Expert Systems With Applications 55 (2016) 469–479

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Predictive analytics for delivering prevention services

Seokgi Lee[a], Yuncheol Kang[b,*], Nicholas S. Ialongo[c], Vittaldas V. Prabhu[d]

[a] Department of Industrial Engineering, University of Miami, 1251 Memorial Drive 281, Coral Gables, FL 33146, USA
[b] Department of Industrial Engineering, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul 04066, Republic of Korea
[c] Department of Mental Health, Johns Hopkins University, 624 N. Broadway, 8th Fl., Baltimore, MD 21205, USA
[d] Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, 310 Leonhard Building, PA 16802

ARTICLE INFO

Keywords:
Prevention service
Evidence-based prevention programs
Logical Analysis of Data (LAD)

ABSTRACT

Early diagnosis and prevention of problematic behaviors in youth play an important role in reducing treatment costs and decreasing the toll of antisocial behavior. Over the last several years, the science of preventing antisocial behavior in youth has made significant strides, with the development of evidence-based prevention programs (EBP) using randomized clinical trials. In this paper, we use a real case implemented by schools in an urban school district of 80,000 students in a mid-Atlantic state to show how predictive analytics can help to improve the quality of prevention programs and reduce the cost of delivering associated services. Data patterns are extracted from conduct disorder assessments using the Teacher Observation of Classroom Adaptation (TOCA) screening instrument and evaluated using the results of the Diagnostic Interview Schedule for Children (DISC). A mathematical method called Logical Analysis of Data (LAD) is used to analyze data patterns. Experimental results show that up to 91.58% of the cost of administering DISC would be saved by correctly identifying participants without conduct disorder and excluding them from the DISC test.

1. Introduction

Over the last several years the science of preventing antisocial behavior in youth has made significant strides through the development of evidence-based prevention programs (EBP) using randomized clinical trials (Greenberg, Domitrovich, & Bumbarger, 1999). Effective EBP can reduce delinquency, aggression, violence, bullying and substance abuse among youth, which in turn could potentially lower the risk of violent incidents. For instance, Miller and Hendrie (2009) calculated that the annual cost of substance abuse in the U.S. was $510.8 billion in 1999 and that effective school-based prevention programs implemented nationwide could save $18 for every $1 spent on intervention programs (Miller & Hendrie, 2009). Both healthcare and education are currently recognized as important service sectors, together accounting for 27% of the U.S. GDP (Larson, 2009), so any potential cost savings could have important implications, from both the social welfare and economic perspectives.

EBP services are commonly delivered to youth, and sometimes their families, individually or in groups through school and community settings (Bumbarger & Perkins, 2008; Bumbarger, Perkins, & Greenberg, 2010; Greenberg, 2004; Greenberg, Domitrovich, Graczyk, & Zins, 2005). Designing and implementing prevention services for youth requires incorporation of data and research from multiple fields, including mental health and education. Researchers and providers of prevention programs have developed strategic plans and methods for implementing prevention services, including identification of those at-risk. The goal is not only to substantially reduce the long-term risks associated with the occurrence of antisocial behavior, but also to ensure the economic viability of delivering the associated prevention services. It should be emphasized that planning and implementing prevention services can be improved through better decision-making, based on analysis of the data collected in associated processes.

The Baltimore City Public Schools (BCPSS), in collaboration with Johns Hopkins University, implemented a number of such programs to provide prevention services in their urban school environment. This article details one of these prevention programs and highlights the unique perspectives of the participants and providers. The prevention program used the Teacher Observation of Classroom Adaptation (TOCA) by Kellam, Branch, Agrawal, and Ensminger (1975) as a universal screening instrument to determine

* Corresponding author. Tel.: +82-2-320-3072.
E-mail addresses: sgl14@miami.edu (S. Lee), yckang@hongik.ac.kr (Y. Kang), nialong1@jhu.edu (N.S. Ialongo), vxp7@engr.psu.edu (V.V. Prabhu).

http://dx.doi.org/10.1016/j.eswa.2016.02.023
0957-4174/

the overall average for risk behavior. TOCA was designed to measure the classroom behaviors of children and adolescents (ages 6 to 18 years) according to various criteria and to serve as an efficient and self-administered version of the measure used in large-scale school-based research trials (Koth, Bradshaw, & Leaf, 2009). As an intensive diagnostic instrument, the Diagnostic Interview Schedule for Children (DISC) is used to establish the presence or absence of conduct disorder; the results are used primarily for epidemiological research but are also useful in clinical settings (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000). It should be noted that the TOCA test is for screening participants who are potentially at risk of conduct disorder, and the cost is relatively low compared to an intensive diagnostic test like the DISC. On the other hand, the DISC test aims to identify conduct disorder in symptomatic individuals, resulting in higher costs associated with establishing and justifying a diagnosis.

From an implementation and management point of view, prevention programs and services usually rely on labor-intensive, error-prone, and cumbersome paper-based processes for data collection and reporting. For example, assessments of participants' behavioral characteristics by TOCA and DISC are often examined and managed through paper records, which are not suitable for a large-scale data environment. Therefore, there is an acute need for robust and scalable software-based solutions for EBP that can be used for cost-efficient and accurate data collection and for monitoring of implementation and compliance in the natural world. The need for automated and computerized data collection and management led to the development of the INtegrated System for Program Implementation and Real-time Evaluation (INSPIRE; Domitrovich et al. (2008)). INSPIRE is a web-based data management system developed at the Pennsylvania State University in 2004 to help schools with the process of administering, evaluating, and sustaining evidenced-based interventions and practices at a high level of quality. It provides users, including teachers, coaches, principals, and student counselors, with educational processes and data management applications. Using a customized version of TOCA for BCPSS supported by INSPIRE, users can rate participants' academic skills, record and analyze their behavior characteristics in real time. Furthermore, DISC data collected by different computer systems can be transferred through the INSPIRE data interface and integrated with the INSPIRE database.

In this paper, DISC and TOCA assessment results for 926 participants in BCPSS interviewed during the 2006–2007 academic year are analyzed using a mathematical method called Logical Analysis of Data (LAD). LAD is a supervised learning method based on combinatorics, optimization, and the theory of Boolean functions (Alexe, Alexe, Bonates, & Kogan, 2007; Hammer & Bonates, 2006) and has been proven for its powerful mathematical analytics in terms of handling the type of inseparable data that are typical in medical applications, such as cancer diagnosis or prognosis, and feature analysis in genomics and proteomics (Alexe et al., 2006; 2003; Reddy et al., 2008). In spite of those successful implementations, the mental health area has not been handled by LAD yet; Kim and Muthén (2009) conducted similar research on identification of latent classes of children in terms of their level of aggression, but the method used in their research was based on a statistical approach called two-part factor mixture modeling.

In this paper, we tailor the LAD procedures to identify primary attributes and their patterns from TOCA for explaining participants' problematic or non-problematic behaviors and utilize patterns to predict the binary outcomes from DISC.

A key motivation of this paper is to use TOCA data to identify participants who are highly likely to have a negative diagnosis and avoid the burden and associated cost of DISC for these participants. Furthermore, we are also interested in "false negative" observations obtained by LAD, because they would eliminate the cost of DISC but increase social risks in the future. Thus, this paper evaluates the performance of the LAD approach and describes the pros and cons of using this approach.

The rest of this paper is organized as follows. In Section 2, the fundamentals of LAD are explained. In Section 3, the general data structure of DISC and TOCA is explained. In Section 4, LAD patterns and analytic models are explained. Lastly, in Section 5, possible advantages and limitations of using the LAD approach in prevention services are discussed and analyzed from an economic perspective.

2. Logical analysis of data

One significant advantage of LAD compared to other classification approaches based on statistics, such as regression and mixture modeling, is that the data used in LAD do not need to belong to any statistical distribution, so no prior statistical analysis or assumptions must be made. Latent class and clustering analysis are also capable of classifying subgroups, but there are usually functional limits to the analysis of within-group variability, due to the small degrees of freedom resulting from small sample sizes compared to the number of variables, and violations of the underlying statistical assumptions, which make it difficult to design effective individual treatment (Kreuter & Muthén, 2008). It should also be emphasized that, in contrast to black box type methods such as *Support Vector Machines* and *NeuralNetworks*, LAD provides a strong justification for the reasons for each positive and negative classification, and thus offers the possibility of implementing individualized treatment (Hammer & Bonates, 2006).

There exist technical limitations of LAD; that, for example, the computational complexity would exponentially increase due to the internal optimization procedures. Another example is that overfitted classifiers would be possibly obtained by outliers in observations. Such technical issues, however, have been successfully addressed by both heuristic or optimal approaches (Han, Kim, Yum, & Jeong, 2011; Ryoo & Jang, 2009), and LAD is now known as being very comparable with that of the best methods used in data analysis such as support vector machine, decision trees, nearest neighbors, neural networks (Hammer & Bonates, 2006).

In this section, overall procedures of LAD are briefly explained, including notation, mathematical meaning, and objectives of LAD steps. For more detailed and in-depth knowledge of LAD, a reader is referred to the following works: Alexe et al. (2007), Boros, Peter, Ibaraki, and Kogan (1997) and Boros et al. (2000).

2.1. Observation and pattern

The primary goals of LAD are to generate data *patterns* that are capable of explaining binary outputs, positive or negative, of previously observed data and to use them to predict classifications of new observations. In this section, we define an observation and a pattern for the purpose of the LAD analytic approach.

Let us define Θ as an archive of past observations, and the element of Θ is an n-dimensional vector whose components are called *attributes*, or *features*, or *variables*. Each observation is classified as a binary result (positive or negative), and Θ is partitioned into two disjoint subsets Θ^+ and Θ^-, i.e., $\Theta^+ \cup \Theta^- = \Theta$ and $\Theta^+ \cap \Theta^- = \emptyset$, which contain the positive and negative observations, respectively.

Composition of an archive (Θ^+, Θ^-) is enabled by a *partially defined Boolean function* (pdBf) f, which is a mapping $\Theta \to \{1, 0\}$, and any completely defined Boolean function which maps all observations in Θ into $\{1, 0\}$ is called the *extension* of f. The goal of LAD is therefore to determine f or the extension of f to correctly classify unknown observations.

A basic construction of a pattern is made from a literal which is either a Boolean variable or its negation. A conjunction of literals is