Nursing Informatics 2018 A.K. Rotegård et al. (Eds.) © 2018 International Medical Informatics Association (IMIA) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/978-1-61499-872-3-265

Predicting Depression Among Community Residing Older Adults: A Use of Machine Learning Approch

Jeungok Choi^a, Jeeyae Choi^b, Woo Jung Choi^c

^a College of Nursing University of Massachusetts, Amherst, MA, USA ^b School of Nursing, University of North Carolina Wilmington, NC, USA ^c Department of Epidemiology and Biostatistics, Boston University, MA, USA

Abstract

The study demonstrated an application of machine learning techniques in building a depression prediction model. We used the NSHAP II data (3,377 subjects and 261 variables) and built the models using a logistic regression with and without L1 regularization. Depression prediction rates ranged 58.33% to 90.48% and 83.33% to 90.44% in the model with and without L1 regularization, respectively. The moderate to high prediction rates imply that the machine learning algorithms built the prediction models successfully.

Keywords:

machine learning, depression, prediction model, logistic regression model with and without L1 regularization model

Introduction

In recent years, rapid advances of information technologies have generated vast amount of data, and nursing is not an exception. Nursing generates data on nursing science and patient care at the exponential rate. In the clinical side, patient data are entered to electronic health records (EHRs) every shift by nurses. Consumers' data from biosensors, wearable and mobile devices are also added to the EHR every day [1]. These health data are defined as "big data" because they are: 1) high volume, 2) high velocity and 3) high variety of data that require sophisticated computational methods of data processing and management. Data mining using machine learning approches has gained popularity in recent years due to its limitless application and the ability to handle large and complex data. The purpose of this study was to demonstrate an application of machine learning techniques in building a depression prediction model for community-residing older adults.

Methods

The National Social Life, Health, and Aging Project (NSHAP) Wave 2 data were used (n=3,377) [2]. The NSHAP is a longitudinal, population-based study of health and social factors, aiming to understand the well-being of older, community-residing Americans. The Wave 2 data includes variables on physical health and illness, medication use, cognitive function, emotional health, sensory function, health behaviors, social connectedness, sexuality, and relationship quality. A total of 261 variables were included in the analysis. We used supervised machine learning algorithms, a logistic regression with and without L1 regularization [3]. L1 regularization was employed to prevent over-fitting of the model and also improve the performance of the model. To evaluate the performed. MATLAB (version 2012a) was used for training models and computations. Common point and Expectation-Maximisation (EM) imputations were used to replace missing value.

Results

A logistic regression model used 261 variables while a logistic regression with L1 regularization model used 32 variables for computations. Table 1 shows the depression prediction rates by the two mdoels.

Table 1- Prediction Rates By Two Models

Depression	Prediction rate at training	Prediction rate at testing
Logistic Regression without L1 Regularization		
mild	94.46%	90.44%
severe	100.00%	83.33%
Logistic Regression with L1 Regularization		
mild	91.59%	90.48%
severe	60.05%	58.33%

Discussion and Conclusions

The moderate to high prediction rates for the models imply that a logistic regression with and without L1 regularization algorithms built the prediction models successfully. The logistic regression with L1 regularization model showed lower performance than the model without L1 regularization. The results demonstrate that machine learning techniques are feasible and appropriate to use in nursing research when discovering new knowledge from large datasets, such as developing a depression prediction model in this study. However, these prediction models cannot be generalized until they are further validated with other health datasets.

References

- P.F. Brennan, S. Bakken, Nursing needs big data and big data needs nursing. J Nurs Scholarsh. 47(2015), 477-84.
- [2] National Social Life Health and Aging Project. NORC at the University at Chicago, Retrieved October 24 2017 from <u>http://www.norc.org/research/projects/pages/national-</u> <u>social-life-health-and-aging-project.aspx</u>
- [3] N. Prashant, J.C. Frenzel, D.H. Smaltz, *Demystifying Big Data and Machine Learning for Healthcare*, CRC Press, FL, 2017.

Address for correspondence

Jeungok Choi, jeungokc@nursing.umass.edu