Application of Data Mining Techniques for Establishing Sizing System for Army Officers in Sudan

Elawad F Elfaki[1],

Akram H M Ali[2]

Department of Textile Engineering,

College of Engineering and Industries Technology,

Sudan University of Science and Technology, Khartoum, Sudan

Abstract

This paper aimed to use data mining techniques in establishing standard sizing system for Sudanese army officers garment based on an anthropometric body measurements variables. The anthropometric data was collected from Sur military clothing factory in Sudan. The data set collected for 841 army officers was used .For each individual 13 anthropometric variables were involved. In this research data mining methods (WEKA and SPSS) were used for clustering and establishing sizing system. The K-means algorithm was implemented to determine the final cluster classification in two stages in the first stage, the (WEKA) clustering method was applied on the original database to categories the size figures. The (SPSS) method, as quantitative approach, was used in the second stage in order to generate descriptive statistics for the raw data following a step by step procedure. Cluster analysis using chest and waist as a control anthropometric variables revealed eight distinct clusters namely; XS, S, M, L, XL, XXL, XXXL and XXXXL respectively. Size codes and upper and lower size limits are generated.

KEYWORDS: Data mining, anthropometric variables, garments industry, sizing chart, clustering

1. Introduction

According to (P.Lyman et al, 2008) survey the world produces between 1 and 2 Exabyte's of unique information per year, which roughly 250 megabytes for every man, women, and child on earth. Due to the wide range of computer devices nowadays, our ability to produce and collect data has impressively increased. As a result, huge amounts of data are collected from different application domains like business, science, telecommunication and health care systems. Also, the World Wide Web overwhelms us with information. A part from their huge quantity, modern data is also characterized by low level of abstraction, high degree of diversity and complexity. Furthermore, data are not only produced in a centralized but also in a distributed way, which imposes new challenges regarding their management. Due to these reasons, it is impossible for humans to thoroughly investigate these data collections through a manual process. Therefore, the compulsory need for data mining emerged.

Data mining is the use of an appropriate set of technologies to exploit patterns of information from massive customer focused databases. However, data mining is not a single technology. Rather, it is a collection of tools that are used to extract information from data. Data mining is not just a technology but it is also a process. It cannot be fully automated as individuals must be active in the process to be sure that the information that is extracted is accurate.

2. Historical Background:

Data mining is used for a variety of purposes in both the private and public sector in business, science and engineering, and manufacturing etc.; in manufacturing: The first applications of artificial intelligence in engineering in general and in manufacturing in particular were developed in the late 1980s. Kusiak (2006) presented a comprehensive overview of data mining application in manufacturing especially in the areas of production processes, control, maintenance; Customer Relationship Management (CRM), decision support systems (DSS), quality improvement, fault detection, and engineering design,

Garment sizing systems were originally based on those developed by tailors in the late 18th century. Before that all garments were hand- made to order. Tailors measured the body dimensions of each customer, and then drew and cut patterns for each garment; for specific customer after many original patterns had been accumulated the tailors discovered correlations between bodily dimensions, regardless of the individual differences. Tailors gradually developed these patterns into a system of garment storage, which could be used to make clothes for people with similar figures.

Emanuel et al, (1959) established a set of procedures for formulating a standard sizes for figure types. Accordingly, people of all figure types were first classified into one of four body weight groups, which are subdivided by height. Therefore people were divided into eight categories based on similar sizing systems with classification based on two or three sizing variables for male and female as well.

The sizing variables for male garment are height, chest girth and waist girth. The sizing variable most commonly used by female garments are height, bust girth, and hip girth. McCulloch et al, (1998) introduced criteria by which the employed systems could be evaluated and they justified that sizing systems should:

a) Cover the greatest number of people and,
b) Require the fewest number of sizes.

At times these criteria conflict with each other. Depending on circumstances, one may take priority over another. Since the late 1800s, anthropologists used tape measures and calipers which are still being utilized for measuring the human body, (Detong et al, 1993). During that period most clothing was custom – made by tailors. Various measuring methods were developed by professional dressmakers and craftsmen. Their techniques for measuring and fitting their clients were unique. Body measurements, were divided into four groups these are: stature, segment length, body breadth, and circumference. Beazley (1996) suggested a procedure by undertaking a size survey using International Organization for Standardization (ISO) 8559(E) (1989) which include a natural sequence of body measurement comprising three types of data: horizontal, vertical, and others. Beazley (1997) reported that; in Japanese Body Size Data (1992-1994), the definition, equipment,

methods and procedures of body measurements were discussed. In the 1920s; the demand for the mass production of garments created the need for a standard sizing system.

3. Material and Methods

3.1 Material:

The based dataset collected has total of (841) records. These records have existing measurements authorized by an anthropometric expert using a measuring tape. The measurements for the anthropometric variables were in inches with decimals. Therefore, they were converted into integer centimeters in order to ease the comparison with the commonly used international garment sizing standard units. The measurements of the anthropometric variables were taken from army officers with age ranging from 16 to 60 years. Based on the experience and the advice of the experts the selected variables for the jacket were collar, chest, waist, length, across shoulder, sleeve and sleeve and cuff. For the trouser variables were; waist, hip circumference, thigh circumference, knee girth, foot, and trouser length.

3.2 Methods:

One of the difficulties when deciding to choose a data mining system is to determine which method (technique) is appropriate to establish a sizing system for clothing. This is because there are many methods (techniques) that could be used. Some of the criteria that are important in determining the methods (techniques) to be used are determined by trial and error; depending on the type and objectives of the research and the methods or techniques available. Bearing in mind the strengths and weaknesses of each method the chosen packages were:

3.2.1 WEKA 3.6.9:

In this work Elawad (2015) stated that the first method used was Waikato Environment of Knowledge (WEKA). WEKA is software written in Java and runs on almost any plat form. It is a data mining system developed at the University of Waikato in New Zealand. The WEKA work bench contains collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality. The following points support the reasons why WEKA technique was selected;

1) Portability since it is fully implement in the Java programming language and thus runs on almost any modern computing platform.
2) A comprehensive collection of data preprocessing and modeling techniques.
3) It is easily useable by people who are not data mining specialists, and also due its graphical user interfaces.
4) It has kept up-to-date, with new algorithms being added.
5) It provides many different algorithms for data mining and machine learning.

These are the key features responsible for WEKA, s success.

Sunita et al, (2011) added that all of WEKA's techniques are predicted on the assumption that the data is available as a single flat file or relation, where each data points described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute are also used).WEKA is not capable of multi-relational data mining, but there is separate software for

converting a collection of linked database tables into a single table that is suitable for processing using WEKA.

i. Launching WEKA Explorer:

WEKA can be launched from c:\ program files directly, from the desktop selecting WEKA 3.4, shortcut 2 KB icon or from the windows task bar "start" → programs → WEKA 3.4. When WEKA GUI chooser, window appears on screen, one of the four options button of the window can be selected, R. Kirkby, (2002). WEKA 3.6.9 could be launched by following the same steps mentioned for WEKA 3.4. When WEKA GUI chooser' windows appears on screen and one of the following steps could be selected. The steps one; (preprocessed, classify, cluster, associate, select attribute, and visualization) as shown in Figure1.

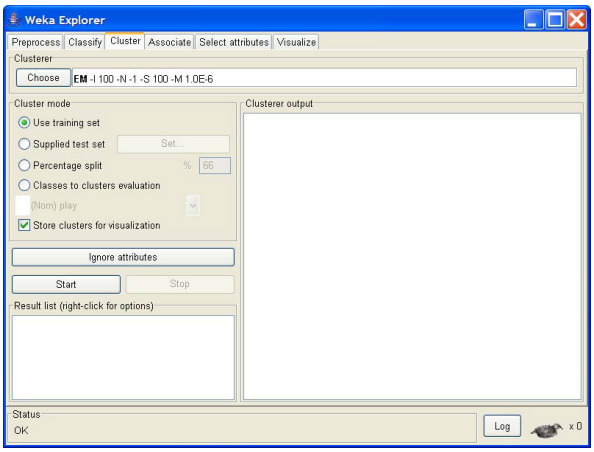

Figure1: Using 'Cluster' tab at the top of WEKA Explorer window

ii. WEKA Cluster Analysis:

Clustering is the portioning of a dataset into subset (clusters), so that the data in each subset (ideally) share some common trait. In the clothing cluster which is typically grouped by the similarity of its member's body shape can be considered as a size category or a figure type. Simple K-means algorithm was used as the clustering approach. This is a prototype–based clustering technique that attempts to find a user–specified number of clusters (Pang, 2006). "K is a certain number of clusters fixed as a priori K centroids are defined for each cluster. Each sample is then assigned to the closest centroid, and each collection of samples that assigned to a centroid is a cluster".

3.2.2 The Statistical Package for Social Sciences (SPSS) Version 18.0

This package was the second method selected because it is a simple clustering method and shows optimal results. However all variables must be independent and normally distributed. Clusters were identified according to the different body types. It was employed for anthropometric data analysis. It reduces the large samples in same groups contains similar number. Finally the WEKA and the (SPSS) methods could be used to establish new sizing systems.

4. Results and Discussion

The clustering model shows the centroid of each cluster and statistics on the number and percentage of instances assigned to different clusters. Cluster centroids are the mean vectors for each cluster; therefore, each dimension value and the centroid represent the mean value for that dimension in the cluster. The results of the cluster samples were analyzed statistically using (SPSS) version 18.0 package.

4.1Size Interval in a Size Chart:

Before the eleven figures types were classified it was necessary to select the size interval in order to get some flexibility for easy allowances during the seaming process. Kunick (1984) stated that, size interval is the division of sizes in a size chart. The BSEN 13402–3,(2004) recommended that an interval of 4 cm or 6 cm for both bust and waist1 and 4 cm or 5 cm for hip in order to have a flexible link between the bust, waist and hip. Beazley (1998) used 4 cm interval for the key dimensions (bust, waist and hip) for size 8 – 14 and 6 cm interval for size 16 to normalize the intervals. Many British companies use 5 cm interval between all sizes, Aldrich, (2008). Kunick, (1984) proposed that 6 cm size interval is used by most countries. According to Winks, (1997), most countries applied 4 cm as the interval of chest girth. In this research, the intervals for the key anthropometric dimensions, chest, waist and hip were 4 cm. As mentioned before, from the total data (841), only (28) samples were excluded (outliers). Therefore, the percentage covered by the proposed sizing systems for the chest and waist was 97%.

4.2 Establishment and Evaluation of the New Sizing System:

4.2.1 Establishing of the New Sizing System:

It would be a very tedious task using all 13 anthropometric variables to establish sizing systems. The development of the size chart was carried out by using values obtained from the statistical information of body dimensions. Winks (1997) states that, the mean value can be a convenient indication of obtaining central tendency. The mean values are the most widely used value for size steps and it is equivalent to the average size (mean) and also equivalent to size 12 of every size chart, Boakye S. et al, (2012). Eleven size steps approach was used to establish the new size chart. The eleven size steps used as a base for the determination of the outliers. The values that were less than the smallest size and those higher than the biggest size were eliminated and classified as outliers. In order to obtain eleven steps for eleven categories of body size, (1STD), (2STD), (3STD), (4STD) and (5STD) values were added to the mean and subtracted from the mean respectively. This was carried out in order to obtain five values that are higher and five values that are lower than the mean. According to Ashdown (1998) by subtracting (-1STD), (-2STD), (-3STD), (-4STD) and (-5STD) from the mean, the values obtained represents size 2, 4, 6, 8 and 10 respectively. When (1STD), (2STD), (3STD), (4STD) and (5STD) are added to the mean, the values obtained represent sizes 14, 16, 18, 20 and 22 respectively. The mean and standard deviation values were all rounded up to 0.1 decimals. Values above 0.15 were rounded up to 0.2 cm, and values below 0.15 have been reduced to 0.1 cm. This was done to make the comparison between SUR, EUR, and US size charts with the new established size charts more easy and understandable. To get the new established sizing charts from the eleven steps, the values of the 5XL figure size were omitted based on the results of the outlier from cluster 8, where there were no classes

represented in this figure size. Therefore, the new established sizing systems chart which consists of 8 figure size is given in Table1.

4.2.2 Evaluation of the New Established Sizing Systems:

As stated by the International Standards Organization (ISO), and those others Organizations mentioned in the literature; "the use of control dimensions and size interval can effectively facilitate to recognize the parameters for developing sizing systems" Winks, (1997). After the eleven figure sizes were classified by the WEKA and SPSS software's, the new established size system of the eight figures size were determined and the results are given in Table 1. Figure2: shows the relevant scatter plots of chest on the X-axis verse the waist on the Y-axis and the interval was 4 cm to demonstrate the distribution of all figures type. It has been reported that, Cooklin, (1992) the chest is the most important anthropometric variable in establishing sizing systems in the field of garment making. The waist is also an important variable for sizing male garments in many countries.

Table1: the Proposed New Established Size System

	XS M-2STD	S M-1STD	M Mean	L Mean+1 STD	XL M+2STD	XXL M+3STD	XXXL M+4STD	XXXXL M+5STD
Cuff	17	18.3	19.6	20.9	22.2	23.5	24.8	26.1
Sleeve	56.6	60.1	63.6	67.1	70.6	74.1	77.6	81.1
Shoulder	40.6	43.6	46.6	49.6	52.6	55.6	58.6	61.6
Length	68.6	73.4	78.2	83	87.8	92.6	97.4	102.2
Waist	73.5	85.1	96.7	108.3	119.9	131.5	143.1	154.7
Chest	92	100.5	109	117.5	126	134.5	143	151.5
Collar	36.1	39	41.9	44.8	47.7	50.6	53.5	64.9
Length 1	97.1	101.7	106.3	110.9	115.5	120.1	125	129.3
Foot	40	42.9	45.8	48.7	51.6	54.5	57.4	60.3
Knee	46.1	50	53.9	57.8	61.7	65.6	69.5	73.4
Thigh	63.8	69.5	75.2	80.9	86.6	92.2	97.8	103.7
Hip	93.2	99.2	105.2	111.2	117.2	123.2	129.2	135.2
Waist 1	72.6	83.3	94	104.7	115.4	126.1	136.8	147.5

n = 813 all values are in centimeters ___ jacket ___ trouser

Figure3: illustrates the differences between the eight types for the new established sizing systems. The figure was plotted as a line graph to yield a better insight into the differences between the new established sizing systems. The eight figures types are exhibited by clear differences in chest and waist. The eight figure types also follow the order, SX, S, M, L, XL, XXL, XXXL and XXXXL.

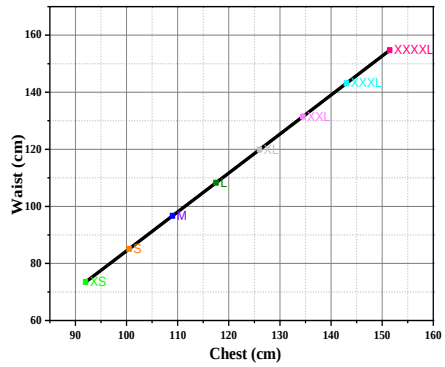

Figure2: Scatter plot of chest verse waist for the proposed new established size system

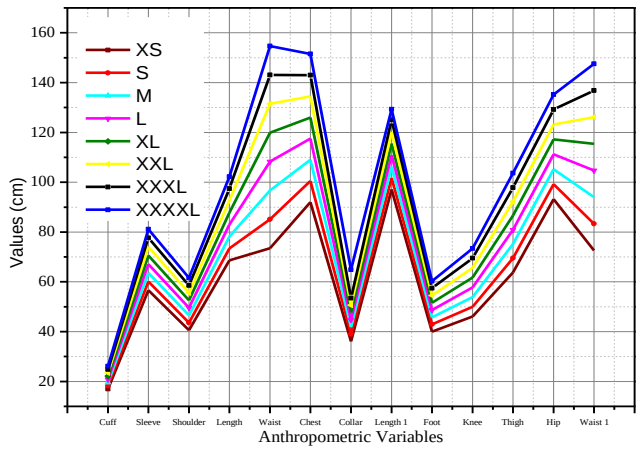

Figure 3: The distinct anthropometric variables between the proposed new established size systems

4.3 The Size Code:

The size codes are based on the numerical coding methods. In this work the size codes were determined after generating the eight size steps values from the figure sizes. The new size code was termed (SUD) abbreviated for Sudan, and then the code was added to establish the 8 size figures for Sudanese army officers (poshirt uniform). Therefore, the new size codes are, SUD XS, SUD S, SUD M, SUD L, SUD XL, SUD XXL, SUD XXXL and SUD XXXXL.

4.4 Lower and Upper Limits of Sizes

In order to establish the limits of each size and demonstrate the extent of coverage for inter size ranges and one of the important step to determine the lower and upper limits of sizes. The value obtained for each size code is used as a midpoint and the lower and upper limit are determined from it. The half value of the standard deviation of each body dimension added or subtracted to the midpoint value. A value of 0.01 is subtracted from the figure obtained below the midpoint to create limits between the lower value of the next size and the upper value of the previous one. To avoid overlapping of figures with the next size value of 0.01 is subtracted from the upper limit making it less than the next value. Beazley, (1998) and other researchers used this procedure. Table 3 tabulates and presents the lower and upper limits of the body dimensions.

Table2: Lower and Upper Limit of Size Code

	SUDXSM-2STD	SUDS Mean-1STD	SUDM Mean	SUDL Mean+1 STD	SUD XL M+ 2STD	SUD XXL M+ 3STD	SUD XXXL M+ 4STD	SUD XXXXL M+ 5STD
Cuff	16.35	17.65	18.95	20.25	21.55	22.85	24.15	25.45
	17	18.3	19.6	20.9	22.2	23.5	24.8	26.1
	17.64	18.94	20.24	21.54	22.84	24.14	25.44	26.74
Sleeve	54.85	58.35	61.85	65.35	68.85	72.35	75.85	79.35
	56.6	60.1	63.6	67.1	70.6	74.1	77.6	81.1
	58.34	61.84	65.34	68.84	72.34	75.84	79.34	82.40
Shoulder	39.10	42.10	45.10	48.10	51.10	54.10	57.10	61.10
	40.6	43.6	46.6	49.6	52.6	55.6	58.6	61.6
	42.09	45.09	48.09	51.09	54.09	57.09	60.09	63.09
Length	66.20	71.00	75.80	80.60	85.40	91.20	95.00	99.80
	68.6	73.4	78.2	83	87.8	92.6	97.4	102.2
	70.99	75.79	80.59	85.39	91.19	94.99	99.79	104.59
Waist	67.70	79.30	90.90	102.50	114.10	125.70	137.30	148.90
	73.5	85.1	96.7	108.3	119.9	131.5	143.1	154.7
	79.29	90.89	102.49	114.09	125.69	137.29	148.89	160.49
Chest	87.75	96.25	104.75	113.25	121.75	130.25	138.75	147.25
	92	100.5	109	117.5	126	134.5	143	151.5
	96.24	104.74	113.24	121.74	130.24	138.74	147.24	155.74
Collar	34.65	37.55	40.45	43.35	46.25	49.15	52.05	54.95
	36.1	39	41.9	44.8	47.7	50.6	53.5	64.9
	37.54	40.44	43.34	46.24	49.14	52.04	54.94	67.69
Length 1	74.50	99.40	104.00	108.60	113.20	117.80	124.40	127.80
	97.1	101.7	106.3	110.9	115.5	120.1	125	129.3
	99.39	103.99	108.59	113.19	117.79	122.39	127.79	131.59
Foot	40.35	41.45	44.35	47.25	50.15	53.05	55.95	58.85
	40	42.9	45.8	48.7	51.6	54.5	57.4	60.3
	41.44	44.34	47.24	50.14	53.04	55.94	58.84	61.74
Knee	44.15	48.05	51.95	55.85	59.75	63.75	67.55	71.45
	46.1	50	53.9	57.8	61.7	65.6	69.5	73.4
	48.04	51.94	55.84	59.74	63.74	67.54	71.44	75.34
Thigh	60.95	66.65	72.35	78.05	83.75	89.35	95.95	100.65
	63.8	69.5	75.2	80.9	86.6	92.2	97.8	103.7
	66.64	72.34	78.04	83.74	89.44	95.04	100.64	106.54
Hip	90.20	96.20	102.20	108.20	114.20	120.20	126.20	132.20
	93.2	99.2	105.2	111.2	117.2	123.2	129.2	135.2
	96.19	102.19	108.19	114.19	120.19	126.19	132.19	138.19
Waist 1	67.25	77.95	88.65	99.35	110.05	127.75	131.45	142.15
	72.6	83.3	94	104.7	115.4	126.1	136.8	147.5
	77.94	88.64	99.34	110.04	127.74	131.44	142.14	152.84

4.5 Application of the New System for Establishment of Garment Measurements:

The new established size systems need to be checked to find out its accuracy. In order to verify the size chart, garments measurements should be established for the preparation of the patterns and subsequently the garment for trials. For establishing the garment measurement, ease allowance was added to each body dimension on the established size chart. By using measurements information from the new established size system, patterns were constructed manually. The basic blocks constructed were grade in all figure sizes indicated, i.e. SUD XS, SUD S, SUD M, SUD L, SUD XL, SUD 2XL, SUD 3XL and SUD 4XL. As mentioned earlier, the basic block for the base size M was decreased four steps down and increased four steps up to obtain the other sizes.

5. CONCLUSION:

The WEKA and SPSS methods were used for clustering and establishing sizing system by implementing simple K-means algorithm to determine the final cluster classification.

Cluster analysis using chest and waist as a control anthropometric variables reveal eight distinct clusters. Each cluster showed distinct difference between clusters but similar within each cluster. The study also showed that within the army officers aged 16-60 years there were eight types of body shapes namely; XS, S, M, L, XL, XXL, XXXL and XXXXL respectively.

The size codes were based on numerical coding method abbreviated from Sudan (SUD) and added to the sizes obtained. Therefore the size codes obtained are, SUD XS, SUD S, SUD M, SUD L, SUD XL, SUD XXL, SUD XXXL and SUD XXXXL respectively.

The value obtained for each size code is used as the midway point for determining the lower and upper limit which helps in establishing the limit of each size in order to demonstrate the extent of coverage for inter size ranges.

The total applicability of the new established sizing system was 96.7% which is very high coverage. Scatter plots of chest on the X-axis; verses waist on the Y-axis, using an interval of 4 cm demonstrates the distribution of all figure types of the established sizing systems. A line graph was plotted and it shows significant differences in chest and waist among the eight figure types of the new proposed established sizing systems.

This work contributes largely to knowledge of size chart by providing a detailed procedure involved in establishing standard sizing system based on anthropometric variables and will serve as the basis for other future research in garments industry in Sudan.

REFERENCES

(1) P .Lyman and H. R. Varian How Much Information? http://www2.sims.berkeley.edu/ research/ projects/how – much-info/, 2003 (valid as of May 2008)

(2) Kusiak, A., Srinivas, S., Shahbas, M., and Harding, J., A., November 2006, Vol. 128/969. Data Mining in Manufacturing a Review. J. of Manu. Sc. and Engineering.

(3) Emanuel. I., M. Alexander, E. Churchill and B. Truett, 1959. A height-weight sizing system for flight clothing. Wright Air Development Center, WADC Technical Report 56-365, Wright-Patterson Air Force Base, Ohio.

(4) Elawad F. Elfaki (2015) Application of Data Mining Techniques for Establishing Standard Sizing Systems for Army Officers Garments in Sudan, PhD Thesis July 2015 (unpublished).

(5) Kirkby, WEKA Explorer User Guide for Version 3-3-4, University of Waikato, 2002.
(6) Abdu-Boakye S., Power, Jess, Wallace, T. and Chen, Z. (2012). Development of a sizing system for Ghanaian Women for the Production of Ready-to-Wear Clothing. In: The 88th Textile Institute World Conference 2012, 15th – 17th May 2012, Selangor, Malaysia.
(7) Ashdown, P.S. (1998) an investigation of the structure of sizing systems: 'A comparison of three multidimensional optimized sizing system generated from anthropometric data with ASTM standard D55855 – 94'. Journal of Clothing Science and Technology. Vol. 10, No. 5, pp. 324 – 341.
(8) Kunick, P. (1984). Modern sizing and pattern making for women's and children's garments. London: Phillip Kunick Publication.
(9) British Standard Institution (2004). BS En 13402-3. Size designation of clothes. Measurements and Intervals. British Standard Institution.
(10) Beazley, A. (1997) 'Size and Fit: Procedures in undertaking a survey of body measurements'. Journal for Fashion Marketing and Measurement. Vol. 2, No. 1, pp 55 – 85.
(11) Winks, J. M., Clothing Sizes: International Standardization, Redwood, U.K. 1997.
(12) Cooklin, G., Pattern Grading for Men's Clothes, Blackwell, Oxford. 1992.