Expert Systems with Applications 40 (2013) 4226–4231

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A case study of muscle dysmorphia disorder diagnostics

Marina V. Sokolova [a,b,*], Irene González-Martí [c], Onofre Ricardo Contreras Jordán [d], Juan Gregorio Fernández Bustos [d]

[a] Instituto de Investigación en Informática de Albacete (I3A), Universidad de Castilla-La Mancha, 02071 Albacete, Spain
[b] Southwest State University, Kursk 305040, Russia
[c] University of Castilla-La Mancha, Faculty of Education, Physical Education Department, 16071 Cuenca, Spain
[d] University of Castilla-La Mancha, Faculty of Education, Physical Education Department, 02071 Albacete, Spain

ARTICLE INFO

Keywords:
Muscle dysmorphia
Decision tree
Modeling

ABSTRACT

Muscle dysmorphia is a mental disorder which mainly affects young people and which may have physiological consequences. Though the nature of muscle dysmorphia has been studied, there is still a lack of scientific works explaining the hidden patterns and creation of models for this disorder. With this aim, using regression techniques, rule learners and decision tree classifiers, several models which can be useful for diagnosis and prevention of this disorder, has been obtained.

1. Introduction

Muscle Dysmorphia (MD) is a mental disorder consisting in that a person becomes extremely dissatisfied with his/her musculature and reaches pathological levels by a distortion of her/his body image (Pope, Phillips, & Olivardia, 2000). This distortion is a reflection of the social vision changes in the way the male body image is perceived (Pope et al., 2000; Tylka, 2011). Over the last twenty years, the concept of the male body ideal image has suffered significant changes, as it has shifted away from one with excessive body fat to an increasingly muscular physique/build (Nowell & Ricciardelli, 2008; Pope et al., 2000; Gruber, Pope, Borowiecki, & Cohane, 1998). Individuals who suffer from MD often perceive themselves as small and weak, even when this is not true. Generally speaking, obsessions with one's musculature often leads to compulsive weightlifters (Olivardia, 2001). However, the consequences of MD can be weakened in social and physical domains (Morgan, 2008; Pope et al., 2000). The main trouble when approaching the treatment of this disorder is that MD is not classified as a mental disorder in the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders *DSM – IV – TR* (American Psychiatric Association, 2002). As a consequence of this fact, the authors who investigate in this area classify MD in function of the principal characteristics which present the people affected.

Although MD affects both sexes, many studies have demonstrated that males are more affected than females (Morgan, 2008; Olivardia, 2001; Pope et al., 2000). However, recent research shows contradictory results (González-Martí, 2012). The average age to develop the disease is the range of onset between 15 and 30 years (González-Martí, Fernández, Contreras, & Mayville, 2012; Pope et al., 2000). People affected by MD acquire certain behavioral changes, including muscular dissatisfaction, dependence on bodybuilding, use of supplements or illegal drugs without medical prescriptions, and repetitive muscle checking after workouts, among others.

These behavioral features can be used as indicators for MD diagnosis, because not only do they describe a psychological profile of a person, which can be suffering from this disorder, but also they show the degree of the disease. The main purpose of this study is to discover hidden patterns among the data from different measures administered on bodybuilders, weightlifters, and powerlifters as this group of sport people is more likely to have MD disorder. In addition, we present the results of modeling and comparative analyses based on this data and determine the most important symptoms which could allow correct and efficient diagnosis of MD.

The set of models calculated in this study include regressions (both linear and non-linear), association rules and decision trees. Hence, we propose a new method of MD detection and evaluation. We show that the results obtained after analyzing a large number of real cases indicate that our method can be used by specialists in order to diagnose people with MD disorder.

The article is organized as follows. In Section 2, we overview recent research and studies dedicated to medical and psychological information modeling with data mining algorithms. In Section 3, the proposed method of MD detection is introduced. Next, in Section 4, the experimental data, outcomes of the classification are presented and discussed. Finally, in Section 5 we compare our method with other popular techniques, and present our conclusions.

* Corresponding author at: Instituto de Investigación en Informática de Albacete (I3A), Universidad de Castilla-La Mancha, 02071 Albacete, Spain. Tel.: +34 967599200; fax: +34 967 599224.

E-mail addresses: marina.v.sokolova@gmail.com (M.V. Sokolova), Irene.GMarti@uclm.es (I. González-Martí), Onofre.CJordan@uclm.es (O.R. Contreras Jordán), JuanG.Fernandez@uclm.es (J.G. Fernández Bustos).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.01.023

2. Data mining for medical and psychological diagnostics

Continuous processing and maintenance of the information requires substantial efforts from the practitioners and professionals, not only while handling and storing data, but also when interpreting it. Actually, it seems very hard to handle all the data without using data mining methods, which can autonomously dig out all the valuable knowledge that is embedded in a database without human supervision, providing a full life-cycle support of data analysis. Techniques such as clustering, classification, logical and association rule-based reasoning, and other methods are highly demanded for comprehensive environmental data analysis (Sokolova, Fernández-Caballero, & Gómez, 2010). Since its definition, the problem of mining knowledge from experimental data is being an area of active research, and many efficient techniques have been introduced and applied to various domains, including body dysmorphic disorder, MD and body image (Rodríguez-Solano, Laita, Roanes-Lozano, López-Corral, & Laita, 2006; Schmidt & Waligora, 2006).

For instance, data mining techniques for knowledge discovery and early diagnostics are used for early intervention in developmentally-delayed children (Chang, 2007). In reference Rodríguez-Solano et al. (2006) a computer-based expert system for diagnosis of depressive situations is presented. The system is based on three-based and modal logic, and computer algebra. The authors demonstrate the functionality of the expert system by introducing potential factors, which describe patients. In Bhattacharyya (2011) a new mining framework based on cohesion that is considered by the author as a representation of the properties prevalent among items in the domain of the dataset is introduced.

Many authors resort to artificial neural networks when managing problems of medical classification. For example, in Qasem and Shamsuddin (2011) there is a proposal of a time-variant multi-objective particle swarm optimization of radial basis function. Within the medical diagnostics presented in this paper, hepatitis is reviewed. There is a number of attributes including anorexia, fatigue, steroid, and others. The authors present the results with high validation and testing accuracy for the experimental data sets. On the other hand, some authors tend to stick to traditional methods, as the results obtained with them are easily understood and interpreted. Purposely, a method that uses Artificial Immune Systems algorithm to extract classification rules from hybrid trained neural network (Kahramanli & Allahverdi, 2008) is introduced in Kahramanli and Allahverdi (2009). This method was used to diagnose heart diseases and hepatitis.

3. Muscle dysmorphia detection

With respect to the current case study, its main goal is to discover the relationships between psychological individual characteristics of the participants and MD and to get classification patterns. In order to resolve these tasks, we widely apply data mining methods such as decision trees, rule classifiers, and statistical methods.

The flow chart of the experiment is given in Fig. 1. The first step is the collection of the experimental data, the second one is its preprocessing, which includes data clearing and training and testing sets preparation. Next, the modeling is undergone, and, after that, regression models and decision trees classifiers are evaluated.

Preprocessing includes clearing data from the outliers, gaps filling, data normalization and smoothing. Next, the correlation analysis is used to evaluate statistical relationships among quantitative variables. The mutual correlation between the variables both dependent and independent of a model leads to unstable and unreliable results. This phenomenon occurs because the variables interact with each other and impede the calculation of accurate model coefficients.

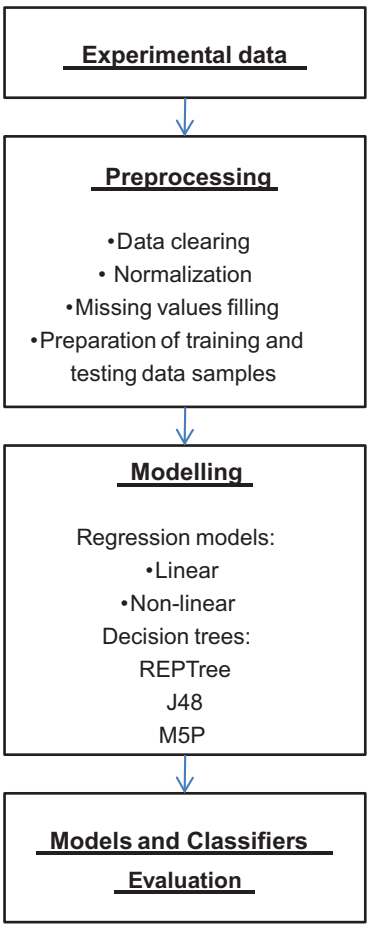

Fig. 1. The flow chart of the experiment.

The following step, according to Fig. 1, is modeling.

After having revised a number of available data mining solutions (Chang, 2007; Sokolova & Fernández-Caballero, 2009; Sokolova & Fernández-Caballero, 2012) it was decided to use regression analysis because regression models can be easily interpreted by a specialist, and decision trees and rule learners because they enable to create woody structures of "if-then" conditions, and facilitate clear perception of a created model as well as allowing to remove excessive rules using pruning procedure.

Linear and nonlinear regression models are created. The explanatory variables within each regression model do not correlate (the maximum correlation coefficient does not exceed the value of 0.7 according to the recommendation given in Sokolova (2010)). Decision trees is one of the standard classification methods which belongs to a "white box" family of data mining methods. Decision tree induction have been applied in many case studies thanks to its easy-to-interpret representation of models (Chandra & Paul Varghese, 2009; Chang, 2007). Final models have a form of tree structures where branches represent test outcomes, and terminal nodes (leaves) hold class labels (Han, Kamber, & Pei, 2011). These methods adopt a greedy approach, creating decision trees from top to bottom in a divide-and-conquer manner (Han et al., 2011).

The *M5P* and *REPTree* algorithms that were introduced by Quinlan (Witten, Frank, & Hall, 2011) are used in the study. *M5P* constructs a tree whose leaves are composed of multivariate regression models, and the nodes are the attributes which maximize error reduction over the standard deviation of the output