UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF GRADUATE STUDIES AND ACADEMIC ADVANCEMENT

Faculty of Computer Science and Information Technology

Clustering Technique Applied in Order to Establish a Sizing System for Sudanese Trousers: Using k-means Clustering Algorithm

by

Abubaker Mohammed Ahmed Abdelrahman

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements in PartialFulfillment of the Requirement for the Degree of Master of Science in Computer Science

Supervisor

Dr.Mohamed Elhafiz Mustafa

November 2015

Abstract

There are many international standard garment size systems. All the systems are based on non Sudanese sizes. The aim of this study is to establish a Sudanese garment sizing system and to compare it with international systems.

About one thousand and eight hundred records were collected for this study. The researcher uses clustering technique (K-means algorithm) to build Sudanese trousers sizing system.

Four experiments have been performed to find the best sizing system of Sudanese garments (male trousers). The results show that: the lengths of the Sudanese range from XS to XL which represents (86.75%) while the waists range from XS to XL representing (74%) of the data.

The largest cluster in length found is similar to the international standard L with percentage (29%),and the largest cluster in waist is also similar to L with percentage (24%).

One of the important findings of the study is that the Sudanese waist is generally big when compared with the length (one cluster in our experiments is equal to the international waist standard L and the international length standard M).

المستخلص

هناك العديد من الأنظمة الدولية لقياسات الملابس .وتستند جميع الأنظمة على أحجام غير سودانية .والهدف من هذه الدراسة هو إنشاء نظام الملابس السودانية ومقارنتها مع الأنظمة الدولية.تم جمع حوالي ألف وثمانمائة سجل لهذه الدراسة .استخدم الباحث خوارزمية (K-meansclustering technique)للوصول الي قياسات مناسبة للبنطلون بالنسبة للرجل السوداني.

وقد أجريت أربع تجارب لايجاد أفضل نظام للبنطلون (للرجل السوداني) .وأظهرت النتائج أن :نسبة السودانيين الموجودين في المقاس (XSإلىXL) تمثل نسبة (86.75%) بالنسبة للطول و(74%) بالنسبة للخصر، وأنأكبرنسبةمنالسودانييناالموجودينفي المقاسLبنسبة تساوي (29%)بالنسبة للطول و L بالنسبة للخصر بنسبة تساوي (24.75%).

واحدة من النتائج الهامة لهذه الدراسة هي أن الخصر السوداني كبير بالمقارنة مع الطول(مجموعة واحدة في تجاربنا تساوي Lللخصر بمقارنة مع المعايير الدولية وللطولM بمقارنة مع المعايير الدولية).

1.1 Introduction

The information industry has a very large amount of data. Till data is not converted into useful information it is of no use. It is necessary to analyze this huge amount of data and extract useful information from it. The extraction ofinformation is followed by several other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. After all these processes are completed, we are now in a position touse this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration etc .

1.2 Clustering

Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar themselves and dissimilar compared to objects of other groups .

1.3 Background of the Problem

Anthropometry is knowledge of body dimension measurement .

In the field of garment manufacturing, the planning and control of production and inventory are rather complicated procedures. This is the reason that establishing standard sizing systems is necessary and important for garment manufacturers in Sudan.

In garment production engineering, sizing system plays an important role for manufacturing of clothing. The standards for defining the size label are a critical issue.

Clustering technique applied on the original dataset, to categorize the size labels. Then these clusters are used for supervised learning in the second stage for classification. A sizing system classifies a specific population into homogeneous subgroups based on some key body dimensions .

1.4 Thesis Problem

When we choose our clothes, we face an appropriate size problem.

We always hear about specific English, French, German or American size clothes Standards and others, even the size label fixed on clothes, which we find on it specific numbers like 40, 12 or 6, and all of it are European standards, American sizes and others, which suitable with the population of these countries whose bodies size different from Sudanese bodies.

For Sudan, it depend on importing clothes from abroad, thus the exporter may provide us with return items of different sizes not required in the market because not designed for us. So it was necessary to put suitable measurements of Sudanese bodies, to be considered as an ideal model or standard for any new product gauge.

The expression of a particular size of ready-made-clothes differs from country to another, hence this research aims to find out legalize average standard of Sudanese's body size, and then to put a guiding table for Sudanese standards.

1.5 Research Objective

The intent in of this research is to use data mining technique (clustering) to assess the suitability of the standard sizing system to Sudanese (men trousers).

1.6 Methodology

Using k-means clustering algorithm the final result of the research will be reached by analyzing (Clustering) collected data following the steps below:

Collecting Data(exclusively men).

1- Collecting Data (exclusively men)[5].

2- Data Cleansing[5].

(i)Missing data handling.

(ii)Removing or estimating missing values in the data.

(iii)Database balancing.

(iv)Correcting imbalances in the target field.

(v)Others as appropriate.

3-Data Preprocessing[5].

(i)Data Entry.

(ii)Converting data from type to other.

4- Data Analyzing (Clustering): clustering Sudanese male sizes data

5- Interpretation and analysis: the clusters found in previous step will be checked with integration standards to find suitable standard for Sudanese.

1.7 Structure of Thesis

This thesis is divided to five chapters first one is introduction it contain introduction to data mining, and problem statement, and objective of this research. Second chapter is a literature review about data mining, and used algorithm in this thesis.Third chapter is a literature review about garment,and used program in this thesis, and then describes data set information. Chapter four is a methodology used in this research, and an implementation .and last chapter is chapter five which is a discussion about the results and conclusion.