BIOTECHNOLOGY & BIOTECHNOLOGICAL EQUIPMENT, 2018
https://doi.org/10.1080/13102818.2017.1423515

Taylor & Francis
Taylor & Francis Group

ARTICLE; BIOINFORMATICS

OPEN ACCESS Check for updates

Establishment of a PCA model for skin health evaluation

YiFan He[a], RuiZhen Wang[a], Hong Meng[a], Li Li[a], Zhemin Wu[b] and YinMao Dong[a]

[a]Department of Cosmetic Science, School of Sciences, Beijing Technology and Business University, Beijing, PR China; [b]Bard Medical Device (Beijing) Co, Ltd., Beijing, PR China

ABSTRACT

The aim of this study is to establish an objective skin evaluation system through testing skin data and scoring conducted by general dermatology experts. First, 18 data points including left cheek face water evaporation, and left cheek oil content are collected. Second, comprehensive skin information is obtained from volunteers by the mathematical model of principal component analysis (PCA), following scoring the status of the volunteers' skin by dermatology experts. Then, 99 volunteers are randomly divided into two groups, namely model information and test information. Specifically, skin model information is obtained from 80 volunteers, and test information is obtained from 19 volunteers. PCA is used to extract 18 independent variables of skin test data by means of R Studio software, followed by principal component regression by expert scoring and the extracted principal components. Finally, based on the expert score and model score, skin grading classification can be conducted by MATLAB software, in order to intuitively and effectively evaluate skin status.

ARTICLE HISTORY
Received 7 March 2017
Accepted 30 December 2017

KEYWORDS
Skin evaluation system; comprehensive skin information; principal component analysis

Introduction

A variety of analytical methods are used to describe qualitative characteristics of the skin creating a large amount of heterogeneous data. Statistically analyzing these data through classical methods gives important information on a number of different variables [1]. Traditional statistical analysis does not provide global knowledge regarding the relationships among the different variables, and does allow the grouping of samples with homogeneous characteristics [2]. Therefore, it is useful to synthesize the trend of a particular phenomenon through a few elements. A common and legitimate complaint directed at the multivariable control literature is that while the theory appears to be strong, it is not accompanied by strong numerical tools [3]. To meet this need, we can make use of multivariate statistical methods, including principal component analysis (PCA), which makes it possible to identify the most important directions of variability in a multivariate data matrix [4]. PCA is often the method of choice for reducing a large set of correlated variables to a smaller number of uncorrelated components. PCA is a statistical method that has been used in representative features to transform the input space into a new lower dimensional space [5]. It has been widely used to identify and summarize multiple inter-relationships among original variables [6–9]. Intercorrelated variables are combined into a smaller number of new variables called principal components [10]. The first principal component accounts for much of the variability in the data, and each succeeding component accounts for the remaining variability [11]. The uncorrelated variables are linear combinations of the original variables and the last of these variables can be removed with minimum loss of real data in order to identify new meaningful underlying variables [12]. The PCA technique has previously been investigated by researchers [13].

The main purpose of exploratory data analysis is to learn about interrelationships between objects and variables [14]. In large, complex data-sets, visualization is needed to achieve this aim. A well-known method for visualization of the data patterns is PCA [1–9]. In this study we use volunteers to obtain accurate, real-time skin information.

As explained by Nӕs et al. [15], the results for both variables and objects are presented. The abscissa corresponds to the first principal component and the ordinate corresponds to the second principal component. Samples to the right in the score plot have high values for variables to the right in the loading plot. The same holds for samples to the left, at the top, or at the bottom. Objects close together have similar characteristics; variables close together are positively correlated, while variables lying opposite to each other in the

CONTACT YinMao Dong ymdong2008@163.com