CONSUMER LENDING USING SOCIAL MEDIA DATA

Davis Bundi Ntwiga[1] and Patrick Weke

School of Mathematics, University of Nairobi, Po Box 30197-00100, Nairobi, Kenya

[1]Correspondence author: dntwiga@gmail.com

ABSTRACT

Consumer credit has been around for a long period of time but the dynamics observable from the consumers makes it hard to credit score and lend to the consumers. This difficulty results in the poor being excluded from receiving credit as they lack financial history. We analyze the limitations of the traditional consumer lending models due to use of historical data, and look at the benefits that could arise by incorporating social media data in credit scoring process for consumer lending. A review of the research progress made in using social media data for consumer scoring and lending process is presented. We found that social media data offers rich, vast and attractive information on changing trends and shifting demographics in credit underwriting of existing consumers and new consumers with minimal or no financial history. This data advances the lending process by widening the data set available and capture of new markets that are excluded from financial services

KEYWORDS: Social media data, consumer lending, credit scoring, credit risk

1. INTRODUCTION

Consumer credit dates back around 3,000 years ago since the time of the Babylonians. In the last 70 years, consumer credit moved to mass market contributing to increasing availability and widespread of consumer credit in modern society. This has increased the rate of personal bankruptcy due to default on credit repayment granted to finance and purchase commodities and services; financing a car, a home, appliances, buying furniture, traveling and others are indicated as for personal development or the instant loans (Horkko, 2010). Credit is granted based on a credit score. Scoring is a process that uses recorded information about individuals and their loan requests to predict their future performance regarding credit repayment (Robert, Raphael, Paul and Glenn, 1996). Lenders use credit scores to differentiate consumers on risk categories, to facilitate strategic planning decisions, and offer credit facilities to their customers (Khanbabaei and Alborz, 2013).

Therefore, modeling credit risk has become a mandatory tool for risk management as the dynamics on consumer lending continues to change. The dynamics in consumer financial market calls for innovative use of advanced analytics and big data from social networks. Big data aids to loop in; households and young consumers who do not interact with formal financial institutions; and to gain more insights in the demographic changes of the active borrowers (Daniel and Grissen, 2015). One source of the big data is from the social networks, which can provide more insights on how social links evolve over time, how this affects people and the interaction choices made (Ntwiga, 2016; Raghavan, Steeg, Galstyan and Tartakovsky, 2013).

The social networks provide the social media data which refers to the wide range of internet based and mobile services data. These channels allow users to participate in online exchanges, join online communities or contribute user created content. Internet services commonly associated with social media include but not limited to the following: blogs, wikis, social bookmarking, social networking sites, status update services, and virtual world content and media-sharing sites, among others (Dewing, 2012). The social media data contained in blogs, online forums, discussion groups, etc, can easily be gathered, analyzed and processed (Siva, 2010) to support decisions about the consumer in the credit underwriting process by improving the scoring process (Ntwiga, 2016).

In this review paper, we contribute by taking a fresh approach to changing credit trends. We consider some limitations of current credit risk models for consumer lending. In addition, look at the promising prospects or benefits in using the social media data in consumer lending with a view of increasing financial inclusion and loop in those marginalized from the formal financial services. A summary of some of academic research work in credit scoring using social media or social network data is highlighted.

Section two analyzes the various limitations of the traditional consumer credit risk modeling. Section three considers some of the reasons and factors that that make the ground of using social media a fertile and innovative approach in consumer lending. Section four reviews some of the progress made in this area and other factors key to its success. Section five offers concluding remarks and areas of future research in consumer lending aided by social media data.

2. CURRENT CREDIT RISK MODELS

In credit lending, a credit rating system is expected to accurately assess the credit risk of the obligors (Ntwiga, 2016). Banks and other financial institutions are applying increasingly sophisticated methods and different techniques to assess the risk in their loan portfolios. The methods can be classified into three groups; statistical models; artificial intelligence models and hybrid models being a combination of either of the former two methods (Ntwiga, 2016; Thomas, Oliver and Hand, 2005). Table 1 shows the classification of the three groups of credit scoring methods as outlined in (Khanbabaei and Alborz, 2013; Li and Zhong, 2012).

Table 1: Classification of Credit scoring methods

Statistical Models	Artificial Intelligence models	Hybrid models
Linear Discriminate analysis	Artificial neural networks (ANN)	Hybrid SVM
Logistic regression (LR)	Support vector machine (SVM)	Hybrid neural discriminant technique
Multivariate adaptive regression splines (MARS)	Sequential minimal optimization (SMO)	Classification and regression tree
Bayesian model	k-Nearest neighbor	ANN and MARS
Decision tree	Case-based reasoning	
Markov model (and Hidden Markov Model)	Genetic algorithm (GA)	
Recursive partitioning		

These credit scoring models performance deteriorates over time. Periodic validation can arrest this shortcoming of the models to retain their accuracy, completeness and timelines of the information used to generate the scores (Robert et al., 1996). The use of historical data contributes to this deterioration as it assumes that credit quality is time independent. An example is the financial crisis of the year 2008 which revealed the intertwined nature of the financial network systems where dependencies led to failure of the credit risk model to estimate the impact of the crisis as the network structural changes were not captured (Capuano, Chan-Lau, Gasha, Medeiros, Santos and Souto, 2009).

The failures highlight the importance of looking into psychological and economic factors that lead to consumers defaulting. A need therefore arises to incorporate social media data in consumer credit risk models as this data is scantly catered for in the traditional consumer models (Ntwiga, 2016). The different models accuracy in credit scoring is highlighted by (Thomas, 2000) in table 2, an indication that more needs to be done in improving existing models accuracy. The new credit scoring models being developed are getting more and more complex and difficult with the implementation cost becoming much higher. A need remains to have models with generalization capability and applicability but this is becoming more elusive (Li and Zhong, 2012).

Table 2: Classification accuracy of credit scoring methods

Author	Linear reg	Logistic reg	RPA	Linear Pro	Neural nets	Genetic Alg
A	43.4	43.3	43.8	-	-	-
B	77.5	-	75.0	74.7	-	-
C	87.5	89.3	93.2	86.1	-	-
D	68.4	-	62.3	-	62.0	64.5
E	66.5	67.3	67.3	-	6.4	-

Another factor is that no mathematical system model is perfect as these models only depict those characteristics of direct interest to the modeler. Inclusion of social media data to complement the existing models not only adds to more accurate estimates but also increases the inclusion of the population with no or scanty financial histories (Ntwiga, 2016). Since we have social media data, advanced powerful models in place, and information technology to “harden” soft information, the next big step is to enhance consumer lending process. We know true default probability is elusive because default estimation depends on the information that is available, but we should continue to enhance the credit scoring systems in the face of globalization (Ntwiga, 2016; Siva, 2010; David, 2004).

3. SOCIAL MEDIA DATA

We highlight some of the benefits geared toward encouraging the use of social media data in the world of consumer lending. Social media has continued to gain widespread acceptance. For example, in the year 2012, Facebook had 1 billion users worldwide while in the same year, Twitter had an estimated 517 million users (Dewing, 2012). Factors that have contributed to this growth are rapid embracement of social media services; increased broadband availability, improvement of software tools and development of more powerful computers and mobile devices (Dewing, 2012; Mustafa and Hamzah, 2011). Social media continue to permeate our social and economic lives and influence how quickly we learn. Data is easily aggregated while the social interactions over time leave a trail of history that informs others about our abilities and dispositions (Jackson, 2008; Chen, Goldberg, Malik and Wallace, 2008).

As social media users continue to generate content and interactions at unprecedented rate, availability of data that waits to be harnessed to improve on the existing consumer credit risk models cannot be ignored (Ntwiga, 2016). Another attribute of social media is that the content posted on social media sites can be found easily using online search tools as it remains there permanently by default (Dewing, 2012). The availability of powerful data science tools for mining internet content offers rich sets of data ideal for consumer credit lending (Ntwiga, 2016). Communication across space and time has expanded by allowing anyone to interact with individuals or groups outside of the physical environment to create, maintain and enhance their social relationship (Mustafa and Hamzah, 2011).

The ability to single out credit worthy consumers can be used to predict other consumers with scanty financial history on their credit worthiness based on the users they interact with in the social media. This is possible because the rise of social networking has allowed people to indicate whom they trust and distrust creating links in the network. This reduces the uncertainty and vulnerability associated with the decisions users make online. Trust assists the users to decide whom to accept information from and with whom to share information

with. Such set of information are crucial in enhancing consumer lending process (Ntwiga, 2016; Dubois et al., 2011; Tang, Hu and Liu, 2014).
Another example is in online trading communities. The seller reputation has a significant influence on online auction process and it is derived from the underlying network and visible to all the people in that system. A good reputation system is supposed to collect, distribute and aggregate the feedback about the agents past behavior. This has been applied successfully by eBay with over four million auctions active at a time through its reputation system, called the feedback forum. The feedback forum assists in knowledge transfer and better understanding of the properties and behavior of the individuals in the network (Resnick et al., 2000). The robustness of a social network assumes that social links are dynamic and allowed to change with and without constraints or restrictions. Such success stories can be improved to fit in well in the process of consumer lending.

4. SOCIAL MEDIA AND CONSUMER LENDING PROGRESS

Evidence suggests that soft information can add value to financial institutions by supplying the missing pieces that complement existing hard information used by banks to make lending decisions. Availability of tools and techniques to analyze unstructured social media data is increasing the possibilities to use this data to design more optimal and efficient strategies on consumer behavior (Siva, 2010).
Social media data has the ability to capture these dynamics and PWC (2015) highlights five key benefits of using an applicant social media data in loan underwriting process: capture new customer segment that has scanty financial history; provide a differentiated customer experience where the analytical client knowledge will demonstrate an understanding of the customer needs; strengthen existing underwriting processes due to availability of more data points and hence helping to limit losses; prevent fraud by using available information to cross-check information provided in loan applications; and develop a competitive edge through better management of new clients with scanty financial history, penetration of new markets, better pricing of risk and improved credit scoring strategy.
Three researchers who have used social media or social network data for credit scoring are highlighted. The mobile phone usage data is used to predict loan repayment in a developing country (Daniel and Grissen, 2015). The duo used behavioral signatures in mobile phone data to predict default with accuracy than the approach of credit scoring using financial histories. They quantified the so called soft information to complement the current methods in use. The approach was found to be promising even for the poor borrowers and those excluded from the main stream financial systems. A subscriber with the knowledge of the algorithm used can manipulate the score but this problem can be overcome by combining with other underwriting techniques. Features used in the study are in table 3. If the mobile phone data was used, the bank can reduce defaults by 41% while still accepting 75% of the borrowers (Daniel and Grissen, 2015).
Another approach in use of social network data to estimate consumer credit risk is the work of (Ntwiga, 2016). Individuals, referred to as agents are assumed to be part of a social and economic network. The data available is time dependent and incorporates the cyclical interdependencies of the agents. A set of variables, the credit risk analysis factors are extracted from the network and are tabulated in table 3. Two other variables are the behavioral scores, distance to default and the credit score. Hidden Markov model (HMM) is used for training and learning the factors to emit the credit score of the agents. A threshold is also estimated from the HMM to

assist in estimating the likely defaulters in the loan portfolio (Ntwiga, 2016). This study offers some promising insights on use of social network data in credit risk management.

The use of social data from Russia's most popular social network to discriminate between solvent and delinquent debtors of credit organizations is the work of Masyutin (2015). The social network data was found to better predict fraudulent cases rather than ordinary defaults, thus ideal to use in enriching the classical application scorecards. Table 3 highlights the social network data used in the study by Masyutin (2015). These are the only academic related studies in the use of social network data in consumer credit scoring.

Table 3: Social network and mobile phone data variables used in credit scoring

Ntwiga (2016)	Masyutin (2015)	Daniel and Grissen (2015)
Age	Age	Age
Gender	Gender	Gender
Trust	Marital status	Top up and depletion patterns
Interactions	Number of days since last visit	Mobility
Risk factor	Number of subscriptions	Pattern of handset use
Sociability	Number of days since the first post	Strength and diversity of social network connections
Relationship strength	Number of user's posts with photos	Intensity and distribution over space and time
Distrust	Number of user's posts with video	Loan size
Private data	Number of children	Loan term in days
Return on private data	Major things in life	
	Major qualities in people	

A number of models to compare the accuracy of customer scoring obtained with and without network data is examined by Wei, Yildirim, Bulte and Dellarocas (2015). They analyzed the benefits of collecting information from consumer's network where people with an above-average chance of interacting with others of similar creditworthiness creating the social scoring. An increase in inclusion of population with limited personal financial history to be offered credit increased due to social scoring. The scoring methods are yet to gain popularity, but as they continue to do so, consumers may adapt their personal networks that may affect their scores positively. The benefits of using social media data in credit scoring increases when it involves networks of ties that exhibit great homophily (Wei et al., 2015).

Social networks are crucial in diffusion of information, ideas, opinions and the influence of individuals in the diffusion process. The social interactions structures that emerge tend to separate the individuals into small interaction groups that offer insights on consumer segmentation for consumer lending (Ntwiga, 2016). As more and more people are getting connected to the internet every day, user interactions and online content continue to increase (Dewing, 2012; Dubois et al., 2011). Social networks continue to accumulate huge amounts of information which can provide valuable insights on people's behavior (Masyutin, 2015). This is the data the financial service industry can use it effectively to make robust and informed decisions in the field of consumer credit risk.

5. CONCLUSION

In this review paper, we presented some of the limitations of the consumer credit risk models. The models are sophisticated but they have inherent weakness as they decay in performance over time. In the year 2008 world financial crisis, they failed to capture the interdependences in the financial markets. The new perspective of social media data offers new impetuous to current efforts to advance the modeling of consumer credit risk (Capuano et al., 2009). The success of some of the online communities not withstanding enables us to replicate this success to consumer lending (Dubois et al., 2011).

Our analysis of the social media data shows promising avenues to assist the existing methods and techniques in consumer loans underwriting. The data has abundant social parameters like trust, interaction intensity between the users and the social groupings found in these networking. This offers an arena for the capture and utilization of the soft factors key in the default process of a loan portfolio. The research work on how social media data can enhance the consumer lending have been highlighted; the use of mobile phone data to predict loan repayment, use of networks in credit scoring and social network analysis to model consumer credit risk.

As the impact of social media on how people interact and share personal information is not fully understood (Dewing, 2012; Dubois et al., 2011), a critical research on this issue arises. Other extensions to this work are to focus on the risk inherent in social media data and how to handle the risks therein. There is need to develop accurate tools to estimate network variables available in social media data and use them alongside the traditional consumer underwriting process to test for their efficacy.

REFERENCES

Capuano, C., Chan-Lau, J., Gasha, G., Medeiros, C., Santos, A., and Souto, M. (2009). *Recent advances in credit risk modeling. International Monetary Fund* Working Paper, No. WP/09/162

Chen, H.J., Goldberg, M., Malik, M. and Wallace, W. (2008). Reverse engineering an agent based hidden Markov model for complex social systems. *International Journal of Neural Systems*, pp. 1-24

Daniel, B., and Grissen, D. (2015). Behavior *revealed in mobile phone usage predict loan repayment*. Department to Economics, Brown University, Working Paper

David, R.K. (2004). *Professional Risk Managers Handbook. A comprehensive guide to current theory and best practices*, Volume III: Risk Management Practices, Professional Risk Managers International Association, United States of America

Dewing, M. (2012). Social media: An introduction. *In Brief, Library of Parliament, Social Affairs Division, Parliamentary Information and Research Services*, Publication No. 2010-03-E, Ottawa, Canada

Dubois, T., Golbeck, J., and Srinivasan, A. (2011). Predicting trust and distrust in social networks. University of Maryland, College Park, USA

Horkko, M. (2010). *The determinants of default in consumer credit market*. (Unpublished Master's Thesis), Department of Accounting and Finance, School of Economics, Aalto University

Jackson, M.O. (2008). *Social and economic networks*. Princeton University Press

Khanbabaei, M., and Alborzi, M. (2013). The use of genetic algorithm, clustering and feature selection techniques in construction of decision tree models for credit scoring. *International Journal of Managing Information Technology*, vol. 5, No. 4

Li, X. L., and Zhong, Y. (2012). An overview of personal credit scoring: Techniques and future work. *International Journal of Intelligence Science*, 2, 181-189

Mustafa, S.E., and Hamzah, A. (2011). Online social networking: A new form of social interaction. *International Journal of Social Science and Humanity*, Vol. 1, No. 2, pp.96-103

Ntwiga, D.B. (2016). *Social network analysis for credit risk modeling*. (Unpublished Doctorate Thesis), School of Mathematics, University of Nairobi, Kenya

PWC (2015). Is it time for consumer lending to go social? How to strengthen underwriting and grow your customer base with social media data. Consumer Finance Group [www.pwc.com/consumer finance]

Raghavan, V., Steeg, G., Galstyan, A., and Tartakovsky, A.G. (2013). Modeling temporal activity patterns in dynamical social networks. Department of Mathematics, University of Southern California, USA

Resnick, P., Zeckhouse, R., Friedman, E., and Kuwabara, K. (2000). Reputation system. *Communications of the Association of Computing Machinery*, Vol.43, No.12

Robert, B.A., Raphael, W.B., Paul, S.C., and Glenn, B.C. (1996). Credit risk, credit scoring, and the performance of home mortgages. *Federal Reserve Bulletin*

Siva, W. (2010). Business intelligence and predictive analytics for financial services. The untapped potential of soft information. Center for digital innovation, technology and strategy, School of Business, University of Maryland

Tang, J.,Hu,X., and Liu, H. (2014). Is distrust the negation of trust? The value of distrust in social media. *Association of Computing Machinery* 978-1-4503-2954-5/14/09, Santiago, Chile

Thomas, L.C. (2000). A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, Vol. 16, pp 149-172

Thomas, L.C., Oliver, R.W., and Hand, D.J. (2005). A survey of the issues on consumer credit modeling research. *The Journal of the Operational Research Society*, Vol. 56, No. 9, pp. 1006-1015

Masyutin, A. A. (2015). Credit scoring based on social network data. *Business Informatics*, No. 3 (33), pp. 15-23

Wei, Y,. Yildirim, P., Bulte, C., and Dellarocas, C. (2015). Credit scoring with social network data. *Marketing Science*, Articles in Advance, pp. 1-25