Intelligent Size Table Generation

R. Ng[a], S.P. Ashdown[b], A. Chan[a]

[a] *Institute of Textiles and Clothing, Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China*
[b] *Department of Fiber Science and Apparel Design, Cornell University, Ithaca NY, 14850, USA*

Abstract

A size table is an efficient tool for improving the fitting of the massively produced garment. Traditionally, the size table is created based on the statistical tools based on linear size steps. Yet the statistical tools are much limited in the sense that it does not allow the user to fine-tune the criteria. So, in the current study, we have applied artificial intelligence technique to build the size table automatically according to the specification of the user. The result is a software system written in Mathematica™. The test data is part of the size USA. The performance of the size table is evaluated according to the cover factor of the size table.

Keywords: Apparel sizing; fit; anthropometric survey; Size table; AI;

1. Introduction

A size table is a chart of body dimensions of a selected population. For example, the USA population can be represented by the SizeUSA; the UK population can be represented by the SizeUK. From these databases of body measurements, it is possible to simply cluster the subjects into different sizes. The resulting classification forms a size table. This reduced body measurements set is most useful for the mass production of garments and other equipments. So, an appropriate size table is the key to one of the greatest challenges facing apparel companies today, which is to provide quality fit to a broadly defined target market.

In the past the lack of current anthropometric data to describe civilian populations has limited resolution of this problem. Apparel companies have relied on sizing systems developed from out-of-date anthropometric studies, adjusted based on their target market customers' feedback and product return data. Due to the lack of valid anthropometric data for the population, there has also been a lack of investigations of methods to convert population data to effective and appropriate size tables.

This situation has now changed as many anthropometric studies have recently been conducted including studies in Japan, England, France, South Korea, Mexico, China, and the U.S. using 3D body scanning technology. These studies are affordable and can provide reliable population data [1] [2]. This new technology will make it possible to conduct regular anthropometric studies to provide the apparel industry with useful current body measurement data. Apparel companies are now investigating ways to use these new data to create effective new sizing systems [3].

Many different methods have been used to develop size sets from anthropometric data in the past. The most common method relies on the selection of one or more key or control body measurements from the range of possible variables. These control measurements are selected based on the garment type and desired range of body types in the population to be accommodated, and are often based on industry practices rather than statistical analyses of the data. The sizing system is designed based on these key measurements. The number of sizes in the system is determined based on the range of control dimensions to be accommodated and the value of the intervals between sizes. The remaining dimensions of the garment are then set based on a set of secondary measurements. These measurements are often calculated using multiple regression analysis [4]. Though the decisions made to create a sizing system using this method are guided by statistics from the population data they are not directly derived from the data. Sizing systems created using this method have the advantage of being simple to implement as the sizes are proportional to one another, grading patterns is a simple process, and sizes can be selected by the consumer based on one or two body dimensions. However they are not directly based on the anthropometric data and they ignore variations in proportion in the population. Such sizing systems are most suitable for limited target market groups from the population with similar body proportions.

Other methods that have been proposed for creating sizing systems from anthropometric data that are more directly derived from the data include principal component analysis (PCM) [5] and various forms of optimization [6] [7] [8]. These methods have the advantage of deriving the size set directly form the data and hold promise to provide sizes that fit the range of variation in the population more effectively than systems created using current methods. Once sizing systems created using these methods are developed and tested, issues related to creating new grading and specification methods for manufacture, and new methods of selecting the correct size from the more complex resulting systems can be addressed. These methods are statistical and optimization in nature.

The proposed method is genetic algorithm (GA), which is one of the popular artificial intelligent methods. This method was chosen from three reasons. First, the user interface of GA is better than PCM, because GA does no require any intervention of the user during the search for the best size partition. Secondly, GA can have a higher computational performance than integer programming. Thirdly, the construction of a size table is basically a clustering problem. There is no expert rule known for the fuzzy logic (FL), and there is no predefined proper answer for training the artificial neural network (ANN). Hence, GA is a better choice among three popular artificial intelligent methods, namely FL, ANN and GA [9].

Furthermore, a size table can be either linear or nonlinear. A linear size table has a constant increment of measurements from one size to the next size. For a nonlinear size table, there are two typical possibilities. Firstly, the nonlinear size table can be combined from two linear subtables. This is a common practice in the fashion industry when the company wants to cover a large range of customers of different sizes. Such arrangement is usually referred as a bilinear approximation of the size spectrum. Secondly, a nonlinear size table can be made up of varying increments of measurements from one size to the next. Logically, the nonlinear size table should be able to perform better than a linear size table. In the current study, only the nonlinear size table is considered.

The evaluation of the performance of the size table building method is primarily based on the cover factor. The cover factor is the percentage of subjects who can be classified into any one of the sizes in a sizing system. It should be noted that subjects with odd sizes can exist in the population. Therefore, even the ideal cover factor of a practical sizing system is typically less than 100%.

In this article, we shall present the historical background for the SizeUSA. We explain how we have selected the women data and the primary body measurements for our study. We also describe how we preprocessed the data by identifying the outliners. Then, we formulate the size table generation problem in mathematical terms. Such a problem is further expressed in a way to match the format of GA. The program of determining the size table using GA was written in Mathematica™. Two size tables, one is produced by statistical method, namely PCM, and another one is produced by GA. The results are compared. Although the result is positive, meaning that the GA approach can produce a better quality size table, more testing will be needed.

2. Data collection

The testing data is based on the subset of the women data in the SizeUSA body measurement database. The data was collected in 2003 [10]. The SizeUSA project was financed by a group of USA industrialists as part of their effort in improving the quality of fitting and facilitating virtual fitting for the US fashion retailing industry.

2.1 Measurement selection

In the SizeUSA, there are 230 body measurements available in the database. Since, some of the measurements are for specific purposes, which are not commonly used in a size table, these measurements were ignored in the computation. Typically, for the size specification of a skirt, four measurements, namely waist, hip, waist-to-hip, skirt length are needed. In the database, there are four measurements of waist-to-hip: front, back, left and right. For retailing purpose, the back waist-to-hip was selected. In this article, these four measurements were used in the computation.

2.2 Data preprocessing

The raw data from SizeUSA was preprocessed. Originally, there are 6533 data. In this article, only 1000 data were used. Firstly, all incomplete subject records with missing values in these four measurements were removed. Secondly, outliners are defined to be lining outside the range of $\pm 3\sigma$ (standard deviation). For current study, all outliners remained in the data set. Thirdly, the descriptive statistics were calculated (Table 1) and the histograms of the data were plotted in Figure 1 to 4 for visual checking.

Table 1. Descriptive statistics of data set

Measurements	Field ID	N	Min (inch)	Mean (inch)	Max (inch)	Standard Deviation
Waist	W108	1000	24.70	35.40	57.32	5.55
Hip	W114	1000	32.47	42.46	63.85	5.03
Waist-to-Hip	W22-W23	1000	2.75	6.28	8.66	0.86
Length	W67R	1000	31.45	39.26	46.24	2.20

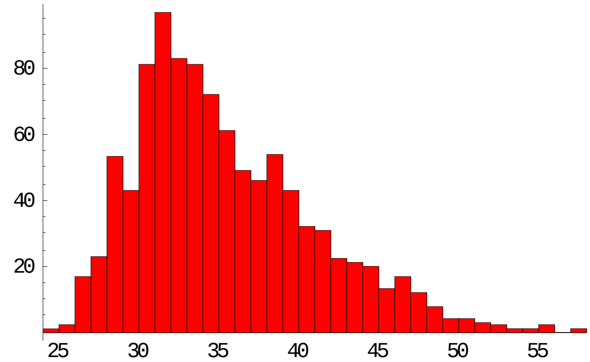

Fig. 1 Histogram of waist girth (Unit: inch)

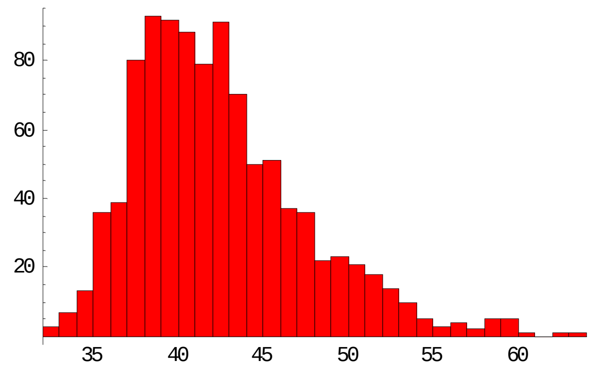

Fig. 2 Histogram of hip girth (Unit: inch)

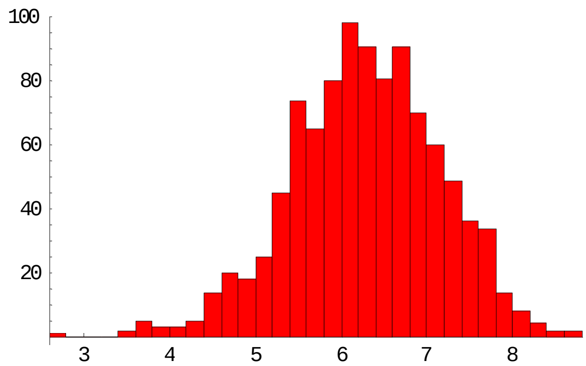

Fig. 3 Histogram of waist-to-hip (Unit: inch)

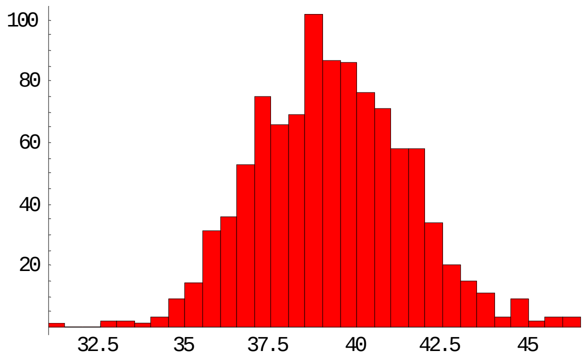

Fig. 4 Histogram of length (Unit: inch)

2.3 Choice of sizes and optimization criteria

In general practice, the number of sizes (*NS*) can vary from 4 to 8. In this article, five cases, from *NS* = 4 to 8 were investigated. Since the distribution of the sizes can affect the production cost, the optimization criteria of the sizes can be complicated as well. One choice is to optimize the number of sizes with respect to the size distribution. Another choice is with respect to the sales forecasts. Finally, the last choice is to maximize the cover factor, which means that most people in the population can find their own appropriate size. Different criteria can serve different purposes. In the current study, the last choice was adopted. Since the number of sizes must be an integer, it is quite trivial to test on different values of number of sizes and then select the one with a maximal cover factor.

Even under such choice of optimization of criteria, the size table may still be either linear or nonlinear, depending on how the parameters are set up. In the case of a linear size table, only the number of sizes, the measurements of the smallest size, and the measurement increments are needed. However, in the case of a nonlinear size table, each individual measurement increment must be known as well. The computation complexity is much higher.

3. Problem formulation

A record of the body measurements of subject j is represented as M_j. The i^{th} measurements is $M_j[i]$. Hence, the sizeUSA database with m subjects and n body measurements is $\{M_j[i], j\ 1..m, i\ 1..n\}$. The size table S is a two-dimensional matrix with p sizes and q primary body measurements. The dimension of S is p by q, where q is less than n. When a subject can be classified by a size s_j, it means that each of the primary body measurements of j fall within the primary body measurement range designated under the same size s_j. Such relationship is represented by a boolean function $F[j, s_j]$ -> {True, False}, indicating whether subject j has a corresponding size s_j. The number of subjects that has a corresponding size is the cover factor of the size table S. Therefore, an ideal size table S must achieve the highest cover factor.

Maximize[Dimension[$F[j, s_j]$ -> True],
where S is to be determined] (1)

Such optimization goal can be decomposed into the following subgoals.

/* spacing matrix between sizes */
$D_g[k] = S[g, k] - S[g+1, k]$ (2)

/* the dimensional ranges matrix */
$e_+[S[g, k]] = (S[g+1, k] - S[g, k])/2$ (3)

$e_-[S[g, k]] = (S[g, k] - S[g-1, k])/2$ (4)

/* measure the deviation of each body measurements */
$$\mathrm{Dif}\,[j, s_j] = \sum_k |\, M_j[k] - S[s_j, k]\,| \quad (5)$$

/* test if the subject fall in the size s_j */
$F[j, s_j] = (\mathrm{Dif}[j, s_j] < err_j)$ (6)

Optimize Dif $[j, s_j]$ by varying $s[s_j, k]$ (7)

where e is the classification range of $s_j[k]$, and err is the tolerance matrix. Finally, the size chart is tabulated as $S[g, k]$, where g is the size ID. Manually construction of size table requires constant spacing between sizes, D_g for all g. The equal spacing implies the dimensions range is also constant. These are linear size tables. Typically, the cover factor of a linear size table is not too high.

4. Algorithm

Both the Cluster Analysis method (statistical) and the GA method (artificial intelligent) will be described in this section.

4.1 Cluster Analysis (Statistical Method)

Cluster Analysis is used whenever one needs to classify the data into smaller clusters of similar properties. In the current study, the data was clustered according to each primary body measurements. Each data has a cluster ID, C_j. In turn, the mean value (average) of the body measurements of each cluster was calculated and used in the size table. Since the statistical package can determine it without any intervention of the user, the resulting size table is nonlinear. One important property of such a size table is the monotonic increasing values of the measurement increment. That means each bigger size is larger than the smaller in all measurements, regardless of the body shape.

The associated evaluation criteria of the cover factor is thus a straight forward comparison of body measurements, one by one and identifying the appropriate size. Hence, there are cases where a subject is having some body measurements within the size s, and some in size s-1 or/and size s+1. As long as there is no jump in the size assignment of any body measurements, the subject is considered as the size s with most body measurement satisfying the size specification.

The program is written in Mathematica™ V5.2 using the built-in function FindClusters[] which is available in the package of Statistics`ClusterAnalysis`. All the technical details can be found at the web site of www.wolfram.com.

4.2 Genetic Algorithm (Artificial Method)

The optimization algorithm is based on genetic algorithm. Since the Mathematica™ comes with a standard optimization function NMinimize[], which has a Genetic Algorithm implementation using the parameter of Method→"DifferentialEvolution". Therefore, if the problem can be formatted to match the syntax of the function NMinimize[], all calculation can be taken cared of by Mathematica™. The number of iteration was set to be automatic. The target accuracy was set to 0.0001. The execution of the program was on an Intel Duo Xeon 2.66G Hz personal computer with 2 Gbyte of RAM. The translation of the existing problem to the genetic algorithm is as follow:

Table 2. Mapping of existing problem to GA format to find the clusters

Existing Problem	Modification	Genetic Algorithm
$M = \{ M_j[i] \}$	$\{ C_j, M_j \}$	Chromosome
	VRC	Fitness function
$C = \{C_j\}$	C_j	Optimization parameter

The measurement array M is of size m by n. Each subject data has a cluster ID, Cj, which forms the chromosome. Since, a fitness criteria of a cluster is the Variance Ratio Criterion (VRC) [11], the fitness function is VRC, which is defined as Equation 8.

$$\mathrm{VRC} = (\mathrm{TRACE}(B) / (NS - 1)) / (\mathrm{TRACE}\ (W) / (m - NS)) \qquad (8)$$

where TRACE is the matrix trace, B is the between-cluster sums of square matrix and W is the pooled within-cluster sums of square matrix, with NS to be the number of clusters desired. The n-dimensional Euclidean distance is used for each pair of subjects.

Once the clusters have been found, the size table can be constructed. The representative measurements of each size can be found by minimizing the distance between the representative measurements with all the data within the same cluster.

Table 3. Mapping of existing problem to GA format to calculate the representative element

Existing Problem	Genetic Algorithm
S	Chromosome
Dif $[j, s_j]$	Fitness function
$S[g, k]$	Optimization parameter

Evaluation of this size table is different from the previous section, because in this formulation, there is no guarantee that the size table is monotonic. That means the final size table is optimized with respect to the subjects as a group of clusters because all measurements are considered at the same time. Each size cluster represents a group of subjects of similar body shape.

Furthermore, since the penalty function is in the form of sum of deviation of the members with respect to the representing member, namely the center of the cluster, the assignment of appropriate size must also be constrained by a upper bound, *err*, on the deviation allowed. For current project, the aggregate deviation bound, *err*, is defined in Equation 9. If more complicated measure is desired, *err* can be a matrix too.

$$err = q \qquad (9)$$

5. Performance evaluation

Although the full data set contains 5544 subjects, the first 1000 subjects were used for the comparison in this article, because of the computation cost is much lower. The results are tabulated in Table 4 to 8. Each table contains the information of the optimized size table using Clustering method and the GA method. The cover factor is main comparison index for the performance of the size table. The distribution of the cover factor of each size is also displayed for comparison. It should be emphasized again that although the ideal size table can have a cover factor percentage of 100%, in general, it is not achievable when real data are being used. Also, the 1000 USA subjects do have a very large variation in terms of body shape. Hence, the cover factor is not very high.

Table 4. Comparison of optimized size table (NS = 4)

Cluster Analsys					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
S	29.89	37.63	4.88	36.30	12.58%
M	34.21	41.89	5.83	38.46	25.87%
L	39.48	46.54	6.58	40.32	22.37%
XL	46.62	53.36	7.44	42.62	7.08%
				Cover Factor	67.90%
Genetic Algorithm					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
1	30.14	37.64	6.41	38.25	25.00%
2	33.92	41.52	6.27	39.66	26.30%
3	38.98	45.24	6.14	39.47	17.00%
4	45.97	52.08	6.28	40.45	5.20%
				Cover Factor	73.50%

Table 5. Comparison of optimized size table (NS = 5)

Cluster Analsys					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
XS	28.71	36.54	4.67	35.84	5.15%
S	32.16	39.69	5.65	37.72	15.68%
M	35.81	43.23	6.29	39.38	19.78%
L	40.21	47.85	6.86	41.21	12.13%
XL	46.88	54.49	7.58	43.55	2.95%
				Cover Factor	55.70%
Genetic Algorithm					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
1	30.18	37.53	6..39	38.06	23.30%
2	32.69	41.20	6.29	40.84	19.20%
3	36.40	42.71	6.21	38.68	20.20%
4	40.81	47.05	6.20	39.80	10.60%
5	46.93	53.17	6.24	40.35	4.00%
				Cover Factor	77.30%

Table 6. Comparison of optimized size table (NS = 6)

Cluster Analsys					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
XS	28.31	36.24	4.52	35.33	2.40%
S	31.57	39.28	5.48	37.11	8.11%
M	34.76	42.53	6.01	38.65	12.48%
L	38.58	46.13	6.39	40.03	9.78%
XL	42.92	50.80	6.86	41.66	5.57%
XXL	48.63	57.80	7.58	44.09	1.15%
				Cover Factor	39.50%
Genetic Algorithm					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
1	29.14	36.86	6.40	37.77	16.40%
2	31.65	39.77	6.27	40.46	19.50%
3	34.42	40.18	6.33	37.10	13.50%
4	35.22	43.40	6.33	40.71	16.00%
5	39.76	45.59	6.11	39.10	13.90%
6	46.06	52.29	6.27	40.50	5.20%
				Cover Factor	84.50%

Table 7. Comparison of optimized size table (NS = 7)

Cluster Analsys					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
XXS	28.16	35.61	4.52	35.20	1.57%
XS	30.98	38.40	5.48	36.87	4.62%
S	33.16	40.81	6.01	38.15	6.02%
M	35.70	43.17	6.30	39.25	6.10%
L	38.96	46.33	6.58	40.46	4.98%
XL	43.14	50.92	6.93	41.84	3.18%
XXL	48.79	57.80	7.58	44.11	0.93%
				Cover Factor	27.40%
Genetic Algorithm					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
1	28.61	35.98	6.41	37.02	9.80%
2	30.79	38.68	6.38	39.40	21.00%
3	33.40	42.17	6.28	41.09	13.90%
4	34.85	40.51	6.29	37.33	13.50%
5	37.78	44.51	6.13	40.02	16.10%
6	42.09	47.79	6.26	39.17	7.80%
7	47.12	53.59	6.19	40.53	3.50%
				Cover Factor	85.60%

Table 8. Comparison of optimized size table (NS = 8)

Cluster Analsys					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
XXS	28.16	35.61	4.52	34.74	1.18%
XS	30.95	38.19	5.48	36.37	2.83%
S	33.02	40.22	6.01	37.56	3.90%
M	35.36	42.41	6.29	38.75	3.15%
L	38.13	44.78	6.30	39.78	2.85%
XL	41.13	47.76	6.58	40.88	2.68%
XXL	45.32	51.57	6.93	42.11	2.00%
XXXL	50.87	57.93	7.58	44.36	0.50%
				Cover Factor	17.20%
Genetic Algorithm					
Size	Waist	Hip	Waist-to-Hip	Length	Cover Percentage
1	28.49	36.17	6.37	37.44	10.30%
2	30.71	38.58	6.37	39.47	16.10%
3	32.43	41.09	6.10	41.41	10.80%
4	33.54	39.87	6.39	37.30	13.40%
5	35.46	43.65	6.32	40.61	12.90%
6	38.32	42.88	6.19	38.03	10.3%
7	40.97	47.39	6.18	39.82	9.30%
8	46.93	53.17	6.24	40.35	4.0%
				Cover Factor	87.10%

The performance of the size table generated by the GA method (referred as GA size table) is far more superior than that of the size table generated by the Cluster method (referred as Cluster size table). Firstly, the cover factor of the GA size table is higher than that of the Cluster size table. Such result is logical because Cluster size table assumes the subjects to behave as if there is a monotonic increasing body framework, which is not true in the reality.

Secondly, as the number of sizes increases, the cover factor of the GA size table can still maintain a very high and stable level, whereas that of the Cluster size table tends to drop very fast. The drop in cover factor of the Cluster size table can be explained because finer size partition implies the exclusion of the subjects whose body shape deviate from the size standard.

Thirdly, as the number of sizes increases, the cover factor of the GA size table also increases. When more sizes are used, the subjects can be classified more finely, because the classification is based on the overall body shape of the subjects, rather than individual measurements. Such increases matches with the expectation that more sizes should flavor better fitting of the size table to the population.

6. Conclusion

A good size table should cover as many people in the population as possible. Each customer can then find his/her own size easily. Loss of sales due to inappropriate size can be reduced. Traditionally there are many different ways to generate a good size table. In this article, an artificial intelligent method, namely the Genetic Algorithm, is adopted. The performance of this method is compared to a statistical method, namely the Cluster Analysis based on individual measurements, which implies monotonic increasing body frame.

Instead of using simulated data, we used real data, which consists of 1000 women's body measurements of the SizeUSA.

The comparison is presented. Based on this set of data, GA method can perform better than the Clustering method in two different ways. Firstly, the GA size table can achieve higher cover factor. Secondly, the GA size table can improve further when the number of sizes increases.

The current study is not without limitation. Firstly, when the number of subject increases, the computational cost increases rapidly. Secondly, the cover factor of the GA size table is measured using a aggregate deviation, beyond a threshold upper bound of the deviation implies unclassified. More sophisticated size assignment algorithm can be used to pin out which body measurements of a subject is out of range. This can be implemented using a deviation matrix. Thirdly, theoretical analysis should also be conducted to derive the properties of the GA size table with respect to changes of different parameters. Fourthly, when the number of data increases, the computational cost increases drastically. So, more powerful computers or parallel computer clusters must be used if real-time performance is required.

As a pilot study, the current study is considered successful and some insight of the sizing system has been discovered. More evidence are needed to prove the application of nonlinear size table in improving the fitting of the garment By then, a robust automatic sizing system generation software can be achieved. Even without a rigorous proof, the industrialists are already using nonlinear size table in capturing their niche market.

Acknowledgements

This project is financially supported by the Hong Kong Polytechnic University Research Grant (A/C PG98). The authors would like to express their gratitude to the funding body.

References

[1] Zernike, K. Sizing up America: Signs of expansion. The New York Times. (March 1, 2004).

[2] [TC]^2 (Textile Clothing Technology Corporation), http://www.tc2.com/news/news_sizedata.html (March 5, 2004).

[3] [TC]^2 (Textile Clothing Technology Corporation), http://www.tc2.com/newsletter/arc/071206.html#two (July 12, 2006).

[4] International Organization for Standardization. Standard sizing systems for clothes (No. ISO/TR 10652:1991).

Geneva, Switzerland: International Organization for Standardization (1991).

[5] Salusso Deonier, C. J., DeLong, M., Martin, F. B., & Krohn, K. R. A multivariate method of classifying body form variation for women's apparel. Clothing and Textiles Research Journal, 4(1), 38-45(1985-1986).

[6] Tryfos, P. An integer programming approach to the apparel sizing system. The Journal of the Operational Research Society, 37(10), 1001-1006 (1986).

[7] Vidal, R. V. V. On the optimal sizing problem. The Journal of the Operational Research Society, 45(6), 714-719 (1994).

[8] McCulloch, C. E., Paal, B., & Ashdown, S. P. An optimization approach to apparel sizing. Journal of the Operational Research Society, 49, 492-499 (1998).

[9] Coppin, B. *Artificial intelligence illuminated*, Sudbury, Mass.: Jones and Bartlett Publishers (2004).

[10] $[TC]^2$ (Textile Clothing Technology Corporation), http://www.tc2.com/ TC2 - SizeUSA - US Anthropometric Survey.mht. (Apr 10, 2007)

[11] Cowgill, M.C., Harvey, R.J., Watson, L.T., A Genetic Algorithm Approach to Cluster Analysis. Computers and Mathematics with Applications, 37, pp99-108 (1999).