Subspace Discovery for Disease Management: A Case Study in Metabolic Syndrome

Josephine Namayanja, University of Maryland, Baltimore County, USA
Vandana P. Janeja, University of Maryland, Baltimore County, USA

ABSTRACT
This paper identifies key subspaces for better disease management. Disease affects individuals differently based on features such as age, race, and gender. The authors use data mining methods to discover which key factors of a disease are more relevant for particular strata of the population using bin wise clustering. The authors use a case study on Metabolic Syndrome (MetS). MetS is a combination of abnormalities that occur in the body during the processing of food and nutrients. A number of definitions have been studied to classify MetS. No clear criterion exists that can generally fit into a single satisfactory protocol. This domain encompasses a variety of demographics in society, leading to an implication that different criteria may be appropriate for different demographic strata. The authors address this issue and identify the cross section of demographic strata and the disease characteristics that are critical for understanding the disease in that subset of the population. Findings in real world NHANESII data support this hypothesis, thus the approach can be used by clinical scientists to narrow down specific demographic pools to further study impacts of key MetS characteristics.

Keywords: Clustering, K-Means, Metabolic Syndrome, Subspace Discovery, Sum of Squared Errors

INTRODUCTION
In various healthcare applications a large amount of data is gathered for individuals. However a few key attributes may be more critical than others for the disease management. In addition different attributes may be critical for different demographics based on age, race or gender. Thus, the focus of this paper is to identify key subspaces in large healthcare datasets for better disease management for varying demographic subsets in the data. Essentially we focus on using data mining methods to discover which key factors of a disease are more relevant for particular strata of the population using bin wise clustering. We focus on a case study in Metabolic Syndrome (MetS). MetS can be described as a combination of abnormalities that occur in the body during the processing of food and nutrients (Wright, 2005). A number of definitions have been studied to classify MetS; however, there is no clear criterion that can generally fit into a single satisfactory protocol. This is primarily because this domain encompasses quite a
Motivation

Metabolic Syndrome (MetS) can be described as a combination of abnormalities that occur in the body during the processing of food and nutrients. First recognized in 1988 (Wright, 2005; “The metabolic syndrome,” 2006), it was narrowed down to the existence of any 3 out of 5 of the following: insulin resistance, high blood pressure, high triglycerides, high cholesterol and a large abdominal circumference (“Schizophrenia and the metabolic syndrome,” 2006; Sherman, 2009).

MetS has become an epidemic to society with an impact of life threatening diseases like Cardiovascular Heart Disease, Type 2 Diabetes and Stroke. According to the World Health Organization (WHO) 180 million people suffer from Diabetes worldwide, with 19.3 million only from the U.S. which is an indicator that insulin resistance plays a key role in the possibility of MetS (Pradhan, 2007). Statistics have also shown that approximately 47 million people in the U.S. have become victim of MetS with men facing a higher risk at 24% and women following closely at 23% (“The metabolic syndrome,” 2006). Next we discuss several aspects of understanding and studying MetS that are relevant to our research from a data analysis perspective.

Definitions: A number of definitions have been brought to the table that try to classify MetS, however there is no clear criteria that can generally fit into a single satisfactory protocol of diagnosing MetS. Research particular to this domain encompasses quite a variety in demographic strata; as a result there is a possibility that different criteria are appropriate for different strata. The National Cholesterol Education Program’s Adult Treatment Panel III (NCEP-ATP III) (Grundy et al., 2004), which is the Third Report of the Expert Panel Detection, Evaluation and Treatment of High Blood Cholesterol in adults focuses on the use of a clinical trial approach to find new treatments for lowering cholesterol. It sets the criterion for MetS that is widely followed by researchers and other organizations and is estimated to be a more practical tool (Grundy et al., 2004). Table 1 shows a comparison of criteria for MetS as defined by different health organizations (adapted from Grundy et al., 2004). These include; National Cholesterol Education Program’s Adult Treatment Panel III (ATP III), World Health Organization (WHO), and American Association of Clinical Endocrinologists (AACE).

Effect of Race, Age and Gender: In clinical trials focusing on selected groups originating from different walks of life, factors like gender, race and age impact the health risks that one is bound to face. A study by CATIE (“Schizophrenia and the metabolic syndrome,” 2006) found that whites are more prone to the risk of MetS than blacks with numbers estimated at 44% and 30% respectively because of cholesterol imbalance. Additionally, the results of this study stated that people suffering from schizophrenia (a chronic, severe and disabling brain disease), especially women, faced a high risk of MetS (“Schizophrenia and the metabolic syndrome,” 2006). In a National Health Statistics report, the association between MetS and race and ethnicity varied by sex. Their results indicated that 25% non-Hispanic black males compared
The Value of Quantitative EEG Measures in the Early Diagnosis of Alzheimer’s Disease
