Computers & Industrial Engineering 66 (2013) 1085–1095

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

A Bernoulli–Gaussian mixture model of donation likelihood and monetary value: An application to alumni segmentation in a university setting

Pablo L. Durango-Cohen [a], Elizabeth J. Durango-Cohen [b,*], Ramón L. Torres [c]

[a] Civil & Environmental Engineering and Transportation Center, Northwestern University, 2145 Sheridan Road, Room A332, Evanston, IL 60208, USA
[b] Stuart School of Business, Illinois Institute of Technology, 565 W. Adams Street, Room 447, Chicago, IL 60661, USA
[c] Civil & Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 15 December 2012
Received in revised form 30 May 2013
Accepted 9 August 2013
Available online 20 August 2013

Keywords:
Market segmentation
Finite mixture models
Alumni giving
Fundraising
Bernoulli–Gaussian Distribution
Expectation–Maximization Algorithm

ABSTRACT

Advances in computational power and enterprise technology, e.g., Customer Relationship Management (CRM) software and data warehouses, allow many businesses to collect a wealth of information on large numbers of consumers. This includes information on past purchasing behavior, demographic characteristics, as well as how consumers interact with the organization, e.g., in events, on the web. The ability to mine such data sets is crucial to an organization's ability to deliver better customer service, as well as manage its resource allocation decisions. To this end, we formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of repeat transactions. In addition to presenting the model, we derive an instance of the Expectation–Maximization Algorithm to estimate the associated parameters, and to segment the consumer population.

We apply the model to an extensive dataset of donations received at a private, Ph.D.-granting university in the Midwestern United States. We use the model to assess the effect of individual traits on their contribution likelihood and monetary value, discuss insights stemming from the results, and how the model can be used to support resource allocation decisions. For example, we find that participation in alumni-oriented activities, i.e., reunions or travel programs, is associated with increased donation likelihood and value, and that fraternity/sorority membership magnifies this effect. The presence/characterization of unobserved, cross-sectional heterogeneity in the data set, i.e., unobserved/unexplained systematic differences among individuals, is, perhaps, our most important finding. Finally, we argue that the proposed segmentation approach is more appealing than alternatives appearing in the literature that consider donation likelihood and monetary value separately. Among them and as a benchmark, we compare the proposed model to a segmentation that builds on a multivariate Normal mixture model, and conclude that the Bernoulli–Gaussian mixture model provides a more coherent approach to generate segments.

1. Introduction

As the world economy steers through a global recession, the "once-booming nonprofit sector [in the United States] is in the midst of a shakeout" (Banjo & Kalita, 2010), a consequence of government funding cuts and declines in donations from individuals. As a case in point, institutions of higher education experienced declines in contributions by individuals in 2009 and 2010 of 6.2% and 11.9%, respectively (Hall & Joslyn, 2011) – record drops in 50 years of recordkeeping. Overall impacts are discussed in studies such as Nicas and McWhirter, 2012, who argue that the severity of the cuts is adversely affecting educational quality, and even the solvency of institutions. This landscape has, in part, motivated universities to increase the volume and sophistication of their fundraising efforts.

Central to university fundraisers' efforts are two primary goals: meeting fundraising targets set by university administrators, and increasing participation rates in fundraising campaigns among alumni. Solicitation strategies are often intended to foster "loyalty/commitment" with younger alumni by encouraging nominal contributions to annual funds, e.g., \$50/year. The importance of increasing participation is magnified by the role it plays in university ranking systems. US News & World Report, the most widely recognized college ranking service, for example, assigns a 5% weight to an "Alumni Giving" category, which the magazine

* Corresponding author. Tel.: +1 312 906 6579; fax: +1 312 906 6503.
E-mail addresses: pdc@northwestern.edu (P.L. Durango-Cohen), durango-cohen@iit.edu (E.J. Durango-Cohen), ramon@u.northwestern.edu (R.L. Torres).

0360-8352/$ - see front matter
http://dx.doi.org/10.1016/j.cie.2013.08.007

defines as the percentage of undergraduate alumni of record who donate money to the college or university in a given year.[1] The percentage of alumni donors serves as a proxy for student satisfaction. At the same time, fundraisers seek to increase the monetary value of contributions as alumni advance professionally and financially. This goal is increasingly critical because a substantial portion of the funds raised by universities come from individual donors; in 2011, for instance, \$13.45 billion (44%) of the \$30.30 billion raised by colleges came from donations made by individuals (Council for Aid to Education, 2012). The development of effective solicitation strategies aimed at increasing the value and frequency of donations relies on understanding the contribution behavior of alumni, and how it evolves over time. In this paper, we present a model that advances this objective.

Specifically, we formulate a finite mixture model that *jointly* describes the likelihood and monetary value of donations in a university setting. The proposed model, a Bernoulli–Gaussian (BG) mixture model, relies on the assumption that the population is comprised of a finite set of latent classes/segments in unknown proportions. Each segment is characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in any given year. In addition to presenting the model, we derive an instance of the Expectation–Maximization (EM) Algorithm to estimate the associated parameters, and to assign donors to the segments in the population. For a given individual, segment membership is established based on the probabilities that her contribution sequence is consistent with the distribution functions that characterize the segments. In terms of validation, we use the model to analyze donations at a private, Ph.D.-granting university in the Midwestern United States. The dataset consists of 282,888 annual contributions from 75,922 individuals taking place between fiscal years 2000 and 2010. The results show that the proposed model adequately fits the contribution data. Further, we use the model to assess the effect of individual traits on their contribution likelihood and monetary value, discuss insights stemming from the results, and how the model can be used to support resource allocation decisions. The presence/characterization of unobserved, cross-sectional heterogeneity in the data set, i.e., unobserved/unexplained systematic differences among individuals, also constitutes one of the significant findings/results of our analysis. Finally, we argue that the proposed segmentation approach is more appealing than alternatives that consider donation likelihood and monetary value separately. Among them and as a benchmark, we compare the proposed model to a segmentation that builds on a multivariate Normal mixture model, and discuss the managerial implications of the results.

This paper contributes a new model that can be used to mine extensive panel data sets in situations where longitudinal data, i.e., transactions, are intermittent, and where there is significant cross-sectional (unobserved) heterogeneity among individuals. As such, the model has applicability to any setting where individuals are associated with large numbers of transactions, and where it is of interest to focus on/allocate resources based on a small number of segments instead of a large number of individuals. Advances in technology and computational power are, of course, making these settings increasingly common in research and in practice. Example application areas in Industrial Engineering include (dynamic) catalog mailing decisions (Simester, Sun, & Tsitsiklis, 2006), devising marketing strategies for (online) retailing (Ha, Bae, & Park, 2002; Jonker, Piersma, & Van den Poel, 2004), monitoring usage rates of (subscription-based) services (Samimi & Aghaie, 2011), and others. The remainder of the paper is organized as follows: Section 2 positions our work with respect to the literature. In Section 3 we describe the data used in this study. In Section 4, we introduce notation and assumptions to formulate BG mixture models. We also present an instance of the EM Algorithm to estimate the model's parameters. Results from an extensive empirical study, highlighting the insights gained by processing transaction sequences, and by segmenting the alumni population on the basis of donation likelihood *and* monetary value, are presented in Section 5. We conclude with a summary of the findings in Section 6.

2. Related work

Rather than providing comprehensive overviews of the fundraising or segmentation literatures, our objective is to contrast the approach implied by the proposed segmentation model to others appearing in the literature. In this case, the university fundraising context provides an interesting and relevant application – one that has been the subject of significant recent research. For broader reviews of fundraising and segmentation readers are referred to Bekkers and Wiepking, 2011a; Bekkers and Wiepking, 2011b; Wiepking and Bekkers, 2012 and Wedel and Kamakura, 2000.

To position the proposed model, we rely on the taxonomy of market segmentation models presented in Wedel and Kamakura, 2000. Among others, they rely on the following attributes: segmentation approach, segmentation bases, and type of statistical model used to explain responses. With respect to segmentation approach, models are classified as either *a priori* or *post hoc*. In *a priori* segmentation models, the number and types of segments are determined in advance of the analysis, whereas in *post hoc* segmentation they are determined as a result of the analysis based on goodness-of-fit or other criteria. The conditions used to assign individuals to segments are referred to as segmentation bases. Segmentation bases are either *observed* when they rely on observed/measured trait, response, or institutional variables; or *unobserved*, frequently attributed either to unobserved/missing data or to latent/unobservable (psychographic) variables. Finally, with respect to the capability to explain responses, models are classified as *predictive* or *descriptive*. Predictive models relate explanatory variables to the outcomes of a set of dependent variables. In contrast, descriptive models represent the (joint) distribution of the variables without distinction between outcome and explanatory variables (Green, 1977). Based on this taxonomy, the proposed model can be categorized as a *post hoc* segmentation procedure where the underlying statistical models are *descriptive*, and where the segmentation basis is *unobserved*. We note that this is consistent with the assumption that cross-sectional heterogeneity can be, at least partly, attributed to unobserved, and potentially unobservable factors.

The prevailing approaches in segmentation/econometric studies of (higher education) fundraising involve assigning individuals to a set of pre-established segments based on observed variables and criteria. Some examples in the literature include: undergraduate vs. graduate-degreed alumni segments, and segments of annual fund contributors vs. major donors who contribute at least \$X per year or over a lifetime. These approaches are, therefore, *a priori* segmentation methods relying on observed segmentation bases. Segmentation bases generally consist of either trait variables (e.g.: demographic or socioeconomic characteristics), institutional variables (e.g.: size, private vs. public institutions, etc.), or response variables (e.g.: Recency, Frequency and Monetary Value of donations – RFM data).[2] Numerous studies combine segmentation and econometric analyses, generally, to explain and reveal insights

[1] http://www.usnews.com/education/best-colleges/articles/2012/09/11/methodology-undergraduate-ranking-criteria-and-weights-2.

[2] In the fundraising setting, recency refers to the number of periods since the last contribution, frequency to the total number of contributions over some period of time, and monetary value to an average dollar amount per contribution, or to the dollar value of the last donation.