/۷۶
سرشانه	۱۶	۱۰	۱۲/۶	۱۲/۵	۱۳	۶	۰/۹۴
دور گردن	۴۶	۳۲	۳۸/۵	۳۸	۳۶	۱۴	۲/۳۲
دور بازو	۴۲	۲۴	۳۵/۴	۳۴	۳۴	۱۸	۴/۰۶
دور مچ	۲۶	۱۷	۱۹/۷	۲۰	۲۰	۹	۲/۷۰

واحدها بر حسب سانتی‌متر است.

همان‌طور که ملاحظه می‌شود، یکی از گام‌های این فرایند، داده‌کاوی است. برای فرایند داده‌کاوی باید داده‌های مورد نیاز انتخاب شوند. در این پژوهش، جامعه آماری شامل اندازه‌های اندام زنان یزدی است که از این جامعه، نمونه‌ای با تعداد ۲۰۰۰ نفر بین ســـن ۲۰ تا ۵۰ سال گردآوری شده است.

برای انتخاب مناســـب داده‌ها در این فراینـــد، اطلاعات از مناطق مختلف شهر یزد، توسط بیست‌ونه خیاط باتجربه با استفاده از روش متریک اندازه‌گیری، سازماندهی و ذخیره شده‌اند. شاخص‌های آماری متغیرهای ابعادی بدن در جدول ۱ آمده است.



شکل ۱- فرایند تبدیل داده‌ها به دانش [۷].

در این مرحله، متغیرهای ابعادی بدن و داده‌ها بررسی و تحلیل می‌شوند. مطابق با پژوهش‌های انجام شده متغیرهای ابعادی مربوط به سه ناحیه دور سینه، دور کمر و دور باسن مهم‌تر از سایر متغیرها هستند. سپس، داده‌های خام این متغیرها پردازش شده و به داده‌هایی تبدیل می‌شوند که قابلیت تحلیل بیشتری داشته باشند. برای ادامه کار، داده‌هایی حذف شدند که درصد کمی از جامعه آماری را شامل می‌شدند. هنگامی که داده‌های مورد بررسی مشخص شدند، معمولاً تبدیل‌های خاصی نیاز است، روی داده‌ها انجام شود. تبدیل‌های ساده همچون تبدیل داده‌ها به داده‌های نرمال تعریف شدند. برای نرمال‌سازی داده‌ها از معادله (۳) استفاده شده است. با استفاده از این معادله داده‌های تبدیل شده دارای میانگین صفر و انحراف معیار ۱ است [۱۳]:

$$P_{ij}^{*} = \frac{P_{ij} - \bar{P}j}{S_j} \qquad (۳)$$

$\bar{P}_j$ مقدار میانگین و s^2_j واریانس j ام متغیر از همه نمونه‌هاست که P_{ij} مقدار اولیه متغیر و P_{ij}^{*} مقدار متغیر پس از نرمال‌سازی است. از متغیرهای ابعادی نرمال شده به عنوان ورودی شبکه عصبی مصنوعی استفاده شده است. اصولاً وارد کردن داده‌های خام باعث کاهش سرعت و دقت شبکه عصبی مصنوعی می‌شود. در مرحله بعد، متغیرهای ابعادی بدن خوشه‌بندی می‌شوند.

روش‌ها و الگوریتم‌های داده‌کاوی، راه‌های اجرای عملیات داده‌کاوی هستند. از آنجا که هر عملیات، نقاط ضعف و قدرتی دارد، در بین الگوریتم‌ها و روش‌ها، بهترین وجود ندارد و با توجه به داده‌ها و کارآیی مدنظر، باید ابزار یا ابزارهایی انتخاب و مدل مناسب طراحی و اجرا شود. در این پژوهش، برای خوشه‌بندی گروه‌های اندازه از روش شبکه عصبی مصنوعی خودسازمان‌دهنده به علت قابلیت‌های آن استفاده شده است. در این شبکه عصبی مصنوعی متغیرها بر اساس شباهت خود در خوشه‌های مختلفی قرار می‌گیرند. ساختار یک شبکه خودسازمان‌دهنده مانند ساختار یک شبکه پرسپترون یک‌لایه، دارای یک لایه ورودی و تعدادی نرون خروجی است. آموزش شبکه با n ورودی و m خروجی است. نرون‌های لایه ورودی وظیفه انتقال داده‌ها به شبکه را بر عهده دارند. تعداد نرون‌های موجود در لایه خروجی به مسئله مورد مطالعه وابسته است و توسط کاربر مشخص می‌شود. آموزش شبکه بر اساس کمترین فاصله اقلیدسی است. در واقع از میان نرون‌های خروجی، نرونی به عنوان برنده (خروجی شبکه) انتخاب می‌شود که در میان نرون‌های خروجی، کمترین فاصله اقلیدسی را با نمونه ورودی داشته باشد[۱۲].

در نهایت، ارزیابی مدل با استفاده از داده‌های آموزشی (۱۳۸۱نفر) و داده‌های آزمون (۱۵۳ نفر) با محاسبه مقدار خطا انجام می‌شود. شاخص‌های اعتبارسنجی متنوعی برای سنجش صحت نتایج خوشه‌بندی، مقایسه روش‌های مختلف خوشه‌بندی یا مقایسه نتایج حاصل از یک روش با پارامترهای مختلف پیشنهاد شده‌اند.

در این پژوهش، از شاخص دیویس - بولدین (Davies-Bouldin Index) استفاده شده است [۱۴]. این معیار از شباهت بین دو خوشه R_{ij} استفاده می‌کند که براساس پراکندگی یک خوشه ($var(C_i)$) و عدم شباهت بین دو خوشه c_i-c_j تعریف می‌شود. معمولا شباهت بین دو خوشه به کمک معادله (۴) تعریف می‌شود:

$$R_{ij} = \frac{\mathrm{var}(C_i) + \mathrm{var}(C_j)}{\|c_i - c_j\|} \qquad (۴)$$

که در آن c_i مرکز ثقل خوشه‌ها و ($var(C_i)$) پراکندگی خوشه‌هاست. شاخص دیویس - بولدین به کمک معادله (۵) تعریف می‌شود:

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i \qquad (۵)$$

که k تعداد خوشه‌هاست و R_i به کمک معادله (۶) محاسبه می‌شود:

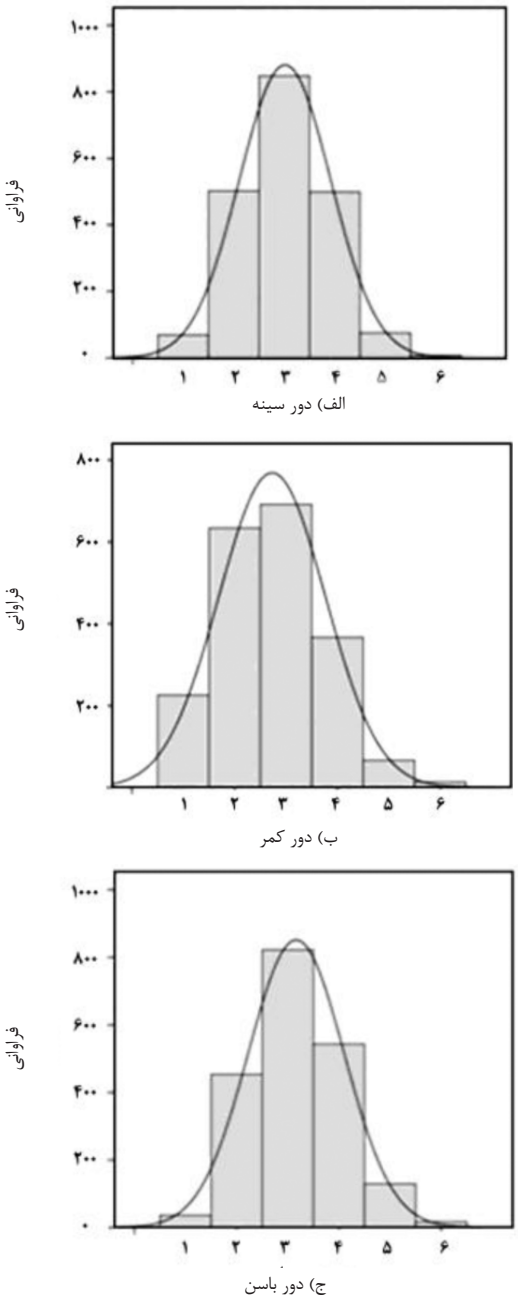
$$R_i = \max R_{ij} \qquad ,i\ ,j = 1,2,...,k\ ,i \neq j \qquad (۶)$$

این شاخص در واقع میانگین شباهت هر خوشه را با شبیه‌ترین خوشه به آن محاسبه می‌کند. هرچه مقدار این شاخص کمتر باشد، خوشه‌های بهتری تولید می‌شود. برای انجام این پژوهش، برنامه‌ای در محیط نرم‌افزار MATLAB با استفاده از سه متغیر ابعادی دور سینه، دور کمر و دور باسن نوشته شده است.

نتایج و بحث

هدف این پژوهش، ایجاد جدول اندازه مطابق با اندام بالاتنه زنان یزدی با استفاده از روش‌های داده‌کاوی و شبکه عصبی مصنوعی بوده است. داده‌ها پس از جمع‌آوری (داده‌های مربوط به سه ناحیه دور سینه، دور کمر و دور باسن)، به شش گروه آماری تقسیم شدند. انتخاب این تعداد گروه بر اساس وجود حداقل مقدار فراوانی‌ها (تعداد افراد) در هر گروه است. ابتدا با استفاده از نرم‌افزار SPSS جدول‌های فراوانی هر متغیر رسم شد. نمودارهای بافت‌نگار داده‌های دور سینه، دور کمر و دور باسن در شکل ۲ نشان داده شده است. با توجه به نمودارهای توزیع فراوانی مربوط به این سه متغیر ابعادی بدن مشاهده می‌شود، این داده‌ها از توزیع تقریباً نرمال پیروی می‌کنند.

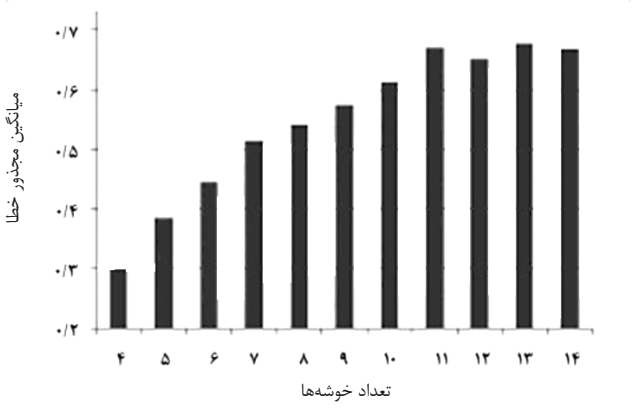
شناسایی ویژگی‌های گروه‌های مختلف متغیرهای ابعادی در تصمیم‌گیری و انتخاب راهکار مناسب درباره چگونگی تهیه جدول اندازه لباس، اطلاعات مفیدی را در اختیار قرار می‌دهد. بدین منظور، داده‌ها پردازش شدند. برای انجام این کار، ابتدا با توجه به اهمیت متغیرهای ابعادی، داده‌های با فراوانی کم مربوط به دورسینه حذف شدند. سپس از داده‌های باقی‌مانده، داده‌های با فراوانی کم مربوط به دور باسن و دور کمر حذف شدند (عدم تناسب معمول اندام). متغیر ابعادی مهم بعدی اندازه‌های طولی (اندازه‌های متناسب با قد افراد) است. در پایان از بقیه داده‌ها، اندازه‌های با فراوانی کم مربوط به قد جلو نیز حذف شدند.



شکل ۲- نمودارهای بافت‌نگار سه متغیر ابعادی بدن.

برای یافتن سایر اندازه‌ها که بیشترین ارتباط را با اندازه دور سینه، دور باسن و دور کمر دارد، ضریب همبستگی ده ناحیه باقی‌مانده با این سه ناحیه محاسبه و نتیجه گرفته شد که فقط کارور جلو بیشترین ارتباط را با اندازه‌های مزبور دارد. نواحی باقی‌مانده رابطه کمی با اندازه دور سینه، دور باسن و دور کمر داشتند. بنابراین، از حذف داده‌های مربوط به آنها صرف‌نظر شد. در نهایت از ۲۰۰۰ نفر، ۱۵۳۴ نفر (با مقدار ۷۶/۷ درصد پوشش) باقی ماندند و تجزیه و تحلیل‌ها روی ۱۵۳۴ نفر انجام شد.

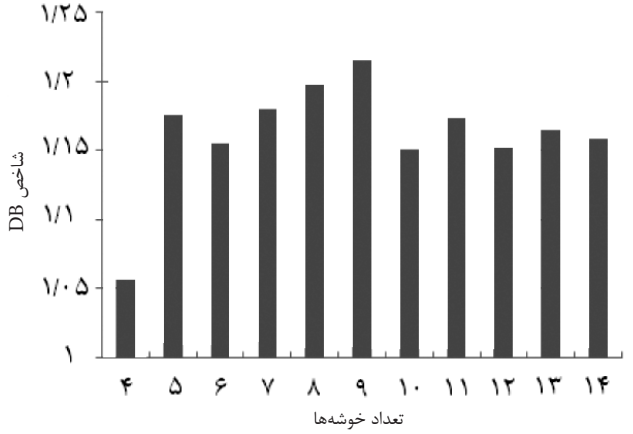
برای طراحی مدل شبکه عصبی مصنوعی، خروجی مدل m (تعداد



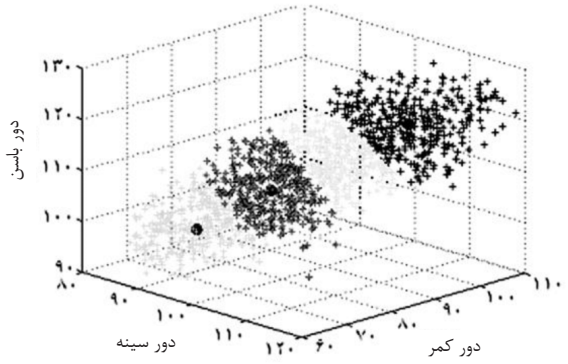
شکل۳- میانگین خطای خوشه‌بندی مدل‌های شبکه عصبی برای تعداد خوشه‌ها (10-fold cross validation).

خوشه‌ها یا گروه‌های اندازه) می‌تواند توسط پژوهشگر معین شود. در این پژوهش، به ترتیب ۴ تا ۱۴ خوشه انتخاب شده است. از الگوریتم کوهنن برای داده‌های آموزش استفاده شده است. برای ارزیابی مدل، داده‌های متغیرهای ابعادی به دو گروه داده‌های آموزش (۹۰٪) و آزمون (۱۰٪) تقسیم شدند و به دلیل انتخاب مقادیر اولیه وزن‌های شبکه به‌طور تصادفی، هریک از مدل‌های شبکه عصبی چند مرتبه اجرا شده‌اند.

روش ارزیابی k-fold cross validation برای پیش‌بینی خطای خوشه‌بندی N مشاهده به‌کار گرفته شده است که داده‌ها به‌طور تصادفی به k (عدد صحیح و به‌طور متداول k=۱۰) بخش تقسیم می‌شود. k-1 بخش برای آموزش مدل و یک بخش برای ارزیابی یا آزمون مدل استفاده می‌شود. این فرایند k-1 مرتبه تکرار می‌شود تا هر بار از یک بخش برای آزمون استفاده شود. برای به‌دست آوردن تخمین خطای کل، از k خطای به‌دست آمده میانگین گرفته می‌شود. از آنجا که انتخاب این بخش‌ها به‌طور تصادفی انجام می‌شود، خطای خوشه‌بندی متفاوتی به‌دست می‌آید و این روش



شکل۴- اعتبارسنجی شاخص دیویس- بولدین برای تعداد خوشه‌ها.



شکل۵- توزیع داده‌های موجود در چهار خوشه برای سه متغیر ابعادی بدن (واحدها بر حسب سانتی‌متر است).

ده مرتبه برای بهبود دقت مقدار خطا تکرار می‌شود [۱۵]. شکل ۳ میانگین خطای خوشه‌بندی تعداد خوشه‌های مختلف را برای داده‌های آزمون نشان می‌دهد. همان‌طور که مشاهده می‌شود، چهار خوشه برای گروه‌های اندازه لباس کمترین مقدار خطا را دارند.

سؤالی که در اینجا مطرح می‌شود، این است که بهترین مقدار برای تعداد بهینه خوشه‌ها کدام است؟ برای تعیین تعداد بهینه خوشه‌ها، همان‌طور که قبلاً اشاره شد، شاخص‌های مختلفی وجود دارند. در این پژوهش، از شاخص دیویس - بولدین استفاده شده است. مقادیر کوچک این شاخص به خوشه‌هایی مربوط می‌شود که فشرده بوده و مراکز آنها کاملاً از یکدیگر جدا شده باشند. در نتیجه تعداد خوشه‌هایی که شاخص دیویس - بولدین را کمینه می‌کنند، به عنوان تعداد بهینه خوشه‌ها درنظر گرفته می‌شوند. شکل ۴ اعتبارسنجی شاخص دیویس-بولدین را برای تعداد خوشه‌های مختلف نشان می‌دهد. همان‌طور که مشاهده می‌شود، کمترین شاخص مربوط به چهار خوشه برای گروه‌های اندازه لباس است.

همچنین از روش تجزیه و تحلیل واریانس‌ها مشخص شد، از خوشه‌های (گروه سایزی) ۶ تا ۱۴ برخی دارای میانگین نزدیک به هم هستند. نزدیک‌بودن میانگین خوشه‌ها نشان‌دهنده این موضوع است که این خوشه‌ها را می‌توان در قالب یک خوشه ارائه کرد. بر این اساس می‌توان پیشنهاد کرد، مدل طراحی شده جدول اندازه لباس با چهار خوشه یا گروه کارایی مناسبی دارد.

شکل ۵ توزیع داده‌های موجود در چهار خوشه حاصل از شبکه عصبی خودسازمان‌دهنده را بر اساس متغیرهای ابعادی دور سینه، دور کمر و دور باسن نشان می‌دهد. با توجه به تجزیه و تحلیل‌های انجام شده روی گروه‌های اندازه لباس و مطابق با شکل ۵ که پراکندگی داده‌ها حول نقاط مرکزی هر خوشه متمرکز است، در نهایت، جدول اندازه لباس با چهار گروه با کدهای ۳۸، ۴۱، ۴۴ و ۴۷ طراحی و در جدول ۲ محدوده این اندازه‌های کلیدی بدن آورده شده است. با توجه به نتایج جدول ۲ می‌توان گفت، دور باسن زنان یزدی ۸ cm از دورسینه آنها بزرگ‌تر است که در جدول مولر [۱۶] این اختلاف ۶ cm و در جدول متریک [۱۷] این اختلاف ۵ cm است. ۸ cm اختلاف نشان‌دهنده بزرگ‌بودن اندام

جدول۲- جدول اندازه لباس تهیه شده برای زنان شهرستان یزد.

اندازه‌های کلیدی بدن	کد اندازه			
	۳۸	۴۱	۴۴	۴۷
دور سینه	۹۱	۹۸	۱۰۵	۱۱۲
دور کمر	۷۳	۸۰	۸۷	۹۴
دور باسن	۹۹	۱۰۶	۱۱۳	۱۲۰
بلندی بالاتنه جلو	۳۸	۳۸/۹	۳۹/۸	۴۰/۷
بلندی بالاتنه پشت	۳۴/۷	۳۵/۷	۳۶/۷	۳۷/۷
کارور جلو	۳۲/۶	۳۳/۸	۳۵	۳۶/۲
کارور پشت	۳۵/۵	۳۶/۹	۳۸/۳	۳۹/۷
بلندی سینه	۲۱/۶	۲۳/۱	۲۴/۶	۲۶/۱
فاصله سینه	۱۸/۵	۱۹/۳	۲۰/۱	۲۰/۹
سرشانه	۱۲/۴	۱۲/۶۴	۱۳/۱۴	۱۳/۶۴
دور گردن	۳۶/۶	۳۷/۸	۳۹	۴۰/۲
دور بازو	۳۰/۱	۳۲/۳	۳۴/۵	۳۶/۷
دور مچ	۱۸/۸	۱۹/۴	۲۰	۲۰/۶
تعداد داده‌های موجود در هر خوشه	۳۶۷	۳۷۲	۴۰۷	۳۸۸

واحدها بر حسب سانتی‌متر است.

زنان یزدی است. همچنین می‌توان پی برد، تمام اندازه‌ها به شکل گام یکسان افزایش یافته‌اند، مثلاً دور سینه هر گروه ۷ cm از دورسینه گروه قبلی بزرگ‌تر است.

نتیجه‌گیری

در این پژوهش، از شبکه‌های عصبی خودسازمان‌دهنده و با توجه به ویژگی‌های کاربردی داده‌کاوی یعنی انتخاب اطلاعات مفید از بین انبوه داده‌ها و گروه‌بندی داده‌های مشابه هم، برای خوشه‌بندی و استخراج گروه‌های اندازه لباس استفاده شده است. نتایج حاصل نشان‌دهنده وجود چهار دسته مؤثر بر گروه اندازه لباس مربوط به بالاتنه زنان یزدی بوده است که براساس مقدار اثر و همبستگی بین آنها رتبه‌بندی و خوشه‌بندی شده‌اند. جدول اندازه ایجاد شده دارای گام اندازه (پرش سایز) یکسان است. به عبارت دیگر، تمام میانگین‌های مربوط به سه متغیر ابعادی، روند صعودی با گام مشخص دارند. بنابراین، داده‌کاوی به عنوان رشته‌ای علمی در زمینه بازیابی و استخراج اطلاعات می‌تواند نقش مهمی در دستیابی به جدول اندازه لباس داشته باشد. در گام بعدی پژوهش، برای اجرای مدل از لحاظ ارزیابی تجربی، چهار نوع لباس مطابق با چهار گروه اندازه بر اساس نتایج جدول تولید می‌شود، سپس نظرهای تعداد مشخصی افراد ثبت و ارزیابی می‌شود.

مراجع

1. Anbahan Ariadurai S., Nilusha T.P.G., Alwis T., and Manori Dissanayake D.M.R., An anthropometric study on Srilanka school children for developing clothing sizes, *J. Soc. Sci.*, 19, 51-56, 2009.
2. Jung Chung M., Fen Lin H., Jiun J., and Wang M., The development of sizing system for taiwanese-elementary-and high-school students, *Int. J. Ind. Ergonom.*, 37, 707-716, 2007.
3. Zheng R., Yu W., and Fan J., Development of A new chinese bra sizing system based on breast anthropometric measurements, *Int. J. Ind. Ergonom.*, 37, 695-705, 2007.
4. Hung Hsu C., Chin Chen S., and Shong Liu B., Employing data mining to identify the significant rules for classifying body types, *Proceedings of the International Conference on Computer Engineering and Applications*, 29-34, 2007.
5. Hsu C.H., Applying a bust-to-waist girth ratio approach to develop body measurement charts for improving female clothing manufacture, *J. Chinese Institute Ind. Eng.*, 25, 215-222, 2008.

۶. باقرزاده، ر. ، *کاربرد داده‌کاوی در صنعت نساجی، مطالعه موردی: تعیین سامانه سایزبندی پوشاک*، دانشگاه صنعتی امیرکبیر، ۱۳۸۷.

7. Herbert A.E., *Introduction to Data Mining and Knowledge Discovery, Two Crows Corporation,* 3rd ed., USA, 1999.
8. Song H.S., Kim J.K., and Kim S.H., Mining the change of customer behavior in an internet shopping mall, *Expert Systems with Applications,* 21, 157-168, 2001.
9. Chen M.C., Chiu A.L., and Chang H.H., Mining changes in customer behavior in retail marketing, *Expert Systems with Applications,* 28, 773–781, 2005.
10. Sambu S. and Klaus O., Self-organizing maps and clustering methods for matrix data, *Comput. Inform. Sci.*, 17, 1211–1229, 2004.
11. Min S.H. and Han I., Detection of the customer time-variant pattern for improving recommender systems, *Expert Systems with Applications,* 28, 189–199, 2005.

۱۲. منهاج، م.، *هوش محاسباتی، جلد ۱ مبانی شبکه‌های عصبی*، دانشگاه امیرکبیر، مرکز نشر دانشگاه صنعتی امیرکبیر، چاپ سوم، ۱۳۸۴.

13. Demuth H. and Hagan B.M., Neural Network Toolbox User's Guide, The Math Works, Copyright 1992 – 2006.
14. Davies D.L. and Bouldin D.W., A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2, 224-227, 1979.
15. Rodriguez J.D., Perez A., and Lozano J.A., Sensitivity analysis of k-fold cross validation in prediction error estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32, 569-575, 2010.

۱۶. هاکوپیان، س.، عزتی، م.ر.، *طراحی و برش لباس زنانه به روش مولر*، ۱۳۸۶.

۱۷. شهیر مفرح، پ.، *الگو (۱)*، شرکت چاپ و نشر کتاب‌های درسی ایران، آموزش فنی و حرفه‌ای، ۱۳۸۵.

Journal of Textile Science and Technology
Vol. 2, No.2, 97-103
Summer 2012
ISSN: 2151-7162

Applying Self-organizing Neural Network Based on Data Mining to Establish a Clothing Size Chart

F. Makooie, F. Mirjalili, M. Hadizadeh* and P. Payvandy

Department of Textile Engineering, Faculty of Engineering, Yazd University, P.O. Box: 89195-741, Yazd, Iran

Received 3 December 2011; Accepted 27 February 2012

Abstract

Extracting appropriate information from infinite number of data requires modern methods. Data mining is one of these tools and approaches that may help in the area of knowledge management in knowledge discovery from the databases. Self-organizing neural network is undoubtedly the most powerful neural networks for data mining. This paper presents a size chart that has been developed based on measurements taken from 2000 women of Yazd. The important input variables have been defined as the chest, waist and hip measurements. In this study, the different size clusters have been investigated through self-organizing neural network model. The results of using the self-organizing neural network indicate that there are four effective size groups that are categorized based on the lowest mean error.

Keywords

artificial neural network,
data mining,
clothing size chart,
cluster analysis

(*) Address Correspondence to M. Hadizadeh, Email: hadizadeh@yazduni.ac.ir