EXperimental
Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 7 UNIT 1
Video 4 Transcript

MIT BDA Module 7 Unit 1 Video 4 Transcript

Speaker key

XD: Xiaowen Dong

In the previous video, we showed an example of using credit card transaction data to predict individual credit worthiness. In this video, we will look at a similar problem but from a merchant perspective, namely, forecasting future financial well-beings of merchants. Our studies focused on small and medium-sized enterprises – in short, SMEs.

SMEs constitute a significant portion of economic activities in most countries. Report estimate that 20 million SMEs in Europe responsible for 67% of the jobs, and in 2012, SMEs accounted for about 75% of net jobs added to the US economy. From a bank or lending agency's perspective, an accurate model for forecasting future well-beings of SMEs is crucial for effective capital allocation. Models that can accurately identify top-performing SMEs will improve decision-making and help reduce financial loss.

00:01:18

On the one hand, traditional approaches for predicting an SME's well-beings rely on financial metrics such as earnings per asset or equity per asset. However, these metrics are usually difficult to acquire and tend to have limited predictive performance. On the other hand, the emerging trend of mobile payment using credit card presents an opportunity to analyze the financial well-beings of SMEs from a new perspective. In this work, we are thus interested in utilizing large-scale transaction data to build risk models for SMEs. Specifically, we are interested in the change in future revenue of an SME as a measure of its financial well-being.

Now, let us look at another case study. In this work, we study credit card transaction records at 1200 retail stores in a large metropolitan area in a OECD country for three months. We summarize our approach in the following flow chart. Instead of traditional approaches of using customers' demographic information or traditional financial metrics, we study the network of merchants based on customer's shopping behavior and the derived features from a network perspective. We then use these features to train machine-learning models to predict the future well-beings of merchants.

Specifically, we define our target variable as 1 if an SME has a change in revenue from the first two months to the third above median changes among all merchants, and 0 otherwise. In other words, if the model predicts a 1 for a specific merchant, that means the performance of this merchant is above average. On the other hand, if the model predicts a 0, then the performance of this merchant is below average. From a lender's perspective, the specified target variable drives decision-making since it provides insight into the future growth in revenue of an SME.

00:03:26

We will now talk in more details about the proposed network features. First, we construct a network where the vertices represent merchants. We draw an edge between merchant i and j if they share common customers in the two months considered, with the weight corresponding to the number of such common customers. This results in an undirected and weighted network.

Given the merchant network, we then compute structural properties of the network as we talked about in Module 4 of the class. For example, various notions of centrality measures on this network capture how far away a merchant is, on average, from every other merchant. We also apply community detection algorithm to group merchants into different clusters. Such structural information, in one sense, captures the importance of a given merchant, and could be indicative of the merchant's future well-being. We use this information as features for constructing our predictive model.

Similar to the case study in the previous video, we evaluate the performance of our predictive model using the AUC metric. The table shows the top five performing features in the three models under comparison, namely, models trained using baseline financial metrics, sociodemographic features, and network features. We see a performance of 0.71 AUC for the network features, which is a 22% improvement compared to the baseline financial model.

00:05:02

A model combining the entire top-ranked features results in a 0.79 AUC, which represents a even bigger improvement of 36% over the baseline financial model. These results demonstrate that features derived from a city-level merchant network, which captures the flow of the customers, improves the possibility to forecast future financial well-beings of these merchants. Our results also suggest that incorporating network-based signals and sociodemographic information into the current financial models will improve lending decisions.

As a recap, in the last two videos, we have shown two examples on how large-scale individual financial transaction data can be leveraged to forecast the well-beings of both individuals and merchants in the future, which provides insights into building better systems for credit assessment from a data-driven perspective.

00:06:07