

上海地区成年女性批量定制服装号型分类研究

齐雪良¹, 袁惠芬^{1*}, 王旭^{1,2}

(1. “纺织面料”安徽省高校重点实验室 安徽工程大学, 安徽 芜湖 241000;

2. 安徽工程大学 纺织行业科技公共服务平台, 安徽 芜湖 241000)

摘要: 为确定批量定制服装合理板型数及提高合体度, 以上海地区520名成年女性为研究对象, 选择身高、胸围、腰围、领围为变量, 提出了基于K-means聚类分析的非国标规格的号型分类方法。探讨了初始聚心的选择和最佳聚类数的确定, 并以Calinski-Harabasz (CH) 指标、变异系数和相对偏差为指标, 对聚类效果进行评价。研究表明: 以国标号型规格为初始聚心, 号型被分为30类; 相同CH指标时, 以非国标规格号型为初始聚心, 号型被分为20类, 即板型数减少, 且变异系数和相对偏差明显降低。

关键词: K-means算法; 批量定制; 号型分类; 聚类分析

中图分类号: TS941-09

文献标识码: A

文章编号: 2095-414X(2016)04-0008-05

服装批量定制是在批量生产条件下满足顾客个性化需求和合体性要求的企业生产新模式^[1], 相比普通定制加工模式具有低成本和快速度等特点。企业在批量定制生产前, 需要对获取的人体测量数据进行号型归档。目前我国定制服装企业, 主要采用传统国标号型规格, 以人工归档方法进行归档, 存在周期长、错归、漏归等问题, 从而制约合同交货期和产品质量^[2]。近年来, 根据测量数据的相似性, 运用聚类分析方法, 采用计算机辅助进行服装号型归档已逐渐成为该领域的研究热点, 其中运用 K-means 算法进行号型分类较为常用。方方等^[3]运用 K-means 算法对 219 名女性进行了号型分类。郑艳等^[4]研究发现以胸腰差和罗氏指数为分类变量时运用 K-means 分类效果较好。目前基于 K-means 聚类的号型分类, 尚未形成标准的、系统的分类和评价方法。本文以上海地区成年女性为研究对象, 结合 K-means 聚类分析算法, 提出适合大批量定制服装号型分类和提高服装合体度的非国标号型分类方法, 并以反映类间类内距离比的 Calinski-Harabasz (CH) 值^[5]、变异系数和相对偏差作为评价聚类效果好坏的指标。K-means 聚类研究表明, 在相同 CH 指标条件下, 以最大最小距离法选择的非国标规格初始聚心的号型分类结果与以国标 GB/T 1335.2-2008 规格初始聚心按一次性择近原则分类结果对比, 板型数从 30 减少为 20。此外, 变异系数和相对偏差, 非国标规格号型也明显优于国标规格。研究结果为确定批量定制服装合理板型数及提高服装合体度具有一定参考价值。

一、实验部分

(一) 实验对象

上海市某行政事业单位 520 名 30 岁-55 岁的成年女性, 身高 150~175cm 之间。

(二) 测量要求

被测试者穿紧身内衣, 不能佩戴首饰和手表, 两脚并拢目视前方, 手臂自然下垂。

(三) 测量项目

根据要求本次测量项目包括: 身高、胸围、腰围、领围。每项目重复测量 3 次取均值^[6]。

二、实验数据分析

聚类分析前的数据预处理^[7], 包括异常值和正态分布检验。基于 Walsh 检验和一般经验法^[8], 本文 520

*通讯作者: 袁惠芬 (1972-), 女, 教授, 研究方向: 服装数字化。

基金项目: “纺织面料”安徽省高校重点实验室开放基金 (2015FZ001); 安徽工程大学研究生创新项目 (2015)。

个样品各测量项目无异常值且满足正态分布。

（一）以国标号型为初始聚心的 K-means 聚类

根据 GB/T 1335.2-2008 女子服装号型和南京某定制服装公司胸围与领围的搭配关系，根据测量数据分布范围，选择国标号型初始聚心 47 个（略）。运行 SPSS 软件，读入初始聚心，以仅分类方式执行 K-means 聚类。聚类结果发现有 17 个号型类别无样品分入，即 520 个样品被分为 30 类，其结果如表 1。

表 1 以国标号型为初始聚心的 K-means 分类结果

身高（cm）	胸围（cm）	腰围（cm）	领围（cm）	类内个数
150	76	60	35	4
150	80	64	35	6
150	84	68	36	8
150	88	72	37	1
155	76	60	35	3
155	80	64	35	41
155	84	68	36	42
155	88	72	37	22
155	92	76	38	6
155	96	80	39	2
160	76	60	35	4
160	80	64	35	42
160	84	68	36	91
160	88	72	37	52
160	92	76	38	29
160	96	80	39	7
160	100	84	40	8
165	80	64	35	8
165	84	68	36	40
165	88	72	37	33
165	92	76	38	21
165	96	80	39	12
165	100	84	40	7
170	84	68	36	10
170	88	72	37	10
170	92	76	38	3
170	96	80	39	5
170	100	84	40	1
175	88	72	37	1
175	96	80	39	1

（二）以非国标号型为初始聚心的 K-means 聚类

1. 非国标号型初始聚心的选择

经典K-means算法聚类结果易受到初始聚心的影响。本文以最大最小距离法产生出非国标规格的初始聚心。基于最大最小距离法的初始聚心选择是一种改进的方法，其思想是取尽可能远离的样品做为聚类中心，避免了随机确定初始聚心时出现由于聚心过于接近导致分类效果不理想的情况。其步骤如下^[9]：

- （1）取距离所有样本中心最近样品作为第 1 初始聚心 Z_1 ；
- （2）取距离 Z_1 最远的样品作为第 2 聚心 Z_2 ；
- （3）计算其他样品 X_i 与 Z_1 和 Z_2 间距离 d_{i1} 和 d_{i2} ，若 $\max\{\min(d_{i1}, d_{i2})\} \geq \theta \|Z_1 - Z_2\|$ ，（ θ 为比例系数， $\max\{\min(d_{i1}, d_{i2})\}$ 表示第 i 次聚类中心最大改变量），则样品 X_i 为第 3 聚心 Z_3 ，否则归入到其中一个聚类中；
- （4）重复上述操作，直到找不到符合条件的新聚心为止。

取 θ 为 0.17 时, 根据改进 K-means 算法选取非国标初始聚心 29 个 (Z1,Z2,...,Z29), 如表 2 所示。

表 2 非国标初始聚心与聚类结果

初始聚心	身高 (cm)	胸围 (cm)	腰围 (cm)	领围 (cm)	类内个数
Z1	160 (159)	84 (84)	72 (71)	36 (36)	99
Z2	158 (158)	108 (108)	88 (88)	41 (41)	1
Z3	174 (170)	96 (96)	80 (82)	39 (39)	6
Z4	152 (151)	76 (77)	60 (64)	34 (34)	8
Z5	154 (155)	96 (93)	80 (80)	39 (39)	8
Z6	170 (168)	88 (88)	74 (74)	37 (37)	17
Z7	165 (164)	100 (102)	90 (88)	40 (40)	7
Z8	163 (161)	92 (92)	84 (81)	38 (38)	12
Z9	150 (154)	86 (84)	73 (71)	36 (36)	42
Z10	164 (163)	80 (80)	64 (67)	35 (35)	22
Z11	155 (157)	100 (99)	88 (88)	39 (40)	7
Z12	153 (155)	80 (80)	68 (67)	35 (35)	37
Z13	160 (160)	92 (92)	76 (78)	38 (38)	25
Z14	163 (164)	100 (96)	82 (82)	40 (39)	12
Z15	158 (157)	76 (76)	64 (65)	34 (34)	13
Z16	168 (165)	84 (84)	68 (71)	36 (36)	51
Z17	155 (157)	88 (88)	78 (75)	37 (37)	38
Z18	167 (166)	94 (92)	79 (78)	38 (38)	15
Z19	165 (162)	88 (88)	78 (75)	37 (37)	62
Z20	159 (159)	80 (80)	68 (68)	35 (35)	38
Z21	163	104	86	40	/
Z22	158	96	84	39	/
Z23	168	96	84	39	/
Z24	150	76	66	34	/
Z25	165	84	74	36	/
Z26	163	85	68	37	/
Z27	160	88	74	37	/
Z28	163	88	72	37	/
Z29	155	84	71	36	/

2. 聚类数 K 的确定

基于 K-means 的聚类分析, 聚类数的选择对聚类结果产生重要影响^[10]。对于大批量定制服装数据, 聚类数 K 的确定直接关系到板型数及服装合体度。本文结合 CH 指标及与国标号型分类情况, 进行聚类数 K 的确定。CH 指标的计算如式 (1)。

$$CH = D_2 / D_1 \tag{1}$$

其中: D_1 表示类内距离和, 如式 (2) ; D_2 表示类间距离和, 如式 (3) ;

$$D_1 = \sum_{i=1}^k \sum_{p \in C_i} |p - x_i| \tag{2}$$

$$D_2 = \sum_{i=1}^k |x_i - \bar{x}| \tag{3}$$

其中: p 为 C_i 类的样品, x_i 为 C_i 类的聚心, $\bar{x}$ 为样本均值。

通常 CH 越大, 聚类效果越好, 但聚类数 K 会增加, CH 越小, 聚类效果越差, 聚类数 K 会减少, 导致服装合体性差。故为节约成本, 可在不影响聚类效果的情况下, 尽量减少板型数。本文以非国标规格为初始聚心的聚类数 K, 是以国标聚类结果 (K=30, CH=0.2166) 为基准, 根据 CH 值接近为原则确定。运行 SPSS 软件, 执行 K-means 算法, 读入 29 个非国标号型初始聚心, 以迭代 (最大 20 次) 并分类方式将

520 个样品依次从 K=2 到 K=29 进行 28 次聚类。注意到当 K=20 时，即 520 个样品分为 20 类时 CH=0.2186，接近按国标 K=30 的 CH 值，故最终确定将 520 个样品分为 20 类，即按照前 20 个初始聚心 (Z1,Z2,...,Z20)，执行聚类，结果如表 2 所示，其中括号内的数据为经过 K-means 聚类后，各类的均值并按四舍五入取整，即产生的 20 个非国标的规格，将作为板型设计的依据。上述结果表明，本文提出的按非国标初始聚心的 K-means 聚类结果，在和国标归档相比具有更少的板型数，且按 CH 指标分类效果接近。

3. 合体度的评价指标

CH 指标对分类数的确定有一定的参考意义，反映了聚类效果的好坏。从定制服装合体性角度，评价每个类内样品测量项目和最终聚心（类均值）的变异系数和相对偏差，在一定程度上更具有指导意义。变异系数 cv 和相对偏差分别是衡量数据变异和偏离均值程度的指标，计算式分别如式（4）、（5）：

$$cv = \frac{S}{\bar{x}} \times 100\%$$

(4)

$$rd = \frac{x_i - \bar{x}}{\bar{x}} \times 100\%$$

(5)

其中： x_i ， $\bar{x}$ ，s 为某测量项目在某类中的测量值、均值和标准差。

（三）不同初始聚心 K-means 聚类效果的比较

为了对比按照国标号型与按非国标规格为初始聚心的 K-means 聚类效果，分别计算了两种聚类结果各测量项目的变异系数和相对偏差，结果如表 3 所示。

表 3 国标与非国标分类变异系数与相对偏差对比								
	国标				非国标			
	身高	胸围	腰围	领围	身高	胸围	腰围	领围
总变异系数	20.81%	29.37%	46.10%	23.53%	20.33%	15.42%	36.58%	19.60%
平均变异系数	0.69%	0.98%	1.54%	0.78%	1.02%	0.77%	1.83%	0.98%
rd > 3%比例	0.58%	8.65%	44.81%	1.54%	0.38%	2.12%	10.38%	1.92%

表 3 中某测量项目总变异系数和平均变异系数，分别为各类别的变异系数之和及变异系数的均值。相对偏差大于 3% 样品所占比例，表示各类别中样品测量值超过该类别均值达 3% 样品所占的比例。

由表 3 可知，4 个测量项目中非国标分类结果总变异系数均小于国标分类结果，表明非国标方式聚类效果优于国标。但平均变异系数国标优于非国标，其原因是由于国标聚类数 30 大于非国标聚类数 20 所导致。测量项目腰围的变异系数最大，说明样本反映的成年女性群体腰围变异较大。

根据国标档差在 4 个测量项目中，相对偏差小于 3% 具有实际意义。各测量项目相对偏差腰围最大，其次是胸围、领围和身高。当相对偏差为 3% 时，国标和非国标腰围超标比例分别减少为 44.81% 和 10.38%，其他 3 个项目身高、胸围和领围超标比例国标和非国标对应关系为 0.58% 和 0.38%，8.65% 和 2.12%，1.54% 和 1.92%。综上分析，非国标分类结果变异系数和相对偏差均小于国标分类结果，表明数据偏离程度小，按非国标的分类结果，板型的合体性更好。

三、结论

（1）以最大最小距离法选择的非国标规格作为 K-means 聚类初始聚心，在相同 CH 指标时，具有比以国标规格为初始聚心更少的聚类数。

（2）以变异系数和相对偏差作为合体性依据，以非国标规格作为 K-means 聚类初始聚心，聚类效果板型更合体。

（3）样本反映成年女性 4 个测量项目中，变异系数从大到小依次为，腰围、胸围、领围、身高。

参考文献：

[1] 项铃铃. 基于服装批量定制的华东青年女性体形分析[D]. 无锡：江南大学，2007.
[2] 毋涛，王银. 服装批量定制量体服务系统的设计与实现[J]. 陕西科技大学学报，2011，29(8)：54-56.

- [3] 方方, 王子英. K-means 聚类分析在人体体型分类中的应用[J]. 东华大学学报: 自然科学版, 2014, 40(10): 593-598.
- [4] 郑艳, 张欣. 我国三地区女大学生体型分类研究[J]. 西安工程科技学院学报, 2004, 18(3): 210-214.
- [5] 刘燕驰, 高学东, 国宏伟, 等. 聚类有效性的组合评价方法[J]. 计算机工程与应用, 2011, 47(19): 15-17, 30.
- [6] 张宁, 王宏付. 基于三维人体测量的江浙女模特体型分类[J]. 纺织学报, 2012, 33(6): 71-75.
- [7] 齐静, 李毅, 张欣. 我国西部地区青年男性体型描述与体型分类研究[J]. 纺织学报, 2010, 31(5): 107-111.
- [8] 郁崇文, 汪军, 王新厚. 工程参数的最优化设计[M]. 上海: 东华大学出版社, 2003. 4-6.
- [9] 周涓, 熊忠阳, 张玉芳, 等. 基于最大最小距离法的多中心聚类算法[J]. 计算机应用, 2006, 26(6): 1425-1427.
- [10] 周世兵, 徐振源, 唐旭清. 新的 K-均值算法最佳聚类数确定方法[J]. 计算机工程与应用, 2010, 46(16): 27-31.

Investigation on Shanghai Adult Female Mass Customization Clothing Shape Classification

QI Xue-liang¹, YUAN Hui-fen¹, WANG Xu^{1,2}

(1. Anhui Provincial Key Lab of Textile Fabric, Anhui Polytechnic University, Wuhu Anhui 241000, China;

2. The Science and Technology Public Service Platform for Textile industry, Anhui Polytechnic University, Wuhu Anhui 241000, China)

Abstract: In order to determine reasonable patterns number of mass customization and improve clothing fitness, this paper analyzed 520 Shanghai adult female body data by K-means cluster algorithm with four classified variables such as height, bust circumference, waist circumference and collar circumference. A new non-GB shape classification method based K-means cluster was proposed. Initial cluster centers and optimal clusters were discussed. Classification result in each cluster was evaluated in Calinski-Harabasz index, coefficient of variation and relative deviation. It showed that the number of clusters was 30 when the GB shape was as initial cluster centers. According to parity of Calinski-Harabasz principle, the number of clusters of non-GB standard shape was 20 and coefficient of variation and relative deviation was decreased obviously.

Key words: K-means algorithm; mass customization; shape analysis; cluster analysis