UNIVERSITY OF SCIENCE ANDTECHNOLOGY

COLLEGE OF GRADUATE STUDIES AND ACADEMIC ADVANCEMENT

Faculty of Computer Science and Information Technology

Towards Finding Standard Clothes Measurements for Sudanese Men Shirts Using Expectation Maximization Algorithm

by

Ahmed Ali Adam Mohammed

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.Mohammed Elhafiz Mustafa Musa

November 2015

Abstract

Clothing industry systems depends entirely on measuring human bodies; which is a thorough survey of human body specifying real dimensions and which is based on the garment industry process.

This research studies the international clothing standards and their congeniality to the Sudanese. To find the gap between international standards and Sudanese bodies.

The research presents (4) experiments to find the real number of groups (clusters) of sizes using a dataset collected for this purpose, moreover, two additional experiments have held to determine the effect of different attributes on the final result.

Through experiments, the researcher finds that biggest cluster is similar to the standard (XL), and the smaller cluster is similar to (3XL), while (17%) from Sudanese have no similar standard sizes. Moreover, there are no clusters for the sizes (4XL) and (5XL).

In order to know the most effective attribute two experiments (5) and (6) were performed; these two experiments showed that the attribute (chest) is the most effective while the attribute (length) has the miner effect.

المستخلص

أنظمة صناعة الملابس تعتمد إعتماداً كلياً علي نظام قياس الأجسام والذي يقوم بعمل مسح دقيق لجسم الإنسان محدداً أبعاده الحقيقية والتي بناءاً عليها تتم عملية صناعة الملابس.

في هذا البحث تم إجراء دراسة لمقاييس الملابس العالمية ومدي ملاءمتها للسودانيين في الطريق من أجل تصميم نظام ملابس سوداني يعتمد علي الدراسات الحقيقية لمقاسات أجسام السودانيين ويغطي حاجة أغلبية الشعب السوداني من الملابس، ومحاولة إيجاد الفجوة الموجودة بين المقاييس العالمية وأجسام السودانيين.

تم إجراء عدد (4) اربعة تجارب لتلمُس العدد المناسب لمجموعات الأجسام حسب البيانات المأخوذة من عدة مصادر، كما أن هنالك تجربتين إضافيتين لدراسة تأثير الخصائص الأساسية علي بعضها البعض.

تم جمع بيانات تمثل 1827 شخص وإجراء عملية التحليل بإستخدام خوارزمية Expectation Maximization (EM).

تم التوصل إلي أن أكبر مجموعة من البيانات المأخوذة تقابل المقاس (XL) بينما أقل مجموعةمنها تقابل المقاس (3XL)، كما أن هناك نسبة (17%) لا يجدون مقاساً مقابلاً لهم في نظام المقاييس العالمية الأوروبي.

إتضح من خلال التجارب (5) و(6) أن العلاقة بين الخصائص العامة للقميص علاقة ليست قوية بحيث أن حذف خاصية الطول بالنسبة للقميص لم تؤثر كثيراً علي النتائج، بينما أثر حذف خاصية الصدر بالنسبة للقميص تأثيراً أكبر علي النتائج .

1.1 Introduction

There is huge amount of data available in Information Industry. This data is no use until converted into useful information [1]. Analyzing this huge amount of data and extracting useful information from it is necessary [1].

The extraction of information is not the only process we need to perform; it also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation [1]. The purpose of data analysis is to discover previously unknown data characteristics, relationships, dependencies, or trends

Once all these processes are over, we are now at position to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration etc .

1.2 What is Data Mining?

Data Mining is defined as extracting the information from the huge set of data. In other words we can say that data mining is mining the knowledge from data [1].

1.3 Literature Review

Anthropometry is knowledge of body dimension measurement [2].

In the field of garment manufacturing, the planning and control of production and inventory are rather complicated procedures [2]. 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 .

1.4 Thesis Problem

There are many international standards (European, Asian …etc) for male clothes. However, there is no standard for Sudanese male clothes. This thesis tries to contribute towards finding Sudanese male clothes standard.

1.5 Methodology

Using Expectation Maximization (EM) clustering algorithm the final result of the research will be reached by analyzing (Clustering) collected data following the steps below:

1- Collecting Data (exclusively men shirts) [4].

2- Data Cleansing [4].

a. Missing data handling.

b. removing or estimating missing values in the data.

c. Database balancing.

d. correcting imbalances in the target field.

e. Others as appropriate.

3-Data Preprocessing [4].

a. Data Entry.

b. 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.

1.6 Research Objective

Firstly this thesis aims to deep studying on clustering techniques generally and Expectation Maximization (EM) specially, the second objective of the thesis is to apply the EM algorithm on the problem of finding Standard Clothes Measurements for Sudanese.

1.7 Expected Results

At the last of this thesis one of two expected result will appear; the first one is that the European standard clothes measurements are accurate to Sudanese, hence the result will be promoted by apply clustering techniques to verify that.

The second expected result is that the European standard clothes measurements aren't suitable to Sudanese; therefore the researcher must try to find the hole of fitting in the standard clothes measurements and make it more appropriate to Sudanese by figure out the suitable alteration to European standard.

1.8 Thesis Structure

This thesis contains five chapters: the first Chapter is introduction. While the second Chapter is Literature review chapter, which defines data mining, clustering analysis techniques, and clustering challenges ended by EM algorithm as a tool for clustering in this research. and the third Chapter is Material and Methods chapter, discusses Garment and sizing system as a main field of study, followed by review of the used software in this thesis for accomplish clustering (WEKA program), ended by description of used Dataset. The fourth Chapter is Results and Discussion chapter, gives detailed information about used hardware and software and. At last of chapter there are six experiments followed by results discussion. While the fifth Chapter is the last chapter, defines conclusion, recommendations, and future work.