EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 7 UNIT 1
Video 3 Transcript

MIT BDA Module 7 Unit 1 Video 3 Transcript

Speaker key

XD: Xiaowen Dong

HY: Hapyak

XD: We have seen some applications of big data in industry so far, following the emergent trend of applying big data analysis in financial industry and marketing applications. We will discuss in this video how large-scale behavioral data can be used for assessing individuals' financial well-being and creditworthiness.

Traditional financial decision systems such as credit scoring system rely on individual traits such as age, gender, job type, and marital status, while being oblivious to the spatial-temporal mobility patterns and the habits of the individual involved. Emerging trends in geo-aware and mobile payment systems and the resulting big data present an opportunity to study human consumption patterns across time and space and assess individuals' financial well-being and creditworthiness.

On the one hand, in 2013, it is estimated that more than 3 billion credit cards are in use, accounting for US$2.2 trillion spending, while mobile payments are expected to exceed US$140 billion in the next five years. On the other hand, reports estimated that 39% of American household carry credit card debt from month to month. Our motivation is, therefore, to leverage the rich individual transaction data to study human purchase patterns for a more accurate and timely assessment of individuals' financial well-being and creditworthiness.

00:01:43

In this work, inspired by the biological phenomenon of foraging, a basic pattern of animal movement for food and resources, we study human purchase behavior through three behavioral traits. First, we look at how diverse is an individual's shopping pattern in terms of both shopping time and location measured by the Shannon entropy. This corresponds to the level of exploration in his or her behavior.

Second, we look at the customer's loyalty towards most frequent shopping time and locations, namely, the percentage of frequent shopping time and locations in all shopping activities. This captures the level of exploitation of known good timeslots and places.

Third, we look at the regularity in an individual's behavior. For example, whether he or she shows similar diversity and loyalty across different months. This captures the plasticity in his or her shopping behavior.

We connect these behavioral traits to an individual's financial well-being and creditworthiness, which are reflected in the following three aspects. Overspending, which means the person spends more money than his or her income; financial trouble, which means the person has defaulted or been identified by the bank with financial troubles; and late payment, which means the person has made late payment to his or her credit card balance.

00:03:12

HY: Which of the following traits do you think is the most important financial indicator?

A. Overspending
B. Financial trouble
C. Late payments

Thank You

XD: Let me explain in more detail how this works. We study 386,000 credit card transaction records for 10,000 customers in a period of three months, which are provided by a major financial institution in an OECD country. For each individual, we analyze his or her behavioral traits by computing diversity, loyalty, and regularity measures. We also obtained the financial trouble indicators for each individual. With both sets of information, we train the machine-learning model that predicts whether customers are likely to have financial troubles or not in future.

HY: Alex Pentland et al. paper on predicting individual financial well-being

XD: If you are interested in the technical details of our experiment, you can go and read this reference paper. We evaluate the performance of our predictive model using the so-called AUC metric, namely, area lying under a receiving operating characteristic or ROC curve, which is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied.

A comparison of the performance of three different models is shown here, including a baseline random guessing model, a model based on demographic features, and our model based on behavioral features. We see that the performance of our model is 30% better than demographic-based models in predicting individual financial troubles and late payment, and nearly 50% better in predicting overspending. We consider these results to be highly significant given that they can impact a trillion-dollar market segment of credit card limit decisions and credit repayment.

00:05:01

As a recap, in this video, we have shown an example on how large-scale behavioral data can be leveraged for assessing individuals' financial well-being and creditworthiness. In the next video, from another perspective, we will examine the possibility of utilizing such data to forecast the well-beings of merchants.

00:05:29