Purchase patterns, socioeconomic status, and political inclination

Xiaowen Dong\textsuperscript{1,}, Eaman Jahani\textsuperscript{1,}, Alfredo Morales-Guzman\textsuperscript{1,2,}, Burçin Bozkaya\textsuperscript{3,}, Bruno Lepri\textsuperscript{1,4,}, and Alex ‘Sandy’ Pentland\textsuperscript{1}

\textsuperscript{1}MIT Media Lab. \textsuperscript{2}New England Complex Systems Institute. \textsuperscript{3}Sabancı University. \textsuperscript{4}Fondazione Bruno Kessler

Abstract

In this paper, we analyze millions of credit card transaction records during several months for tens of thousands of individuals from two different countries and show that, purchase patterns are strongly correlated with important societal indexes such as socioeconomic status and political inclination. Our results suggest the possibility of understanding and predicting the evolution of such societal indexes from purchase behavioral patterns, potentially at high temporal and spatial resolutions.

Keywords— Purchase patterns, socioeconomic status, political inclination, credit card transaction

1 Introduction

The emergence of new research fields such as computational social science \cite{4} has promoted data-driven approaches, through the unique lens of Big Data, to study and understand human and social behavior at unprecedented scale. In particular, several studies have proposed the use of large-scale cell phone data to study important economic questions such as social welfare \cite{3}, poverty mapping \cite{1} and unemployment prediction \cite{6}. In addition to cell phone data, the recent availability of large-scale financial transaction data provides even better opportunity to study human behavior and decision making in urban environment, and their implications in the organization of the society.

Human economic behavior has deep connections with sociopolitical processes \cite{5, 2}. In this paper, using large-scale individual credit card transaction data from two countries, one European and the other Latin American, we study human purchase behavioral patterns in urban environment as well as their sociopolitical implications. In particular, we focus on the pattern of purchase diversity, i.e., how diversely people visit different stores or districts in the city. We propose two diversity measures, “outgoing” and “incoming”, and test their correlations with important societal indexes, namely, socioeconomic status and political inclination at the census district level. We show that, in both countries, purchase diversity is strongly correlated with

\textsuperscript{*}indicates equal contribution. Corresponding author: Xiaowen Dong (xdong@mit.edu).
socioeconomic status of the district; furthermore, we find interesting links between purchase diversity and political inclination of the district, the interpretation of which however depends on the political context of the specific country.

Our results are one of the first to establish links between purchase behavioral pattern and important societal indexes, and test on large-scale transaction data from more than one country. They open new possibilities to understand, from only the purchase behavior, the evolution of socioeconomic status and political inclination of districts in a cost-effective way, and potentially at much higher temporal and spatial resolutions compared to census and election results. Our findings may also foster the prediction of these important measures in a timely fashion. Finally, the proposed “incoming” diversity may help establish the potential links between economic mixing, income segregation and political polarization, which is certainly an interesting direction for future studies.

2 Data and methods

We analyze credit card transaction data provided by two major financial institutions in two low-to-middle income countries, one European and the other Latin American. The European data contain more than 10 million credit card transactions from more than 100 thousand individuals at more than 100 thousand merchants during a period of three months. Similarly, the Latin American data contain about 60 million transactions from about 3 million customers at about 400 thousand stores. Each credit card transaction record comes with metadata such as the time and amount of the transaction, as well as store location and the customer’s home location. Particularly, the geographical information allows us to investigate the spatial distribution of financial transactions and the economic links between various census districts. The data are analyzed under legal restriction against reidentification, which fully conforms to the privacy laws of the countries.

In this paper, we analyze patterns of purchase activity and their socioeconomic and political implications. In particular, we establish links between purchase diversity of census districts and their socioeconomic status and political inclination. We define two measures of purchase diversity:

1. Outgoing purchase diversity: This captures the district-level exploration, namely, how diverse are the stores explored by the individuals who live in a given district.

2. Incoming purchase diversity: This measures the district-level attractiveness, namely, how diverse are the districts from where individual customers are attracted to come and make financial transactions in a given district.

For each individual, we define the outgoing purchase diversity as the Shannon entropy of his/her purchase activities:

\[ H(i) = - \sum_{j=1}^{N} p_{ij} \log(p_{ij}), \]

where \( p_{ij} \) is the probability that an individual \( i \) visits a merchant \( j \) within three months and \( N \) is the total number of merchants. The diversity of a district is then defined as the average
diversity of individuals living in that district. Similarly, we compute the incoming purchase
diversity of a given district based on the set of districts from where individual visitors come. It
is worth noting that our approach is similar to the network-based approach of [3]; the difference,
however, is that the network in our case is essentially a bipartite graph where customers and
stores form two sets of vertices.

We study the relationship between these diversity measures and two important indexes at
the district level. The first is the socioeconomic status of each district, which in the case of
the European country is a composite measure between 0 and 100 that quantifies the relative
prosperity of the district based on a number of indicators such as education, income and housing
conditions. The higher the index, the more prosperous the district is. For the Latin American
country, the measure is defined similarly but as a categorical variable with 5 different levels.
The second is the political conservatism of each district, which is an index between 0 and 100
that is computed based on the percentages of votes parties labeled as “liberal” or “conservative”
obtained in a recent national election. The higher the index, the more conservative the district
tends to be.

3 Results and Discussion

In Table 1, we show the Pearson product-moment correlation coefficients between the two pur-
chase diversities and the socioeconomic/political indexes of about 1000 districts in both countries.
In both countries, purchase diversity measures are positively correlated with the socioeconomic
status of the district, with a particularly strong relationship \( r = 0.77 \) in the case of outgoing
diversity in the European country. This indicates that, in general, the higher the diversity, the
more prosperous the district tends to be. In addition, for the European country, there exists
a fairly strong negative correlation \( r = -0.55 \) between outgoing purchase diversity and political
conservatism, namely, the higher the diversity, the more conservative the district tends to be.
Interestingly, such a negative relationship is reversed in the case of the Latin American country
\( r = 0.26 \) for outgoing and \( r = 0.33 \) for incoming) which may be due to the different economic
characteristics that a label such as political “conservatism” might have in two different contexts.
Fig. 1 illustrates both diversity measures, socioeconomic status, and political conservatism, for
districts in the largest city of the Latin American country. The visual patterns are consistent
with the quantitative results in Table 1.

The positive correlation between district-level purchase diversity and socioeconomic status
is consistent with the finding in [3], where the authors have shown that the diversity of phone
communication of local districts is strongly correlated with the socioeconomic index of these
districts. This confirms the link between explorative behavior and regional prosperity. On
the other hand, the difference between the two countries, in terms of the relationship between
purchase diversity and political conservatism, highlights the different meanings conservatism
represents in different political contexts. Nevertheless, the presence of these correlations, either
positive or negative, suggests that different purchase patterns indicate a separation between the
districts in terms of their political views.

1 In case of the Latin American country, the outgoing diversity is computed directly based on district-level
information.
Table 1: Pearson’s $r$ between purchase diversities (both outgoing and incoming) and indexes about socioeconomic status and political conservatism. All correlations are statistically significant with $p$-value < 0.001.

<table>
<thead>
<tr>
<th></th>
<th>Socioeconomic Status</th>
<th></th>
<th>Political Conservatism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European: Purchase</td>
<td>0.77</td>
<td>-0.55</td>
<td></td>
</tr>
<tr>
<td>European: Purchase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversity (outgoing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European: Purchase</td>
<td>0.35</td>
<td>-0.25</td>
<td></td>
</tr>
<tr>
<td>European: Purchase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversity (incoming)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latin Amer.: Purchase Diversity (outgoing)</td>
<td>0.53</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Latin Amer.: Purchase Diversity (incoming)</td>
<td>0.49</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

Although the correlation-based analysis does not permit causal inference, our results show that purchase patterns at least serve as strong statistical indicators of the socioeconomic status and political view of the districts. This would potentially enable the understanding and prediction of the evolution of such important societal indexes, from only purchase behavioral patterns, in a cost-effective and timely fashion.

Finally, both diversity measures, and in particular the incoming purchase diversity, capture the richness of social and economic information that is introduced to each district. We believe that they have attenuating effect on any extreme economical and political behavior. We hypothesize that districts (or cities) which attract economic activity from various regions tend to be less politically extreme, show high voting diversity, and experience lower economic segregation. We plan to perform such analysis as future work.

References


Figure 1: Purchase diversities (both outgoing and incoming), socioeconomic status, and political conservatism, for districts in the largest city of the Latin American country. The darker the color, the higher the index value associated with a district.