MobiScore: Towards Universal Credit Scoring from Mobile Phone Data

Jose San Pedro[1(✉)], Davide Proserpio[2], and Nuria Oliver[1]

[1] Telefonica Research, Barcelona, Spain
{jose.sanpedro,nuria.oliver}@telefonica.com
[2] Computer Science Department, Boston University, Boston, MA, USA
dproserp@bu.edu

Abstract. Credit is a widely used tool to finance personal and corporate projects. The risk of default has motivated lenders to use a credit scoring system, which helps them make more efficient decisions about whom to extend credit. Credit scores serve as a financial user model, and have been traditionally computed from the user's past financial history. As a result, people without any prior financial history might be excluded from the credit system. In this paper we present MobiScore, an approach to build a model of the user's financial risk from mobile phone usage data, which previous work has shown to convey information about *e.g.* personality and socioeconomic status. MobiScore could replace traditional credit scores when no financial history is available, providing credit access to currently excluded population sectors, or be used as a complementary source of information to improve traditional finance-based scores. We validate the proposed approach using real data from a telecommunications operator and a financial institution in a Latin American country, resulting in an accurate model of default comparable to traditional credit scoring techniques.

1 Introduction

Credit scores are widely used in the financial sector as a way to control risk when lending to potential credit customers. Credit scores convey information about the ability of customers to pay back their debt or conversely to default. Therefore, they are a key variable to build financial user models. Traditionally, credit scores have been computed by modeling the past financial history of users. Credit bureaus, such as Equifax or Experian, provide these scores to interested parties by aggregating all the available financial information from customers across different financial products and institutions.

Scores generated using the past financial history have shown to be very reliable. Hence, lenders prefer to focus on cross-selling to existing customers or catering to those for whom credit history information is more readily accessible [13]. As a result, financially responsible customers could be excluded from credit because they lack any prior financial history. These customers are referred to as *no-file* or *thin-file*.

F. Ricci et al. (Eds.): UMAP 2015, LNCS 9146, pp. 195–207, 2015.
DOI: 10.1007/978-3-319-20267-9_16

While *no-file* or *thin-file* consumers are prevalent in developing economies – the World Bank estimates that there are 2.5 billion *unbanked* adults who lack access to formal financial services[1] – millions of consumers today don't receive credit scores in developed economies either. New immigrants (only in the U.S. there are 1.1 million working immigrants per year), recent college graduates, people recovering from earlier financial missteps and the *underbanked* – individuals who have a bank account but do not use it on a regular basis, accounting for 20% of the U.S. population in 2013 [2] – are also in the no-file category.

The growing need to access credit has motivated the search of alternative sources of information that could serve as accurate proxies of traditional credit scores to model user proneness to credit default. In this direction, a large body of literature has studied demographic aspects and variables related to personality and socioeconomic status that correlate with unreliable financial behavior and lead to default. Among these characteristics we find impulsiveness, education level or marital status [10,11].

In this paper we propose MobiScore, an approach that leverages data passively captured from the mobile phone network, *i.e.*, patterns of mobile phone usage, to build a user model of propensity to credit default by using supervised machine learning methods. This concept is supported by the results of previous studies that leverage mobile phone usage data for modeling users, such as inferring personality traits or socioeconomic status [5,6,17], which have been shown to correlate with financial behavior [9]. Given the ubiquity of mobile phones both in developing countries and developed economies, MobiScore could help currently excluded users in getting access to financial services such as loans or credit cards.

The main contributions of this paper are twofold. First, we introduce MobiScore, a methodology to build a consumer credit default model using as input data call detail records obtained from telecommunications companies. Second, to asses the feasibility of the model, we conduct a large study with real credit default data for a large population of over 60,000 customers from a Latin American country. This study benchmarks the performance of a traditional credit scoring approach along with our proposed methodology. To the best of our knowledge, this is the first work that reports results on such a large scale population.

2 Related Work

Credit scores have been extensively used in the financial industry during the last two decades to improve customer credit collections [13]. However, research reporting the performance of this tools is scarce due to their proprietary nature. In this work, we focus on the particular problem of customer credit scores, namely to predict individual risk. In this area, credit scoring techniques are based

1 http://www.worldbank.org/en/results/2013/04/02/financial-inclusion-helping-countries-meet-the-needs-of-the-underbanked-and-underserved

2 https://www.fdic.gov/householdsurvey/

on statistical models that use different data sources that reflect user behavior, financial or not, to determine their likelihood to default [1,3].

The most widespread way to compute credit scores uses financial and demographic data. Credit bureaus, including Equifax and Experian, or specialized technological companies like Fico, rely on specific financial information, such as credit history, current credit use, or ratio between credit limit and outstanding balance. As previously discussed, these methods need access to the financial history of users, which is not always available. Hence, alternative methods to score users that do not rely on such information have been proposed in the last few years.

There is a vast amount of literature trying to understand the socio-economic and behavioral factors that can be correlated with financial risk and spending behavior, as well as with credit default. Socio-economic factors including education, marital status, net worth, financial knowledge, and household income, as well as several environmental factors (*e.g.* income or home ownership status) and demographic factors (*e.g.* racial background, age, and gender) are all correlated with financial risk [10]. Furthermore, it has been shown that personality affects the predisposition to financial risk. For example in [11] the authors show how impulsiveness is correlated with credit card behavior, while in [12] the authors show that impatient people are more prone to default. Finally [2] shows how user mobility is (inversely) correlated to default rates.

Mobile phone usage logs have been studied for modeling users and community dynamics in a wide range of applications. For example, a framework to discover places-of-interest from multimodal mobile phone data is described in [14], and [7] shows that it is possible to accurately infer friendships based on observational mobile data alone. Socio-economic status has also been inferred from mobile phone activity data [17]. More broadly mobile data has been used to uncover individual and collective human dynamics [4], infer gender [8] and personality [5, 6]. In [16], the authors study the interconnection between social and mobile features and spending behavior. Our work relates with all the above in that we model user behavior from phone usage, and in particular the user's propensity to default on their credit.

The use of mobile data for financial risk assessment has attracted the attention of some entrepenurial efforts, such as Cignifi[3]. However, a comparison is difficult as they use a proprietary technology and the details of their approach are unknown. To the best of our knowledge, this paper is the first to disclose the potential of mobile phone usage data in this problem setting.

3 Dataset

In this section, we describe the dataset used in MobiScore to model the users' financial reliability from their mobile phone usage data. This dataset combines two main sources of information: (1) mobile phone usage logs, which we use as

[3] http://www.cignifi.com/en-us/technology

raw data to find behavioral patterns consistent with unreliable financial behavior; and (2) financial information in the form of credit default reports, which serves as ground truth to train user models of credit default using a supervised learning paradigm. We obtained the data for over $60,000$ people from a Latin American country in which both the telecommunications company and the financial institution operate. The data was completely anonymzed and stripped of all personal information, such that it was not possible to identify any individual. The Telco and Financial institution generated a unique user identifier applying a common hash function unknown to us. This enabled us to anonymously merge the two datasets. Moreover, the analysis of the data was in compliance with the terms and confitions of data usage of that particular service.

3.1 Mobile Phone Usage Dataset

Mobile phone networks are built using a set of cell towers (BTS) which connect phone devices within the network. Each BTS is identified by the latitude and longitude of its geographical location. Whenever a phone in the network makes or receives a call or uses a service (*e.g.* SMS, MMS), this communication event is logged into a raw database known as Call Detail Records (CDRs). CDRs include information about all the different aspects of the communication event: phone lines involved in the event, type of event, timestamp, etc. BTS details are also logged, providing a proxy of the geographical position of the user at the time of the call (at the granularity of the area of coverage of the BTS).

From all the information contained in a CDR, MobiScore analyzes the two most common events: calls and text messages (SMS). We extract the following fields from the CDRs: anonymized originating and destination numbers, time and date of the call or SMS, and an identifier of the BTS that the originating phone was connected to when the call was placed. For calls, we also extract their duration in seconds. Note that the dataset neither contains personally identifiable information nor the content of the calls or the SMS.

Our CDR dataset contains three months of data from January to March 2014, consisting of daily cell phone events from pre-paid and contract subscribers in the Latin American country of study. The collected dataset contains over 35 million call events, and over 11 million SMS events, with an average of around 12 million calls and 3.5 million text messages per month.

3.2 Financial Dataset

Lenders normally report payments in arrears to credit bureaus on a monthly basis. Two payments in arrears, *i.e.* 30 days past due, is the first level reported to bureaus, as dictated by the Fair Credit Reporting Act.[4] We collected such records for a credit card offered by a financial institution operating in the considered Latin American country. The dataset of customers in arrears was collected for a period of 15 months, from January 2013 to March 2014. For each customer we

[4] http://www.ftc.gov/os/statutes/fcradoc.pdf

obtained a monthly record stating the amount of pending balance and the days in arrears of payment.

Customers are considered to be in default only when the lender deems their pending balance to be uncollectible; the exact number of days for this condition to hold depends on the financial product, the lending institution and the legal framework of the country where the institution operates. In the context of this paper we consider two definitions of customers in default that are common in the literature: customers who are more than 30 days in arrears (*Default @30*), and customers who are more than 90 days in arrears (*Default @90*). Using the aforementioned reports, we could determine whether each customer was in default for any of the two definitions used.

4 Methodology

We pose the credit scoring problem as a binary classification task, which we model using a supervised learning strategy. For each potential credit customer, we intend to model a score P which can be interpreted as the probability of default. Customers are represented in a feature space that captures different orthogonal aspects of their use of the phone, which we expect to convey behavioral information with discriminative power in this problem setting.

4.1 Feature Extraction

In Sec. 2, we introduced previous research focused on identifying correlations between financial risk and user socioeconomic status, personality and demographic factors among others. We follow this line of work to extract from the CDRs a set of features of mobile phone usage that convey user behavioral information related to financial risk. Furthermore, we take advantage of additional information collected by the telecommunications company about their customers, known as Customer Relationship Management (CRM), to extract demographics and socioeconomic features.

CDR Features. CDRs amass all the communication actions carried out by customers, and therefore has great potential as a source of information for user modeling. We are especially interested in studying traits of personality and specific behaviors of stationary nature, as to discern which of those could lead to credit default with higher probability. Previous literature has found such correlation in subjects with impulsive behavior or with more impatient personalities. In our context, we do not count with such high level information about users, but previous work on analyzing phone logs encourages us to think that we can find such correlations also in features extracted from CDRs. In order to be as comprehensive as possible, we study CDR logs across three complementary dimensions: consumption, mobility and social network. A detailed list of the CDR-based features is shown in Table 1.

Consumption features are related to the amount of use of the communications network, and provide a high level view of the access frequency for different communication methods, both calls and SMS. We compute the above features for different temporal partitions: by day of the week (weekdays *vs* weekends), and by time of the day (office hours, 9am to 6pm, *vs* evenings). This slicing of the temporal dimension in four partitions (or time windows) allows to profile users at different significant moments, helping to generate a more holistic view of their phone usage. Furthermore, since mobile communications are intrinsically directional, some features are subdivided in incoming (received by the user) or outgoing (originated by the user).

?Social network features focus on capturing information about the characteristics of the graph of connections between customers, which could convey information about empathy and other social-related traits of personality. We include in these features the number of unique correspondents in incoming and outgoing calls/SMS (i.e. in and out degrees) and the in/out degree variation between between temporal windows as defined above (delta degrees).

Finally, *mobility features* capture mobility patterns of customers in their everyday life inferred from the position of the BTS the users connected to. Radius of gyration is computed as the minimum radius that encompasses all the locations (BTS) visited by a user daily. Distance traveled adds the distance traveled to visit all the BTS connected every day. Popular antennas is computed as the number of BTS connected to account for 66% of the communications activity, serving also as a proxy of mobility.

We also include the number of events that are reciprocated. An event is marked as reciprocated if the current customer A calls (or texts) another customer B within one hour from an incoming call (text) from customer B. Reciprocated events such as calls or SMSs are used in the mobile data analysis literature as an evidence of the existence of strong ties between users [15]. While a single call between two individuals may not carry much information, reciprocated calls between two users indicate some degree of relationship (work, family, or other), helping us to better characterize the user's social interactions.

Moreover, we extract regularity features by computing Shannon's entropy for variables related to calls, messaging, and mobility. *Communication time entropy* measures the regularity of customers regarding the time window used to contact individual interlocutors. We consider the multinomial distribution of events over the temporal dimension for each customer and unique interlocutor. That is, for every customer i and interlocutor j, their *communications time entropy* is:

$$H_{ij}(X) = -\sum_{x \in V} p(x) log(p(x)) \tag{1}$$

where X is a discrete random variable taking values from the set V (every one of the four temporal partitions), and $p(x)$ is its probability mass function (the fraction of times user i contacted user j in the given time window x). The final entropy is computed as the mean value $H_{ij}(X)$ for all interlocutors of i.

Communication entropy measures how regular a user is with respect to the interlocutors (s)he has contacted. Using the formulation of Eq 1, V would refer to the complete set of interlocutors of i, and $p(X)$ to the probability that i calls/texts this particular interlocutor x (relative frequency *w.r.t.* the rest of interlocutors). Similarly, *mobility entropy* is computed by considering the fraction of calls made from every BTS in the set of popular antennas.[5]

Customer Relation Management (CRM) Features. In addition to usage logs, telecommunication companies collect additional information related to their customers, typically stored in Customer Relation Management or CRM logs. For instance, demographic information (age, gender) is required for signing the contract of specific products. Socioeconomic information is inferred from the customers home address, required for landlines. Additional data related to socioeconomic status can be extracted from the type and number of products owned by the customer, as well as data related to the devices used (*e.g.* brand of phone). We collected some of this CRM information and use it both as a standalone dataset to predict financial risk and to complement the CDR features. A detailed list of the CRM features used is presented in Table 2.

Table 1. Event log features

Consumption features
By time window and direction
Daily call (SMS) events
Daily duration of call events
Daily time between consecutive call (SMS) events
Daily time between consecutive events (either call or SMS)
Global
Communications time entropy
Communications entropy
Social network features
Number of unique call (SMS) correspondents
Call (SMS) delta degrees
Number of reciprocated call (SMS) events
Fraction of reciprocated call (SMS) events
Median of time between reciprocated call (SMS) events
Mobility features
Radius of gyration
Distance traveled
Popular antennas
Popular antennas entropy

Table 2. CRM Features

Product features
Device brand
Device operating system
Device type
Line type
Line status
Line quantity
Late payments
Month elapsed since activation
Socioeconomic features
Age
Gender
Estimated customer income
High risk ZIP code
Regional area code

4.2 Model Building

We use supervised learning for building user credit score models. To this end, we represent users in the presented feature space, which we extract from the CDRs in the period January 2014 to March 2014. Our financial information for this population predates these CDRs for the most part, as it goes from January 2013 to March 2014. As we are interested in stationary personality traits and behavior, we use the following methodology to generate the ground truth.

[5] SMS were not associated with a BTS in our dataset.

We consider only customers that activated their card during the period January 2013 to March 2014 (our range of financial reports). Given that the probability of default at a time greater than T is $P(t > T) = 1$ for a sufficiently large T, we decided to consider the same period of maturity for all customers. That is, we study the credit default during the first M months since customers activate their card. We use two different values for M depending on the default definition. For the *Default @30* scenario, we assign $M = 7$. That is, we observe the customers' default history up to the beginning of their seventh month of card use, resulting in six full months of observations. For *Default @90*, we assign $M = 9$, which gives users six full months to default so that their debt at 90 days is observable at the beginning of the ninth month. In both scenarios, we label users that defaulted at any point during the first M months as positive samples, disregarding later payments that would have cleared their default status.

To give an example, customers that activated their card during March 2013 (Month 0) are observed during the next six months, April to September, and their default status is obtained from the financial reports in October (Month 7) for the *Default @30* scenario. In the *Default @90*, the same procedure is followed, but the observation period goes from April to November, and the default status is obtained from the financial reports in December (Month 9). Analogously, customers that activated their card during April 2013 will have their default evaluated in November or January, depending on the default definition.

Observing default status after a fixed interval since the credit is granted is a common methodology used by financial institutions to measure the performance of their scores. We fixed the observation interval at 6 months as it provides the best trade-off between the number of subjects and the interval duration given the characteristics of our dataset. Using this setup, we were able to train the model using all the customers that activated their card by August 2013 (*Default @30*) or June 2013 (*Default @90*). Although the CDR data used for part of the population post-dates the used ground truth, we believe that the stationary nature of the behavioral and personality features sought by our method should not experience noticeable variations.

We tried three different classification methods to compare their performance in this specific problem setting: L2-regularized logistic regression (LR), linear Support Vector Machines (SVM), and Gradient Boosted Trees (GBT). We tune the hyper-parameters automatically using a grid-search strategy over a validation set extracted from the training collection. We use fivefold cross validation to predict the customers in default for the whole dataset, where the evaluated fold in each step is not used during the training phase at any point.

5 Experimental Results

In this section we report the experimental results obtained by MobiScore for the credit card default inference task. We considered the following different configurations of the input data and ground truth:

Feature sets. We use three different feature representations of mobile phone behavior: features extracted from **CDR** only, from **CRM** only, and from both (**CDR+CRM**). The goal of this partition is twofold. First we want to determine the financial risk modeling power of the different data sources considered. Second, we want to understand the performance of CDR features, which capture behavioral information, when compared with the higher level socioeconomic and demographic CRM features. To the best of our knowledge, our analysis is the first study to assess the value of CDRs –particularly when compared to CRM information– to build a model of the user's financial risk.

Time dimension. We use three different time windows of CDRs: two weeks, one month and three months. The goal is to study the effect in classification performance of observing mobile data for longer time periods. We consider CRM features to be invariant during the period considered, so this dimension has no effect on them. Two weeks is the minimum amount of data needed to obtain meaningful results given the user activity levels in our CDR data. Conversely, three months is the maximum amount of data that was available in this dataset.

Default definition. We use two default definitions which are common in the financial domain, as described in Sec. 3: *Default @30*, and *Default @90*. Default @30 includes all customers with six or more months of credit card use (only the first six are used to generate their ground truth). This subset had $\simeq 55,000$ users, with a positive sample rate of 19.6%. Default @90 considers customers with nine or more months of credit card use. This subset of the population accounted for $\simeq 30,000$ users, with a lower 12.7% rate of positive samples.

Given the unbalanced nature of the ground truth, we used the following two metrics to evaluate classification performance: *Average Precision*, which refers to the area under the precision-recall curve, and *AUCROC*, which refers to the area under the Receiver Operating Characteristic curve and provides information about the ability of the trained models to rank users according to their probability of default. Ranking customers allows to cluster them in different risk groups (*e.g.* low risk, medium risk, high risk) and make different decisions for each group. For example, lower interest rates could be applied to the low risk group given their increased financial reliability. For this reason, AUCROC is a commonly used performance metric in the credit scoring literature [18].

We also had access to the credit scores provided by a major Credit Bureau agency for our population. Credit Bureaus use proprietary methods and data sources to compute credit scores, which are unknown to us. These scores are commercially available, and were provided to us by the financial partner in this study. We used this information as a baseline to benchmark the performance of the proposed approach given that this is the information used today as the state-of-the-art.

Table 3. Classification performance (AUCROC) as a function of the CDR time window considered

		CDR			CDR+SMS		
		2 weeks	1 month	3 months	2 weeks	1 month	3 months
@30	GBT	63.0	64.5	**67.5**	68.5	69.4	**71.6**
	LR	62.2	64.0	66.4	67.6	68.8	70.7
	SVM	62.1	63.9	66.7	67.9	68.8	70.6
@90	GBT	63.1	64.4	**67.5**	70.2	70.8	**72.5**
	LR	62.4	64.5	67.4	68.7	70.5	72.1
	SVM	63.1	64.1	67.2	69.7	70.3	72.1

In Table 3 we present the classification results for the AUCROC metric, focusing on the two CDR-based feature representations. First, we observe that longer observation periods are consistent with better performance. Our dataset contains a monthly average number of $\simeq 200$ calls made/received and $\simeq 65$ SMS messages sent/received per customer. When using longer periods of CDRs we get access to a larger amount of user activity, fostering the discovery of behavioral patterns and hence yielding better performance. Second, the *Default @90* scenario consistently achieves better performance than *Default @30*; the latter corresponds to a riskier financial profile that is expected to be consistent with behavioral traits that would be more prominently captured in the CDRs. Finally, we observe that the GBT classification approach consistently outperforms both logistic regression and SVMs in the given classification task. This is explained by the higher degree of complexity and flexibility of boosted trees ensembles, which produce non-linear models composed of thousands of weak predictors. Ensembles have been shown to outperform single models in many contexts, including financial modeling problems [18].

We also study the inference power of each feature representation. In particular, we compare the performance of the CDR-based feature representations to two static baselines: CRM-features, and Bureau data. For this comparison, we select the best performing configuration of CDR-based models (GBT classifier using 3 months of CDR data). The results of this comparison for the average precision and AUCROC metrics are shown in Fig. 1.

We observe that CDR-based features provide the highest discriminative power. Interestingly, **CDR** (alone) and **CRM** are similarly discriminative. This result suggests that the aspects of human behavior that can be inferred from mobile phone data (*i.e.* mobility, consumption and social variables) serve to model financial risk with similar performance as the socioeconomic and demographic variables included in the CRM. It is also worth mentioning that the combination of CDR and CRM increases performance noticeably for all metrics considered, which suggests that the information conveyed by these two sources of data is complementary. Moreover, all the scoring models built from mobile phone data outperform the Credit Bureau score by a large margin, supporting the validity of MobiScore as an alternative to traditional credit scoring approaches.

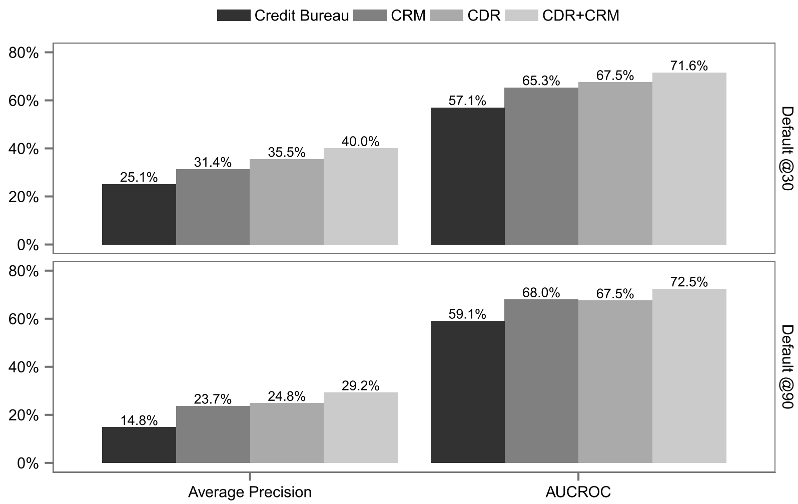

Fig. 1. Classification performance (Average Precision and AUCROC) of CDR-based representations compared to CRM and Credit Bureau baselines

6 Conclusions and Future Work

In this paper we have presented MobiScore, a novel approach to compute credit scores from mobile phone activity data. Standard approaches use prevalently past financial history, and therefore cannot be applied to users with no (or limited) financial history. We carried out experiments with a large-scale population and show that MobiScore achieves better performance when compared to current techniques used by Credit Bureaus. Moreover, even a limited time window of two weeks of mobile phone data can be enough to generate a reliable score. Overall, these results provide strong evidence that mobile phone data is a reliable signal of a person's financial risk.

MobiScore has the main advantage of leveraging passively collected data from the mobile network infrastructure and having wider applicability than traditional credit scores since it only requires the use of mobile phone, which most of the world population has access to. The mobile phone has become such an integral part of our lives that behavioral data captured with it adds informative value to a variety of applications, included the one proposed in this paper. The most important implication of the presented study is the opportunity MobiScore brings to enable access to credits and loans to the millions of *thin-file* individuals.

In the future, we plan to investigate in detail the discriminative power of the different features used in our experiments. Further, we are interested in assessing how our model changes as a function of the geographical areas and financial products analyzed. An additional interesting aspect, common to any project dealing with human behavioral data, is addressing any potential privacy concerns and data protection laws compliance. A straightforward way in which our model can be implemented and provided to users is through an opt-in mechanism, where the user gives to the mobile operator explicit consent to use mobile data to generate and provide the credit score. We plan to experimentally evaluate the attractiveness of such a system in future research.

References

1. Abdou, H.A., Pointon, J.: Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intelligent Systems in Accounting, Finance and Management **18**(2–3), 59–88 (2011)
2. Agarwal, S., Chomsisengphet, S., Liu, C.: Consumer bankruptcy and default: The role of individual social capital. Journal of Economic Psychology **32**(4), 632–650 (2011)
3. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J.: Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society **54**(6), 627–635 (2003)
4. Candia, J., González, M.C., Wang, P., Schoenharl, T., Madey, G., Barabási, A.-L.: Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical **41**(22), 224015 (2008)
5. de Montjoye, Y.-A., Quoidbach, J., Robic, F., Pentland, A.S.: Predicting personality using novel mobile phone-based metrics. In: Greenberg, A.M., Kennedy, W.G., Bos, N.D. (eds.) SBP 2013. LNCS, vol. 7812, pp. 48–55. Springer, Heidelberg (2013)
6. de Oliveira, R., Karatzoglou, A., Concejero Cerezo, P., Armenta Lopez de Vicuña, A., Oliver, N.: Towards a psychographic user model from mobile phone usage. In: CHI 2011 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2011, New York, NY, USA, pp. 2191–2196. ACM (2011)
7. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences **106**(36), 15274–15278 (2009)
8. Frias-Martinez, V., Frias-Martinez, E., Oliver, N.: A gender-centric analysis of calling behavior in a developing economy using call detail records. In: Proceedings of the AAAI Spring Symposium: Artificial Intelligence for Development (2010)
9. Gathergood, J.: Self-control, financial literacy and consumer over-indebtedness. Journal of Economic Psychology **33**(3), 590–602 (2012)
10. Grable, J.E., Joo, S.-H.: Environmental and biophysical factors associated with financial risk tolerance. Journal of Financial Counseling and Planning **15**(1) (2004)
11. Henegar, J.M., Archuleta, K., Grable, J., Britt, S., Anderson, N., Dale, A.: Credit card behavior as a function of impulsivity and mothers socialization factors. Journal of Financial Counseling and Planning **24**(2), 37–49 (2013)
12. Meier, S., Sprenger, C.: Impatience and credit behavior: Evidence from a field experiment. Social Science Research Network Working Paper Series (2007)
13. Mester, L.J.: Whats the point of credit scoring? Business review **3**, 3–16 (1997)
14. Montoliu, R., Gatica-Perez, D.: Discovering human places of interest from multimodal mobile phone data. In: Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, p. 12. ACM (2010)
15. Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L.: Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences **104**(18), 7332–7336 (2007)

16. Singh, V.K., Freeman, L., Lepri, B., Pentland, A.S.: Predicting spending behavior using socio-mobile features. In: 2013 International Conference on Social Computing (SocialCom), pp. 174–179. IEEE (2013)
17. Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E.: Prediction of socioeconomic levels using cell phone records. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 377–388. Springer, Heidelberg (2011)
18. Wang, G., Hao, J., Ma, J., Jiang, H.: A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications **38**(1), 223–230 (2011)