Research on the method of children's body shape recognition based on big data technology of dressed human body

Xiaoning JING
School of Textile Science and Engineering, Tiangong University, 399 west Binshui Road, Xiqing District, Tianjin, China
Email: jingxiaoning@tiangong.edu.cn

Abstract. In this paper, a large number of children samples are collected to establish a database with detailed measurement index information. The data are analyzed and the figure of children are divided into three types. On the basis, the Naive Bayesian (NB) algorithm is applied to the study of figure discrimination, and the discriminant model is established. The discrimination of children figure is realized. At the same time, based on big data theory, the similar body of the test sample in the database is found through querying based on minimum difference algorithm. The unknown detailed measurement data of the test sample can be replaced by the similar body, to provide data support for the relevant application areas or dressed the human body that need to obtain detailed measurement information about the children.

1. Introduction
Classification and discrimination of human figure has been a hot research topic in the field of human body measurement and clothing engineering[1-3]. Many scholars have carried on the research. Many studies mainly aimed at the classification and discrimination of adult human figure[4-5]. Children are in the key period of growth and development, the research on their figure can guide more scientific guidance of children clothing structure design, to meet the physiological needs. On the other hand, the soaring consumption level of children clothing market and the growing demand for children health and safety clothing, also prompt the children clothing industry more concern about the characteristics of children figure to make a reasonable improvement of children clothing. Therefore, research on children figure has become a new hot spot. Naive Bayesian algorithm is a classic classification algorithm[6-7], which is widely used in various fields of research[8-10]. In this paper, this method is used to discriminate the girl's figure to achieve a ideal discriminant accuracy.

2. Data acquisition and pre-processing

2.1. Data acquisition
The measurement data of 1000 children samples are collected by non-contact 3D body scanning system of the French Lectra company.
Measurement object: 4-12 year old girl; Height range: 104.2-160.9 cm; Weight range: 14.0-50.0 kg.
Measurement indicators (27 indexes): height, head height, cervical height, cervical height in sitting, waist height, hip height, crotch height, shoulder height, chest height, neck depth, hip depth, chest depth, length between waist line and crotch, neck girth, over shoulder length, chest width, chest girth, back width, back length, distance between the waist and the hip, waist girth, hip girth, arm length, Wrist...
circumference, maximum thigh circumference, knee circumference, ankle girth.

2.2. Data pre-processing
Through the missing value analysis, singular value analysis, normal test and correlation analysis of the collected data, some data were eliminated, and finally, 868 measurement samples are retained.

3. Figure classification

3.1. Factor analysis to extract figure characteristic factor
Factor analysis is used to extract two feature characteristic factors according to the scree plot.

Table 1. Rotated component matrix.

| Component                  | 1   | 2   |
|---------------------------|-----|-----|
| Zscore(height)            | .365| .926|
| Zscore(head height)       | .232| .729|
| Zscore(cervical height)   | .370| .922|
| Zscore(cervical height in sitting) | .004| .337|
| Zscore(waist height)      | .388| .899|
| Zscore(hip height)        | .361| .876|
| Zscore(crotch height)     | .260| .925|
| Zscore(shoulder height)   | .315| .828|
| Zscore(chest height)      | .374| .900|
| Zscore(neck depth)        | .876| .045|
| Zscore(hip depth)         | .860| .038|
| Zscore(shoulder depth)    | .973| .384|
| Zscore(length between wrist line and crotch) | .279| .574|
| Zscore(shoulder length)   | .706| .401|
| Zscore(over shoulder length) | .857| .317|
| Zscore(chest width)       | .204| .428|
| Zscore(cheast girth)      | .868| .442|
| Zscore(back width)        | .715| .418|
| Zscore(back length)       | .243| .820|
| Zscore(distance between waist and hip) | .111| .503|
| Zscore(waist girth)       | .942| .184|
| Zscore(hip girth)         | .708| .545|
| Zscore(arm length)        | .407| .658|
| Zscore(Wrist circumference) | .822| .269|
| Zscore(maximum thigh circumference) | .939| .458|
| Zscore(knee circumference) | .808| .452|
| Zscore(ankle girth)       | .581| .484|

Figure 1. Scree plot.

The maximum variance is used to make the orthogonal rotation of component matrix, in order to describe the information of common factors. The rotated component matrix is shown in table 1.

The absolute value of index factor greater than 0.5 represents corresponding to the component. Therefore, the first principal component factor contains these indexes: neck depth, hip depth, chest depth, neck girth, over shoulder length, chest width, chest girth, back width, waist girth, hip girth, Wrist circumference, maximum thigh circumference, knee circumference, ankle girth. Obviously, the first principal component is a comprehensive characterization represented the lateral and circumference dimensions of human body.
The second principal component factor contains: height, head height, cervical height, waist height, hip height, crotch height, shoulder height, chest height, length between waist line and crotch, back length, distance between waist and hip, arm length. The second principal component factor represents height and length dimensions of human body. Therefore, according to the connotation of the principal components, the two figure characteristic factors can be named as the lateral body circumference factor and the longitudinal body scale factor.

3.2. Figure classification bases on k-means cluster analysis
According to the standard of two figure characteristic factor’s score, the comparison is made, which probability of no difference between classes is less than 5%. We can get it: when the cluster number is 3 that is the most ideal class. At this time the girl’s figure is divided into three types. The table of ANOVA is as follows when figure is divided into three types.

| Cluster | Error | F   | Sig. |
|---------|-------|-----|------|
| Mean Square | df | Mean Square | df |     |
| REGR factor score 1 for analysis 1 | 287.126 | 2 | .338 | 865 | .848385 | .000 |
| REGR factor score 2 for analysis 1 | 289.369 | 2 | .333 | 865 | .860320 | .000 |

Table 3. Number of cases in each cluster.

| Cluster | Cases |
|---------|-------|
| 1       | 236.000 |
| 2       | 230.000 |
| 3       | 402.000 |
| Valid   | 868.000 |
| Missing |       |

Figure type “3” contains the largest number in the three figure types, occupies 46.3% of the total. Followed by it, the proportion of figure type “1” is 27.2%. Figure type “2” is the least, accounting for 26.5% of the total sample.

3.3. Description of the three figure types
Through single factor variance analysis we can get the specific degree of 2 factors influencing on the 3 figure types. They are shown in the table 4 and the table 5. The factor 1 represents the lateral body circumference factor. The factor 2 represents the longitudinal body scale factor. The higher of graph feature’s score is, the greater of graph type’s influence is. Then the professional description of the figure can be made.

| Cluster Number of Case | N | 1 | 2 | 3 |
|------------------------|---|---|---|---|
| 1                      | 256 | -6507551 | 0.000 | 0.000 |
| 2                      | 462 | -390893 | 1.341367 |
| 3                      | 250 | 1.000 | 1.000 | 1.000 |

Table 4. REGR factor score 1 for analysis 1.

| Cluster Number of Case | N | 1 | 2 | 3 |
|------------------------|---|---|---|---|
| 1                      | 256 | -1.1017044 | 0.000 | 0.000 |
| 2                      | 236 | -1.493136 | 0.000 | 0.000 |
| 3                      | 402 | 0.7791050 | 0.000 | 0.000 |

Table 5. REGR factor score 2 for analysis 1.
According to the results of comparative analysis, the characteristics of three figure types are obtained.

**Table 6.** The characteristics of three figure types.

| figure | figure characteristics |
|--------|-------------------------|
| 1 (F)  | fat                     |
| 2 (ST) | short-thin              |
| 3 (T)  | tall                    |

The lateral and circumference dimensions of figure type “1” are the largest of all. The horizontal and vertical dimensions of figure type “2” are the smallest of all. And the longitudinal dimensions of figure type “3” are the largest of all.

### 4. Figure discrimination

In the premise of 27 measurement index data have been known, based on the classification results of girl's figure, the new figure discrimination can be researched. That is to say, we can know the figure type of the new girl sample.

#### 4.1. Bayesian theory

Bayesian decision theory method is a basic method in statistical model decision.

The basic idea is that:
- Conditional probability density parameter expression and a priori probability have been known.
- Using Bayesian formula to convert it to the posterior probability.
- According to the size of the posterior probability to make the decision classification.
- If \( P(A) \) is a prior probability or edge probability of \( A \).
- \( P(A|B) \) is known as the conditional probability of \( A \) after the occurrence of \( B \). It is called posterior probability of \( A \).
- \( P(B|A) \) is known as the conditional probability of \( B \) after the occurrence of \( A \). It is called posterior probability of \( B \).
- \( P(B) \) is a prior probability of \( B \).

Then,
\[
P(B | A) = \frac{P(A | B) P(B)}{P(A)}
\]

#### 4.2. Figure discrimination algorithm based on naive bayesian classification

Naive Bayesian method is a classification method based on the Bayesian theorem and the independent assumption of the characteristic condition.

According to Bayesian formula:
\[
P(B | A) = \frac{P(A | B) P(B)}{P(A)}
\]

Suppose \( T_i \) is the figure types. It is known by the classification results: \( i=1, 2, 3 \).

\( S_1 \) is the sample of a girl. \( S_1 = a_1, a_2, ... a_{27} \), \( a_1, a_2, ... a_{27} \), is the 27 measurement indexes of the sample.

Then, when the data of 27 measurement indexes in sample \( S_1 \) coexists and belong to the figure type \( T_i \),
\[
P(T_i | S_1) = \frac{P(S_1 | T_i) P(T_i)}{P(S_1)}
\]

The probability is

According to the Naive Bayesian classification principle, hypothesis \( S_1 = \{a_1, a_2, ... a_{27}\} \) is a sample to be classified. \( a_1, a_2, ... a_{27} \) is the 27 measurement indexes of \( S_1 \).

Calculate
\[
P(T_i | S_1), P(T_j | S_1), P(T_k | S_1)
\]

If
\[
P(T_k | S_1) = \max\{P(T_i | S_1), P(T_j | S_1), P(T_k | S_1)\}
\]

then \( S_1 \in T_k \).

The conditional probability estimates of each measurement index in each figure is obtained by those training samples which figure types have been known. That is,
\[ P(a_1 \mid T_i), P(a_2 \mid T_i), \ldots, P(a_{27} \mid T_i) \]

If each measurement index is in condition independent, the following deduction is based on Bias’s theorem. Because the denominator for all categories is constant, so we just would maximize the molecular.

\[ P(S_1 \mid T_i) \cdot P(T_i) = P(a_1 \mid T_i) \cdot P(a_2 \mid T_i) \ldots P(a_{27} \mid T_i) \cdot P(T_i) = P(T_i) \prod_{j=1}^{27} P(a_j \mid T_i) \]

\[ P(T_i \mid S_1) = \frac{P(T_i) \prod_{j=1}^{27} P(a_j \mid T_i)}{P(S_1)} \]

That is,

Namely, the probability of each measurement index of S1 respectively belong to Ti (i = 1, 2, 3) is calculated.

The calculated probability belonged Ti is the maximum, then S1 belongs to figure type Ti (i = 1, 2, 3).

For example, the calculation result of \( P(T_2) \prod_{j=1}^{27} P(a_j \mid T_2) \) is maximum. then S1 belongs to figure type T2.

5. Algorithm validation

The whole samples are verified by naive bayes algorithm to carry out the model of figure discrimination.

The output result of program is as follows:

0| True Class: 0, Predicted: 1, Confidence: 0.9999994944058159
1| True Class: 0, Predicted: 1, Confidence: 0.9999629044874304
2| True Class: 0, Predicted: 0, Confidence: 0.998503983737399
3| True Class: 0, Predicted: 0, Confidence: 0.999960513692685
4| True Class: 0, Predicted: 0, Confidence: 0.9878530888292812.

50 errors were made, 5.768905472636816% error rate

The error rate of the 50 samples is about 6%, and the correct rate of the algorithm is about 94%.

6. Similar body data acquisition

The main measurement indexes of new test sample are provided. That are height, chest girth, waist girth, hip girth, arm length, over shoulder length, back length and back width. The minimum difference algorithm is used to search the sample in the database which is closed to the data of test sample, that is the similar body sample.

Query threshold value is set according to size variation.

The corresponding data of similar body samples are compared with the actual data of test samples. The fitting diagram are drawn to verify the similarity. As shown in figure 2, 3, 4, the main measurement index data of test samples a, b, c are very consistent with the similar body samples. Under the premise of the database sample capacity is large enough, according to the principle of full sample analysis[11-13] and prediction principle of large data[14], It can be thought that we can use the similar body instead of the test sample, to provide 27 measurement indexes data information.

In this way, the cumbersome and inconvenient measurement work is omitted and the cost is reduced. The detailed information of girl’s body can be quickly output to support the related application fields of human body engineering and garment engineering.
Figure 2. Fitting of main measuring indexes of sample “a” and similar body.

Figure 3. Fitting of main measuring indexes of sample “b” and similar body.

Figure 4. Fitting of main measuring indexes of sample “c” and similar body.

7. Conclusions
In this paper, a large number of girl’s measurement data are collected. The figure classification is made. Based on Naive Bayesian algorithm, figure discrimination model is established. And the discrimination accuracy of the model is verified. The similar body of the test sample in the database is searched out according to the minimum difference algorithm. The main measurement index data of them are fitted. According to the similarity, the similar body can be used instead of the test sample, to provide more detailed data to support the data needs of relevant application areas.

References
[1] Summons K and Istook C L 2004 Female Figure Identification Technique (ffit) for Apparel PartI: Describing Females Shapes ( Journal of Textile and Apparel Technology and Management vol 4) pp 1-16.
[2] Vuruskan A and Bulgun E 2011 Identification of Female Body Shapes Based on Numerical Evaluations ( International Journal of Clothing Science and Technology vol 23) pp 46-60.
[3] Connell L J and Ulrich P 2002 Body Scan Analysis for Fit Models Based on Body Shape and Posture Analysis ( National Textile Center Annual Report vol 11) pp 31-37.
[4] Yang L and Chunting Y 2010 A Novel Method for Body Shape Analysis in E-mtm (Applied Mechanics and Materials vol 20-23) pp 1482-1486.
[5] Song H K and Ashdown S P 2011 Categorization of Lower Body Shapes for Adult Females Based on Multiple View Analysis (Source: Textile Research Journal vol 81) pp 914-931.
[6] Yaner RR 2006 An Extension of the Naive Bayesian Classifier (Information Sciences vol 176) pp 577-588.
[7] Flach P A and Lachiche N 2004 Naive Bayesian Classification of Structured Data (Machine Learning vol 57) pp 233-269.

[8] Derouin F and Sarfati C 2007 Naive Bayesian Classifier for Incident Duration Prediction (Transportation Research Board Meeting vol 51) pp 800-807.

[9] Ratanamahatana C A and Gunopulos D 2003 Feature Selection for The Naive Bayesian Classifier Using Decision Trees (Applied Artificial Intelligence vol 17) pp 475-487.

[10] Wang X Z and He Y L 2014 Non-naive Bayesian Classifiers for Classification Problems with Continuous Attributes (IEEE Transactions on Cybernetics vol 44) pp 21-39.

[11] Big data successfully predicted USA election [EB/OL]. [2014-02-25]. Http://www.china-bigdata.com.

[12] Mayerschonberger V and Cukier K 2013 Big data: a revolution that will transform how we live, work and think (England: Hodder Export Press) pp 39.

[13] Big data [EB/OL]. [2014-05-07]. Http://www. wikipedia.org.

[14] Alexander M and Konrad B 2014 Psycho-informatics: big data shaping modern psychometrics (Medical Hypotheses vol 82) pp 405-411.