The Journal of Academic Librarianship xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

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

Knowing What the Patron Wants: Using Predictive Analytics to Transform Library Decision Making

Ryan Litsey*, Weston Mauldin

Texas Tech University Libraries, 2802 18th St., Lubbock, TX 79409, United States

ARTICLE INFO

Keywords:
Machine learning
Predictive analytics
Interlibrary loan
Collection development
Access services

ABSTRACT

Predictive analytics and machine learning are burgeoning areas of professional practice for large corporations especially businesses that offer products and services to customers. The power to better understand the movement of large amounts of data in a company and the capability to deploy that data to meet a customer's needs is invaluable from a services standpoint. Some in libraries have theorized that this type of data usage could possibly be used in a library service environment as well. In this article, we demonstrate how you can develop and use machine learning algorithms and predictive analytics to proactively understand library behavior. Although libraries are good at data collection, we often rely on statics or old data for assessment. Utilizing a machine learning system, called the Automated Library Information Exchange Network (ALIEN), we can better understand the movement of the items in the collection and better serve the needs of our customers the library patrons.

Introduction

Libraries are excellent collectors of information. In one single day, any given library will track, catalog and store gigabytes of information. Typically, this information is static and stored until it is called upon to generate some type of behavioral observation. For example, we store circulation data for a given time and then run a series of reports to determine which books are checked out the most or determine which user status circulates which books. We use this information to help make collection development decisions or determine workload for stacks shelving. This type of data collection happens on a constant basis throughout the library and even within the services offered by many library vendors. A second example is the book profiles that many libraries use to determine the gaps in a collection. These profiles gather cataloged collection data and display the titles that are missing from a certain genre, or specifically the new titles that might be worth purchasing. All this type of data is more or less static data, but what does that mean?

Static data is data that is stored for analysis at some point in the future. An example of static data is the circulation data mentioned above. Circulation data is downloaded from the library system and stored in an Excel spreadsheet. From there, the data can be manipulated to produce results. Although static data has good uses, it can quickly become outdated and irrelevant.

Since static data must be accessed and assessed, it cannot account for micro level changes in library behavior. Libraries need to develop new methods of data analysis beyond static data. New technologies in the field of machine learning can help guide libraries to develop new tools that can analyze, interpret and provide direction in near real time. This type of data analysis can help library managers, selectors and library services operate at peak efficiency. This article seeks to demonstrate how this can be accomplished through a system we have called ALIEN.

The Automated Library Information Exchange Network (ALIEN) is the first library developed machine learning/predictive analytic tool that can analyze a variety of disparate library data points, combine those points into a holistic picture of library behavior, as well as recommend and predict services and actions that can help improve the overall function of the library. While still in its infancy, ALIEN demonstrates how libraries can leverage the data collected daily and deploy it to serve a business needs that can help improve patron services. Imagine for a moment that you could predict the most requested books for Interlibrary Loan and have those books requested by the system ahead of time. Predicting the requests allow those books to be on the shelf before the patron knows they want them. Or imagine you could see the changing preferences of library users and enhance the collection in a new direction. That is the power predictive analytics and machine learning can give libraries; more than just informed decision making using numbers, but contextually relevant decision making.

ALIEN is still in development, but with early success comes new lessons learned. Some of the potential uses of a machine learning system in libraries has been discussed recently in library, computer

* Corresponding author.
E-mail addresses: Ryan.litsey@ttu.edu (R. Litsey), W.mauldin@ttu.edu (W. Mauldin).

http://dx.doi.org/10.1016/j.acalib.2017.09.004
Received 16 May 2017; Received in revised form 5 September 2017; Accepted 15 September 2017

science and business literature. Through this article we will first demonstrate how the theories of machine learning and predictive analytics can be applied to libraries. Second, we will discuss how such a system can be developed and deployed to answer near real time everyday library questions. Finally, we will discuss the future steps and potential for a system like this to work in concert with library staff to help create a more accurate and near real time holistic picture of library behavior.

Literature review

To better understand the power and possibilities associated with a machine learning predictive analytics it important to examine the literature surrounding these systems and how they might apply to libraries. Bruce Massis (2012) in his article "What's New in Libraries; Using predictive analytics in the library" offers one of the first theoretical applications of predictive analytics to the library environment. Massis writes, "Predictive analytics can also greatly benefit a library's operations as its use as a transactional predictive tool allows for the collection of data on a per-use basis, and therefore, it can provide a wealth of data for close analysis prior to reporting any success or failures" (Massis, 2012, p. 492). In the article Massis alludes to how the use of a predictive analytic can provide real time actionable information to library professionals that can help inform decision making on the fly rather than waiting until "all" the data has been collected. Massis provides some of the first insights into how a system like this would work in a library environment. While Massis provides some good foundations, the article was an overview and did not serve to provide practical applications of the technology or a discussion of some of the problems it could potentially solve. To understand the field of machine learning and predictive analytics, it is important to lay some foundations for the use of these terms. So, what is meant by the term machine learning and predictive analytics?

Machine learning is a branch of computer engineering that endeavors to provide computers the ability to learn independent of programming. As Arthur Samuel (1959) in the article "Some Studies in Machine Learning: using the game of checkers" defines early machine learning. The author writes, "…the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning" (Samuel, 1959, p. 535). Samuel in this early discussion of some of the first machine learning systems provides a good synopsis of the topic. Machine learning is, simply put, a branch of artificial intelligence that seeks to teach a computer system to learn from its base programming in order to execute commands not part of the base program. The machine through different types of algorithms learns from previous processes and provides non-preprogrammed outputs. There are a variety of ways to construct a learning machine. Different mathematical and programmatic models are used to accomplish different tasks. One of these methods is predictive analytics. Predicative analytics uses statistical modeling to provide potential outcomes.

As Eric Siegal's book, "Predictive Analytics: the power to predict who will click, buy, lie, or die" describes, "In commercial, industrial, and government applications – in the real-world usage of machine learning to predict – it's called something else, something that is the very topic of this book: Predictive Analytics (PA) – technology that learns from experience (data) to predict the future…" (Siegel, 2016, p. 8). In the book Siegal makes a very good argument about the power predictive analytics has in helping businesses and governments utilize the data they collect to solve real world problems in real time. Predictive analytics being only one branch of machine learning. The usefulness, however, with this type of machine learning lies in the power of large scale statistical analysis. Essentially the more an institution can gather data, the more the mathematical modeling of the probabilities of an outcome will conform to a standard bell curve distribution. It is at this point that prediction becomes very powerful. While this is a rudimentary explanation, it can serve as a touchstone later for the discussion concerning how to deploy predictive modeling in a university library environment. For now, the question remains, if one was interested in using a type of machine learning, how to begin?

In looking at industry for guidance, the question of where can a library begin with this type of technology comes to mind. James Taylor (2011) writing in Information Management magazine argues that the best place to start in thinking about how to apply predictive analytics is in the day-to-day decisions. These types of decisions as Taylor (2011) describes are transactional and operational. Taylor (2011) writes, "In my experience, the best place to start with predictive analytics is in your day-to-day operations. Operational decisions are about a single customer or transaction." (Taylor, 2011, online). Operational or transactional decisions happen tens of thousands of times a day in the academic library. Taylor (2011) concludes that predictive analytics works best at predicting three settings. First, you can use prior behavior to predict future risk. Second, you can use past interactions to predict future growth. Finally, you can use past fraud to predict future exposure to fraud. It is the second application of predictive analytics - using past interactions or data points in the form of transactions that occur within a library to predict future behaviors - that can serve most of the needs of an academic library.

Fern Halper (2011) in the same issue of Information Management magazine echoes the emerging needs and capabilities of predictive analytics. Halper (2011) identifies five trends for the emerging use pf predictive analytics. The first trend is the expansion of predictive analytics beyond the statistician to now include business intelligence analysts and other corporate entities. The second, expansion is the operationalizing of the models. What Halper (2011) is emphasizing is the movement of the models from data presentation to actionable data modeling that can be incorporated directly into a work process. Third is using unstructured data to support structured data. This is a combination of data that is free text like notes combined with numerical or structured data to provide a boost to predictive mechanism. Fourth is using big data. Big data can be understood as data that occurs in large volumes on a large scale. For example, the checking out of books over the lifetime of a library management system would produce a big data set that could be tapped for analysis. Fifth is the growth in open source predictive systems. The open source nature of machine learning allows libraries to leverage cutting edge technology at no cost. The open source technology is something that will be demonstrated later in the application phase of this paper. The literature above paints a very powerful picture for the use of predictive analytics as a method to solve real time problems in a library environment. The kinds of problems can be solved is discussed in the same issue of Information Management magazine by William McKnight.

McKnight (2011) illustrates the practical applications predictive analytics can have in the business environment. For McKnight (2011) predictive analytics can help inform customer lifetime value, clinical treatments, churn management (the loss of customers), segmenting the next best offer, and fraud detection. By taking these three commercial articles together, it is possible to begin to paint a picture for the role predictive analytics can play in the academic libraries. Predictive analytics can understand the transaction level exchanges and provide operationalized modeling of the customer experience with the library. Specifically, the customer satisfaction with the collection in the form of circulating books. The spread of predictive analytics beyond the business or private sector into academia has started.

In times of shrinking budgets, academia is looking for new tools that can provide better information about the processes they undertake. This has led to an increased interest in business analytics and specifically predictive analytics. Academia has begun to use predictive analytics to help solve practical problems.

In 2005, Philip Goldstein writing for the Educause Center for Applied Research argued, "Producing meaningful, accessible, and timely management information has long been the holy grail of higher