A Discussion Paper on the Grey Area – The Ethical Problems Related to Big Data Credit Reporting

Victor Chang and Jing Lin
International Buiness School Suzhou, Xi'an Jiaotong-Liverpool Universiy, Suzhou, China

Keywords: Big Data Credit Reporting, Ethical Concerns for Big Data, Controversies on Big Data Uses.

Abstract: With the rise and the development of the "credit society", the credit reporting has played a central role in evaluation one's credit statues, including monitoring and updating creditworthiness of individuals. As the emergence of big data, new tools enabling the credit reporting system to develop new level, by collecting the online and offline data to establish more completely score system. This review paper is aimed to present the difference between the new big data credit reporting and traditional credit reporting, and then explain advantages offered by the new data management. Subsequently, ethical problems will be described due to rising concerns. Being "kidnapped" by the credit reporting applications, users' data will be collected and disposed without prior permission. Some data processes may arise with the messy, unreasonable and fake data resource problems to add more complexities to the existing services which are unable to cope with. As a result, individual users could not verify the correctness of the data and did not know which data would be more trustworthy to be verified for payment and billing. To be worse, users even do not know how to improve their creditworthiness if they have done everything correctly. There are some issues about precision marketing, since some data brokers will target the individuals who was vulnerable to the non-performing and short-term loans. Last but not least, the algorithm of big data prone to evaluate the credit score by groups that individual related to, rather than the individual's own merits, which may lead to discrimination issue, and accelerate the wealth gap problem.

1 INTRODUCTION

Every citizen in USA have a unique society security number, which recorded all the credit activities, include pay off the credit card debt, apply for the mortgage or student loans (Komuves, 1997). There are three main credit bureaus in the American, Equifax, Experian and TransUnion, they are all owned and operated by publicly companies, not owned by any governments. According to your credit history, amounts owed and debts on records, they will generate an individual credit reporting, it can be corresponded to your society security number, which means once you were recorded, it would not disappear. With this credit reporting, you can apply for bank loans or jobs. If you were identified as an individual with a poor credit record, the bank will reject to give you loans and the employer will reject you either. Before the wide adoption of big data technology, the credit reporting is mainly used in the financial field to evaluate applicants' ability to repay their loans (Mierzwinski and Chester, 2013), with more social data and information, the credit reporting system grew into the creditworthiness evaluation system, which can be used in all aspects of life (Miller, 2000). One of the big data reporting strengths is the huge amount data resources that it can get connected, queried and linked altogether. Another strength is the machine learning and AI technologies to support large scale data processing, analysis, categorization and management. Unlike the traditional data resource, the 'thin-file' customers who have less credit history due to age, immigrant status or recent credit record lacking, will get a fair credit score in big data credit reporting (Bureau, 2012). Using the big data services, which can combine as much as relevant credit data, such as the social data, property and online activities. By cross checking to find whether there are no bad records, the thin-file customers can get loans by a lower rate.

However, as an idiom has rightly pointed out - the water that bears the boat can be the same that swallows it up. The rise of quantity, variety and complexity in data can lead to chaos, if it is not

Chang, V. and Lin, J.
A Discussion Paper on the Grey Area – The Ethical Problems Related to Big Data Credit Reporting.
DOI: 10.5220/0006823603480354
In *Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security (IoTBDS 2018)*, pages 348-354
ISBN: 978-989-758-296-7

properly managed. In the data collection and transformation, most people will face the all-or-nothing decision. The user will not have access to services if he/she rejects to upload the private information, such as locations, or delivery addresses. Although some application suppliers announce that the application will not use the customers' data for profits, their authenticity is debatable. There are already some cases of erroneous data problems, two women's credit reporting have been swapped because they have similar personal information. In addition, some researchers founded that there exist the fabricated issues in the data input procedures. While putting the subjective views aside, data problem may be structured because of a certain level of error rate in model operations (Pasquale, 2015). The Federal Trade Commission have showed in the report that the error rates up to 26%. The structural problem also related to the machine learning processes, due to the complex algorithms, the machine will decide how much attention an emphasis it needs to give and dynamically adjust the system itself regularly. It would be difficult to perform manual regulations due to millions or billions of users involved.

Additionally, machine learning and AI techniques may not be robust enough to provide secure services. Some developers can use the tools to aim at the vulnerable, high-value targets for non-performing loans, just like sell heroin to drug addicts (Mierzwinski and Chester, 2013). Furthermore, the machine learning can also assess the individual's credit situations by evaluating their circles of friends, groups and communities of politics, religions, and others. Instead of evaluating the individual's own merits, it is unfair to the vulnerable groups. As the situation continues to deterioration, the wealth gap will be widened, for the wealthy people will get more service and better remedial measure to keep their credit reporting more dignity.

2 LITERATURE REVIEW FOR BIG DATA CREDIT REPORTING

Hilda et al. (2016) indicated the benefit of big data used in the credit reporting field, in his review, he illustrated some new data processing methods, which can be profitable in business. Singer (2012) explain some new data source, such as the web browsing data or user purchase records, can be used in the credit evaluation process. The new data added in can help the thin-file people get reasonable evaluation (Bureau, 2012). While in the data collection, user will be forced to make an all-or-nothing choice between the acceptance of free services or refuse to be recorded (Bilogrevic et al., 2014). The Federal Trade Commission (2013) also points out that credit score counting process with machine learning contain a certain percentage of errors. Jacobs (2015) manifests that the online data has inaccuracy problems. Hurley and Adebayo (2016) claims that individuals could not improve their credit score due to the non-transparent issue. Furthermore, some "precision marketing" will target the people who were vulnerable to the high interest rates but short-term loans, which will indulge the high-risk borrowers (Trusts, 2014). As the machine will be prone to assess the individual's credits by the group-referential processing, the discrimination issue will be deteriorated (Meyer, 2015). When such vicious cycles repeat more often, wealth gap will be widened (Hurley and Adebayo, 2016).

Before 1980s, lending a loan is decided by individual loan officers and specialists who assess applicants, which can differ from person to person (Citron and Pasquale, 2014). Then, the decision for lending based on the automated credit scoring systems developed by the Fair and Isaac Corporation(FICO). Moreover, the credit scoring is more useful, it uses sum of a person's apparent creditworthiness to make underwriting decisions, similar to "predict the relative likelihood of a negative financial event, such as a default on a credit obligation." (Yu, 2014). In America today, every citizen has equipped with one unique social security number, which record all the credit activities, such as paying off credit card and loan, repaying the education loans. Customers can get their credit reporting from three main credit bureaus in America, Equifax, Experian and TransUnion, they both publicly-traded, not owned by government. For the most of Americans, without a good credit score will cause many inconveniences, such as fail to get a mortgage, fired by employer, or even lose the opportunity for education. As a result, it is essentially important to pay attention to individual credit score.

Nowadays, Hilda et al. (2016) claimed that use of big data has been adopted by more organizations. This can be done by identifying patterns, clusters and outliers that are not obvious by the traditional methods. The ultimate goal is to bring profits to business. As the big data technology has becomes more mature, it can handle the massive amounts data with high speed. The field of credit reporting has been reforming.

2.1 The Big Data Credit Reporting Different from Traditional One

2.1.1 Data Source

The big data credit reporting is different from the traditional one, the first part is data source, the big data credit reporting can utilize the different kinds of non-traditional data, which include the criminal records, social media data, consumers' retail spending history, the online purchase records, internet browsing history, or even an individual's friend circle on social networking board (Singer, 2012). Although the traditional data sources still have the basic of credit system, the new alternative tools equipped with the big data has rapidly occupying the market. Like the Experian, it has used big data to exploit thousands of 'universal customer profiles' that can integrate both the online and offline data (Tewksbury, 2013). As same as the Fair and Isaac Corporation(FICO) the main planner of lending decisions has exploit a new system combined with non-traditional data to assess the thin-file borrowers, for they lack the traditional data and difficulty to evaluate. The new 'FICO Score XD' is the result of the cooperation between the FICO and credit bureau Equifax, they use the cable and mobile phone accounts to predict the consumers' creditworthiness (Carrns, 2015). Nowadays, the most remarkable credit scoring and underwriting company is the ZestFinance, who helps two groups of people getting loans. One group includes the people who cannot meet the basic requirements of lending because their FICO score is less than the 500, and other group who have high loan cost and low credit score. It uses the proprietary algorithms to analyse several thousand data points per individual in order to arrive at a final score (Lohr, 2015).

2.1.2 Machine Learning

The new method adopted by ZestFinance was based on the machine learning, which is also the different part with the traditional credit reporting system. Machine learning is a method which can automatically explore the data patterns and used the pattern to predict the future data (Robert, 2012). There are two styles of machine learning, supervised and unsupervised. The supervised one try to find out the relationship pattern between the target which the data analyst already confirmed and the other data points collected. Unsupervised try to predict the target variable, assisting analyst understand how the data was generated and discovered the pattern behind them (Calders and Custers, 2013). The big data credit reporting belongs to the use of supervised machine learning tools.

2.2 The Big Data Credit Reporting's Characteristic and Advantage

With numerous data sources and machine learning tool, the big data credit reporting has been rapidly developed. One reason for traditional credit reporting been replaced is that the big data credit reporting can predict the thin-file consumers, the people who cannot get access to the credit score, such as many immigrants or recent college graduates with little or no credit history, or people who have not activated its credit account at least six months. With the big data tool, equipped with huge data and high speed, these people can also get the credit derives from their normal life activities. Another reason is its foresight, it combines as much as possible relative credit data, such as the usage time of the same phone number, connections on Facebook, a stable address or even the use of proper capitalization in filling out a form, they can assess the credit in every aspect. The last but not the least reason is that it can protect the vulnerable groups, for example, LexisNexis had created a new credit score system called RiskView (Bureau, 2012), this product include the traditional public record like the foreclosures and bankruptcies, in addition, it also include the educational history, professional licensure data, and personal property ownership data, as a result, these people who don't have the traditional credit score but have an professional license or pay rent on time, or own a car, may get a better access to credit system than they otherwise would have (Ramirez, 2016). The Future of Privacy Forum (Polonetsky, 2014) report also indicated that business and government use the big data to protect and empower the disadvantaged group, giving them the access to job markets, disclosure the discriminatory practices and improve the education and help those in need.

3 THE ETHICAL PROBLEM ABOUT THE BIG DATA CREDIT REPORTING

3.1 The Big Data Credit Reporting Different from Traditional One

When you use mapping software it might record all your footmarks and sell it without your permission, if you don’t give the application permission to access your information, then you cannot use the function, you cannot control what information would be collected and often be forced to make all-or-nothing decision between receiving free services or refusing to be profiled (Bilogrevic et al., 2014). Not only a fraction of people has such thoughts, the investigations revealed that, seven in ten Europeans are worried about that software operator will use the personal data for profit or other uncontrolled use (Social, 2011). Over the past few years, the online applications have collecting the user’s information to build the individual profiles, then selling them to the advertiser and data broker to make profit, the users have little control with the data collect process.

3.2 The Messy, Unreasonable and Faked Data

Some aggressive big data enthusiasts claimed that we should embrace the ‘messy data’, for the errors in data manipulated process can help develop better pattern result, the more mistakes there are, the more preparations and well-equipped the system has gone through (Alloway, 2015). However, the messy problem had already affected the normal person life. There is one case, the victim is Judy Thomas, who sued Trans Union for regularly mixing her report with Judith Upton. One day in 1996, Judy found there were some mistakes since some bad debts appeared on her credit report. After the investigation, she found the bad debts was belonging to a woman named Judith Upton, the one who has similar first name and same birth year with her, and only one number different in their Social Security numbers. Due to the operation of one careless staff, Judy could not get the loan based on her expectancy value, she struggled to appeal for justice. Finally, Judy was awarded more than 5 million dollars by a Federal Court jury as a compensation (Weisman, 2013). In this case, it reveals problems with the quality of data in credit reporting system as follows, first, erroneous data could be caused by some careless problems, second, there were fake and fabricated data, researchers found that some investigators of the credit system even fabricated derogatory information about individuals. Apart from quality of data, the lack of third party central regulation on credit agencies appear to be an issue. While the credit rating agencies do not have the direct relations with the customs, they lack incentives to treat the individuals fairly, if profits and finding vulnerable individuals appear to be their concerns. There is also another challenge - it is difficult to interchange with and between the credit rating agencies. To sum up all the observations above, the data problem is not just an anecdotal, it is structural (Cetorelli and Gambera, 2001; Kenny, 2014). In a detailed and comprehensive investigation conducted by scholars and Federal Trade Commission, there are almost 3,000 credit reports belonging to 1,000 consumers, it has been found that 26 percent had “material” error problems which were serious enough to affect the individuals credit scores. Even with the conservative estimates, there were still 23 million Americans having errors in their credit reports. Once their credit scores amended, they will obtain credit loan in lower prices (commission, 2013). On the other hand, some researchers and industries might think the big data can improve the situations, since more information can be cross checked and validated for better precision and accuracy. Some data may seem logically related with the credit record, for example, payment historical records matches the identity of the right person. However, there are other “fringe data” like the reading time of user notice to indicate the individual’s degree of care might lack the correlations with the creditworthiness (Yu, 2014). Additionally, lots of evidence confirm that since thousands of data were from the individuals on and off the site activities, it might result in getting a high percentage of inaccurate information. As Jacobs (2015) claimed that mobile location data can be easily lead to inaccuracy. Inaccurate data problems can deeply affect the individual credit activities. While the big data can improve on speed and interconnectivity, it might less efficient to double check accuracy and validity of the data, the problem may be worse.

3.3 Non-Transparency

The “kidnap service” and data source problem both brought out one issue - individuals cannot get access to the data process. The operation processes are opaque to customers, like the case about Kevin Johnson, who was a person with a good and decent credit. He received a letter from the American Express in one day of late 2008, which told him that the credit limit of his credit card had been decreased from $10800 to $3800. The reason given by American Express was that, the market which Kevin recently shopped by have some customers who have bad credit history with American Express. While Kevin casted around for an explanation, the American Express did not want to share more details about that. Even the Federal Trade Commission have sued the three major credit rating agencies in 2000, because they did not answer phones, which means the customers have always been neglected. If users did not know how their scores been calculated and even cannot complain about the unfair scores, the fairness of credit reporting can be questionable. On the basis of electronic privacy information centre, big data credit evaluators are mainly "concerned about collecting a large amount of information about individuals" and the overall quality of the data may be affected (Scoring, 2016). With non-transparent problems exist, individuals cannot identify the unfair credit scores and fail to improve scores or prevent them from falling further (Hurley and Adebayo, 2016). The non-transparent problems exist in two aspects, one is in the data collection and transformation process, another is due to the machine learning algorithms. Guzelian et al. (2015) explain that the credit companies treat their data as the commercial secrets. Therefore, the data collection progress would be opaque, in case the competitor found out data resources, the customers could not get access to the data, they could not find whether the data was accurate either. Even if the customers know the data is accurate, it is difficult to find one error in the millions of entries in the credit scorer is raw data set. Since the machine learning process is designed to find the relationship between data and target, the data collected from customers will be transferred to the language that computers can understand, the process will be involved with mass aggregations and combinations of data points. In this case, if the machine learning algorithm was a “layman”, then the layman could not understand how much a dataset would need and what types of diagnosed emphasis for further actions. As a result, it is not straight forward to get some effective evidence to identify the algorithm validity.

3.4 Heroin for Drug Addicts

Trusts (2014) shows the lenders will use the most advanced technology and complex algorithm to target the vulnerable persons, particularly those attracted by the low-valued, short-term credit products with usurious interest rates and highly adverse terms. The experts at Upturn have also supported this view. There is potentially one possible situation: While some credit reporting system developers will not interest in the system improvement or predicting the consumer creditworthiness, they choose to put efforts on finding vulnerable or high-value targets for non-performing loans. The survey shows the target disproportionately come from poor and minority communities (Hurley and Adebayo, 2016). Angwin (2014) identified that although no unambiguous evidence showed the big data is used in the identified the vulnerable borrower, some major data brokers who work on credit reporting system have been accused for bringing the "sucker lists" to the market, the list which specifically target on vulnerable groups of people, which may cause them distress or more pain to their existing debts. The evidence can be found from 2013 Senate Commerce Committee report which have lists with title like "Hard Times", "Burdened by Debt", "Retiring on Empty" and "X-tra Needy" which were deliberately aimed at the individuals who were trying to buy some unfavourable financial products.

3.5 Discrimination

Facebook is another good example. It is a free social networking website which can allow users to upload their photos or videos, and post their comments, has recently applied for one patent application. The patent is related to one method for "authorization and authentication based on an individual's social network." which means the users’ social information can be used to evaluate their credit score. The patent application explains that, when someone applies for loan, the lender can get access to the individual’s social network profile and get the rank of his social group, if the social group meet the minimum requirement of the lender, then the lender can give a loan to the applicant, otherwise, he will be rejected. Criticasters have indicated that the tools can bring about the new style of digital redlining (Meyer, 2015).

The machine learning may also produce results that have inequity biases, because the person’s final credit score is not based on the individual’s own value but prefer on the basis of the relative group that the system affirm effectively. As Barocas and Selbst (2016) claim that when a model relies on the generalization reflected in the data, the final result of individual maybe statistically sound inference, so that the result may be inaccurate. However, this may happen when customers do nothing with it and they cannot do anything to revert. Like the case of Kevin Johnson mentioned earlier, the model will punish individuals that in a particular environment like the designated community or having some issues with the political or religious grounds. The case of Kevin Johnson is not a special case, in lots of other fields, from the school admission decisions to the insurance selling, the big data tool will judge the person by the shared data rather than individuals' own and true value, which have been proven to aggravate existing prejudice (Lowry and Macpherson, 1988).

3.6 Wealth Gap Widen

Schmitz (2014) claimed that the use of big data in the credit reporting field can foster the discrimination, due to the data analysis subjective rules the machine will give an aid to the smartest customers with the best services and best remedial measure, which will widen the gap between the wealthy and poor individuals. As the Hurley and Adebayo (2016) explained in the report, the big data tools will increase the risk of creating a system of "creditworthiness by association", which combine the individuals' families, religions, sexual orientation and other relevant information. All information will figure out the individual's credit score, whether you can pay for a load is not determined by how much you can earn, but the group you are in, which will further aggravate discrimination. Like the example of Kevin Johnson, he became the proxies of sensitive information like the race and vulnerable attributes. In this example, the big data could not eliminate the prejudice, but aggravate the existed bias.

4 CONCLUSION

In this paper, we report the grey area about the use of big data, particularly for credit reporting system that has biases and selection on vulnerable groups of people. Big data enthusiasts argue that all the data can be managed by credit reporting and the big data will lead us to live happily and comfortably. While we can find is that the data maybe objective, the process of data categorization and evaluation can be subjective. Moreover, the ways data will be collected or diagnosed largely depend on the existing social resources. Individuals who have contributed their data to the credit reporting system, have no or few effective actions to modify the "unreasonable" data. The vulnerable people will be likely to be trapped in a vicious circle by precision marketing or have been considered as low credit groups due to potential discrimination in income, social status, race, politics or religion. On the other hand, the wealthy people can possess the best remedial measures and excellent services. While we need to use the big data to build the completeness and soundness of the credit reporting system. However, we should also pay attention to ethical issues raised by the big data and find betters to manage and treat each individual fairly and equally.

REFERENCES

Alloway, T., 2015. *Big data: Credit where credit's due* [Online]. Available from: https://www.ft.com/content/7933792e-a2e6-11e4-9c06-00144feab7de (Accessed: 2017).

Angwin, J., 2014. *Dragnet nation: A quest for privacy, security, and freedom in a world of relentless surveillance*. Macmillan.

Barocas, S., Selbst, A. D., 2016. 'Big data's disparate impact', *Cal. L. Rev.*, 104, p. 671.

Bilogrevic, I., Freudiger, J., De Cristofaro, E., & Uzun, E. (2014) 'What's the gist? privacy-preserving aggregation of user profiles', *In:* KUTYŁOWSKI, M. & VAIDYA, J. (eds.) *Computer Security - ESORICS 2014: 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II.* Cham: Springer International Publishing.

Bureau, C. F. P., 2012. *Analysis of Differences between Consumer-and Creditor-Purchased Credit Scores*.

Calders, T. & Custers, B., 2013. 'What Is Data Mining and How Does It Work?' In: Custers, B., Calders, T., Schermer, B. & Zarsky, T. (eds.) *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*. Berlin, Heidelberg.

Carrns, A., 2015. *New Credit Score Systems Could Open Lending to More Consumers* [Online]. Available from: http://lippmancpas.com/new-credit-score-systems-could-open-lending-to-more-consumers (Accessed: 2017).

Cetorelli, N., & Gambera, M., 2001. 'Banking market structure, financial dependence and growth: International evidence from industry data', *The Journal of Finance*, 56(2), pp. 617-648.

Citron, D. K., & Pasquale, F., 2014. 'The scored society: due process for automated predictions', *Wash. L. Rev.*, 89, 1.

Commission, F. T., 2013. Section 319 of the Fair and Accurate Credit Transactions Act of 2003: Fifth Interim Federal Trade Commission Report to Congress Concerning the Accuracy of Information in Credit Reports.

Guzelian, C. P., Stein, M. A., & Akiskal, H. S., 2015. 'Credit Scores, Lending, and Psychosocial Disability', *BUL Rev.*, *95*, p.1807.

Hilda, J. J., Srimathi, C., & Bonthu, B., 2016. 'A review on the development of big data analytics and effective data visualization techniques in the context of massive and multidimensional data', *Indian Journal of Science and Technology*, *9*(27), pp.1-13.

Hurley, M., & Adebayo, J., 2016. 'Credit scoring in the era of big data', *Yale JL & Tech*., 18, p.148.

Jacobs, S., 2015.*Report: More Than Half of Mobile Location Data is Inaccurate* [Online]. Available from: http://streetfightmag.com/2015/05/14/report-more-than-half-of-mobilelocation-data-is -inaccurate (Accessed: 2017).

Big data: a tool for fighting discrimination and empowering groups (2014) [Online]. Available from:

https://fpf.org/2014/09/11/big-data-a-tool-for-fighting-discrimination-and-empowering-groups (Accessed: 2017).

Kenny, D. A., 2014. Measuring model fit.

Komuves, F. L., 1997. 'We've Got Your Number: An Overview of Legislation and Decisions to Control the Use of Social Security Numbers as Personal Identifiers', *J. Marshall J. Computer & Info. L.*, 16, p.529.

Lohr, S., 2015. *Big Data Underwriting for Payday Loans* [Online]. Available from: http://bits.blogs. nytimes.com/2015/01/19/big-data-underwriting-forpayday-loans/? r=0 (Accessed: 2017).

Lowry, S., & Macpherson, G., 1988. 'A blot on the profession', *British medical journal (Clinical research ed.)*, *296*(6623), p.657.

Meyer, R., 2015. 'Could a Bank Deny Your Loan Based on Your Facebook Friends?', *The Atlantic*.

Mierzwinski, E., & Chester, J., 2013. 'Selling consumers not lists: the new world of digital decision-making and the role of the Fair Credit Reporting Act', *Suffolk UL Rev.*, *46*, p.845.

Miller, M., 2000. 'Credit reporting systems around the globe: the state of the art in public and private credit registries', *World Bank. Presented at the Second Consumer Credit Reporting World Conference, held in San Francisco, California, October*.

Pasquale, F., 2015. '*The black box society: The secret algorithms that control money and information',* Harvard University Press.

Robert, C., 2014. Machine learning, a probabilistic perspective.

Ramirez, E., 2014. *Big Data: A Tool for Inclusion or Exclusion?* US FTC.

Schmitz, A. J., 2014. 'Secret Consumer Scores and Segmentations: Separating Haves from Have-Nots', *Mich. St. L. Rev.*, p.1411.

Scoring, C., 2016. Electronic Privacy Information Center

Singer, N., 2012. 'Mapping, and sharing, the consumer genome', *New York Times*.

Social, T. O., 2011. Special Eurobarometer 359 Attitudes on Data Protection and Electronic Identity in the European Union.

Tewksbury, M., 2013.*The 2013 Big Data planning guide for marketers.* Experian Marketing Services.

Trusts, P. C., 2014. Fraud and abuse online: Harmful practices in internet payday lending. Washington, DC.

Weisman, S., 2013. *A Guide to Elder Planning: Everything You Need to Know to Protect Your Loved Ones and Yourself*. FT Press.

Yu, R., 2014. Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace.