Chapter 13
Extracting Rules for Decreasing Body Fat Mass Using Various Classifiers from Daily Lifestyle Habits Data

Sho Ushikubo
Faculty of Science and Technology, Tokyo University of Science, Noda-shi, Japan

Katsutoshi Kanamori
Faculty of Science and Technology, Tokyo University of Science, Noda-shi, Japan

Hayato Ohwada
Faculty of Science and Technology, Tokyo University of Science, Noda-shi, Japan

ABSTRACT

This study was performed to extract rules and factors for reducing body fat mass and to compare Inductive Logic Programming (ILP) and common classifiers to provide necessary steps for the healthcare system. Many researchers have focused on lifestyle-related diseases; however, few have attempted to extract rules and factors for decreasing body fat mass. The authors obtained lifestyle habits data. This data includes a variety of features (e.g., sleep, exercise, and nutrient intake). These features are easier for patients to understand. ILP and common classifiers are applied to this data. In terms of accuracy, random forest outperformed all other methods, and random forest is suitable for extracting factors among common classifiers. However, in terms of rules, ILP is more suitable than others, because ILP can extract rules covering many positive and negative examples, and it is easy to apply to the healthcare system because these rules cover a range of features.

DOI: 10.4018/978-1-7998-2460-2.ch013

1. INTRODUCTION

Changes in health and social lifestyle in recent years have led to increasing numbers of patients suffering from lifestyle-related diseases, even among younger people, and thus pose important social problems in Japan (Ministry of Health, Labour and Welfare, 2001). These lifestyle-related diseases include chronic diseases (e.g., cancer, heart disease, and diabetes) that result from excess fat accumulation, or obesity. Daily habits such as nutrition intake, physical activity, and sleep are important factors that are related to lifestyle-related diseases. Moreover, body fat mass is generally related to obesity. Hence, a close study to extract factors and rules related to body fat mass is necessary for preventing obesity and lifestyle-related diseases. Moreover, there is no doubt that a healthcare system based on extracted rules helps maintain human health and prevent obesity. However, not all factors and rules that decrease body fat mass have been clarified.

Factors related to decreasing body fat mass and weight have been studied using statistical analysis such as the t-test. These studies focus on new findings and factors that are common to many people. However, opinions regarding these factors are divided among scientists. For example, a low-carb diet resulted in greater weight loss than a low-fat diet (Bazzano et al., 2014; Nordmann et al., 2006); however, some studies have found that the difference between a low-carb diet and a low-fat diet is small and not statistically significant (Guldbrand et al., 2012; Brinkworth et al., 2009). Moreover, a combination of certain nutrients such as calcium and vitamin D is associated with lower body weight and better metabolic health (Zhu et al., 2013). However, because these clinical studies consider only specific nutrients, they cannot point to specific factors and detailed rules.

Recently, some researchers have applied machine learning classification or regression to clinical or life science data, because machine learning algorithms can deal with mass data including many samples or many features. Plis et al. (2014) applied support vector regression (Smola & Vapnik, 1997) to clinical data in an effort to predict blood glucose levels for patients with diabetes. Their results indicated that support vector regression is more accurate than conventional statistical analyses such as the ARIMA model and t-tests. Babič et al. (2014) extracted important variables classifying patients with metabolic syndrome using decision tree (DT) (Quinlan, 1986). They ran blood tests on patients and obtained corrected data including many features of substances in the blood. Then they used DT to analyze the data, and extracted rules classifying patients. Their study is useful for setting new thresholds distinguishing patients and non-patients. However, extracted thresholds help only doctors; they are useless for the patients themselves, who cannot thoroughly understand blood variables. Hence, the authors assume that making lifestyle habit rules such as nutrition intake, exercise quantity, and sleeping hours is useful for patients to prevent disease or improve health. Previous studies used machine learning on exclusive clinical data; however, few of these studies focused on extracting rules and factors for decreasing body fat mass, which is related to improving diabetes or metabolic syndrome.

Therefore, in the present study, the authors focused on extracting rules and factors for decreasing body fat mass from lifestyle habits data using machine learning algorithms. The authors chose Naïve Bayes (NB) (Rish, 2001), K-nearest neighbor (k-NN) (Cover & Hart, 1967), support vector machine (SVM) (Suykens, & Vandewalle, 1999), logistic regression (LR) (Cox, 1958), DT, and random forest (RF) (L. Brieman, 2001) as common machine learning classification algorithms. In addition, the authors used Inductive Logic Programming (ILP) (Muggleton, 1991), which is commonly used because of virtual screening in drug discovery (Srinivasan et al., 1999; Geppert et al., 2010). Recently, ILP has been used for knowledge discovery of diseases using clinical data (Qiu et al., 2014). ILP can easily and logically