ProHealthSu17 - Kate Sanders - Rob Long

[Extended Abstract]

Author One*
Institute One
Address One
author.one@emails.com

Author Two
Institute Two
Address Two
author.two@emails.com

Author Three
Institute Three
Address Three
author.three@emails.com

Author Four
Institute Four
Address Four
author.four@emails.com

ABSTRACT

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Theory

Keywords

ACM proceedings, LaTeX, text tagging

1. ABSTRACT

We created a Bayesian Network that models and tracks human behavior. The data set that we used consists of health records from 5,115 black and white, males and females, ages 18-30, called CARDIA data.[3] Using the CARDIA data we were able to track behavioral changes amongst patients, through the implementation of our Bayesian Network. Our Network then, modeled these changes and made statistical inferences based off different attributes within our network. These inferences can aid physicians in building a patient specific decision support model. Hospitals can then use these decision support models to produce a more personalized medical approach for treating their patients.

2. INTRODUCTION

comment

Trying to track the behavior of an individual over time is a daunting and time consuming task. Due to the tendency for behavior to drastically change over time. These short term behavioral alterations play a huge role in the patients health status, but are extremely hard to observe.This poses a huge problem when trying to make accurate representations of behavioral data. Thus, if a model were to be created displaying these changes it could help patients make better decisions about their own lives.[w4:sp]field

There are complications involved in the process of producing a behavioral model as such. Just as tracking the behavior of an individual was time consuming, so is creating a longitudinal model. Longitudinal models are sometimes intractable, making it hard to deal with. Also, there is copious amounts of background research that needs to be done before constructing and implementing such a complex idea.[w5:sp]field

In order to generate an elaborate model as such we used a Bayesian Network. The Bayesian Network enabled us to represent the dependence between variables and gave a concise specification of the joint probability distribution.[9] Within the network each node or variable is equipped with a conditional probability table (CPT). The nodes are then connected by edges that represent direct influence of one node to another. This connection permits the network to make statistically based inferences, in modeling the CARDIA data set.[9] As a guide for implementing this model we used Bayes rule which is: P(A|B) = P(B|A)P(A) / P(B).

Within our Bayesian network, we used the variable elimination. Variable elimination is a standard algorithm for computing probability of evidence with regard to a given Bayesian network [14] This to successfully utilize local structure in the form of determinism [5] and contextâĂŞspecific independence[2] to execute inference more systematically. Also, it enables us to answer multiple queries at the same time.

Currently, there are Bayesian Networks that are used to show connections between behavior and disease, but none that specifically represent behavior. To fix this problem, we have created a preliminary Bayesian Network that will aid in tracking and modeling human behavior using CARDIA data.

CARDIA or The Coronary Artery Risk Development in Young Adults Study, looks at the growth and determinants of clin-

*Now on postdoctoral fellow at ABC University ical and sub-clinical cardiovascular disease and its risk components.[1] This data set records 5,115 black and white, males and females, ages 18-30. [3] CARDIA presents the enrollment and examination methods, the mean levels of blood pressure, total plasma cholesterol, height, weight and body mass index, and the prevalence of cigarette smoking by age, sex, race and educational standing. [3]

In this paper, we make the following contributions: our research can be used to build a decision support model for hospitals to aid in several tasks including; resource allocation, treatment planning, and prospectively provide physicians with valuable resources required for making informed decisions. Resulting in a more personalized medical approach for hospitals to use when treating their patients. Also, the methods we used to model these behavioral changes can serve as a backbone for further modeling related work.

3. RELATED WORKS

Previous research with the CARDIA database statistically links behaviors and socioeconomic status with psychosocial vulnerability [11, 10], physical fitness [7, 12], and relative risks of heart disease [6].

Bayesian Networks have been used to model the longitudinal links between behavioral and socioeconomic data and Coronary Artery Calcification [13]. Machine learning techniques have also been used to predict whether an individual has a rare disease based solely on behavioral data [8]. However, this paper contains the first implementation of a Bayesian Network for modeling solely longitudinal behavioral data.

4. METHODS

The data used to create this network was gathered through the Coronary Artery Risk Development in Young Adults (CARDIA) study. The specifics of the study procedures are detailed elsewhere[3]. This study followed 5115 subjects from 1985-6 until present. Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA served as centers for data collection. Each location recruited participants evenly distributed between the subgroups of gender, race, education level, and age-group (18-25 or 25-30). Data gathered from participants included physical measurements, clinical tests, and an in depth questionnaire about lifestyle and socioeconomic status. This study has served as the backbone for major research in the development of heart disease.

We consider only the longitudinal lifestyle and socioeconomic data in our research. Socioeconomic factors included age, sex, race, education level, marital status, employment status, difficulty paying for basic necessities, home ownership, and health insurance status.

To model how behaviors change over time, we applied the CARDIA data to BN learning algorithms, which used a hill-climbing approach. To score structures we use BIC, Mutual information, and BDe, [4] because these metrics have shown promising in previous BN research [13].

5. RESULTS AND DISCUSSIONS

6. REFERENCES

[1]

[2] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-Specific Independence in Bayesian Networks. In *Proceedings of the Twelfth Conference Conference on Uncertainty in Artificial Intelligence*, volume 4, pages 115–123, 1996.

[3] G. D. Friedman, G. R. Cutter, R. P. Donahue, G. H. Hughes, S. B. Hulley, D. R. Jacobs, K. Liu, and P. J. Savage. Cardia: study design, recruitment, and some characteristics of the examined subjects. *Journal of Clinical Epidemiology*, 41(11):1105–1116, jan 1988.

[4] D. Heckerman, D. Gelger, and D. M. Chickeriag. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data.

[5] F. V. Jensen, K. G. Olesen, and S. K. Andersen. An algebra of bayesian belief universes for knowledge???based systems. *Networks*, 20(5):637–659, 1990.

[6] A. S. Karlamangla, B. H. Singer, D. R. Williams, J. E. Schwartz, K. A. Matthews, C. I. Kiefe, and T. E. Seeman. Impact of socioeconomic status on longitudinal accumulation of cardiovascular risk in young adults: the CARDIA Study (USA). *Social Science & Medicine*, 60(5):999–1015, 2005.

[7] C. E. Lewis, D. E. Smith, D. D. Wallace, O. D. Williams, D. E. Bild, D. R. Jacobs, and Jr. Seven-year trends in body weight and associations with lifestyle and behavioral characteristics in black and white young adults: the CARDIA study. *American journal of public health*, 87(4):635–42, apr 1997.

[8] H. MacLeod, S. Yang, K. Oakes, K. Connelly, and S. Natarajan. Identifying Rare Diseases from Behavioural Data: A Machine Learning Approach. In *Proceedings - 2016 IEEE 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2016*, 2016.

Feature Name	Type	Threshold
Marital Status	Boolean	
Employment Status	Boolean	
Home Ownership	Boolean	
Health Insurance Status	Boolean	
Obesity	Boolean	
Difficulty Paying for Basic Necessities	Boolean	
Sex	Boolean	
Electrocardiogram and Echocardiography	Boolean	
Coronary calcium	Boolean	
Age	Ordinal	18,24,25,30
Race	Ordinal	1,8
Education Level	Ordinal	1,8
Blood Pressure	Ordinal	cell6
Chemistries	Ordinal	
Anthropometry	Ordinal	1,500
Medical history	Ordinal	1,5
Family history	Ordinal	1,5
Physical Activity	Ordinal	1,10
Nutrient intake/dietary history	Ordinal	1,10
Psychosocial parameters	Ordinal	1,10
Pulmonary function	Ordinal	1,3
Carotid intimal medial thickness	Ordinal	1,3
Genetic studies	Ordinal	1,10

[9] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*, volume 9. 1995.

[10] L. W. Scherwitz, L. L. Perkins, M. A. Chesrtey, G. H. Hughes, S. Sidney, and T. A. Manolio. Hostility and Health Behaviors in Young Adults: The CARDIA Study. *American Journal of Epidemiology*, 136(2):136–145, jul 1992.

[11] G. Scherwitz, L; Perkins, L; Chesney, M; Hughes. Cook-Medley Hostility scale and subsets: relationship to dem... : Psychosomatic Medicine. *Psychosomatic Medicine*, 53(1):36–49, 1991.

[12] M. H. Shishehbor, P. Gordon-Larsen, C. I. Kiefe, and D. Litaker. Association of Neighborhood Socioeconomic Status with Physical Fitness in Healthy Young Adults: the CARDIA Study.

[13] S. Yang, K. Kersting, G. Terry, J. Carr, and S. Natarajan. Modeling Coronary Artery Calcification Levels From Behavioral Data in a Clinical Study.

[14] N. L. Zhang and D. Poole. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.