

Fraud Detection Social Security and Social Welfare Data Mining

A.Jenifer Sophia, S.Parthiban

M.E(CSE),

pjeniferarul@gmail.com,

M.E(CSE),

sparthiban2012@gmail.com

Abstract— The importance of social security and social welfare business has been increasingly recognized in more and more countries. It impinges on a large proportion of the population and affects government service policies and people's life quality. Typical welfare countries, such as Australia and Canada, have accumulated a huge amount of social security and social welfare data. Emerging business issues such as fraudulent outlays, and customer service and performance improvements challenge existing policies, as well as techniques and systems including data matching and business intelligence reporting systems. The need for a deep understanding of customers and customer-government interactions through advanced data analytics has been increasingly recognized by the community at large. So far, however, no substantial work on the mining of social security and social welfare data has been reported. For the first time in data mining and machine learning, and to the best of our knowledge, this paper draws a comprehensive overall picture and summarizes the corresponding techniques and illustrations to analyze social security/welfare data, namely, social security datamining (SSDM), based on a thorough review of a large number of related references from the past half century. In particular, we introduce an SSDM framework, including business and research issues, social security/welfare services and data, as well as challenges, goals, and tasks in mining social security/welfare data.

Index Terms—Government data mining, public sector, public service, social security data mining (SSDM), social security, social welfare.

I. INTRODUCTION

MACHINE learning and datamining are increasingly used in business applications, and in particular, in public sectors. A distinct public-sector area is social security and social welfare which suffers critical business problems, such as the loss of billions of dollars in annual service delivery because of fraud and incorrect payments. People working in different communities are increasingly interested in “what do social security data show” and recognize the value of data-driven analysis and decisions to enhance public service objectives, payment accuracy, and compliance. Within the marriage of machine learning and data mining with public sectors, an emerging datamining area is the analysis of social security/welfare data. Mining social security/welfare data is challenging. The challenges arise from business, data, and the mining of the data. Social security data are very complex, involving all the major issues that are discussed in the data quality and engineering field, such as sparseness, dynamics, and distribution. Key aspects contributing to challenges in mining social security data are many, e.g., 1) specific business objectives in social security and government objectives, 2) specific business processes and outcomes, 3) heterogeneous data sources, 4) interactions between customers and government officers, 5) customer behavioral dynamics, and 6) general challenges in handling enterprise data, such as data imbalance, high

dimension, and so on. Since then, many researchers have worked on different topics. The majority of research has been conducted from political, economic, sociological, and regional perspectives, compared with a much delayed effort made on the technical aspects. The main issues involve problem analysis, process and policy modeling, business analysis, correlation analysis, infrastructure development, and emerging data-driven analysis. The dominant fact and trend of policy and economy oriented studies, very limited research can be found in the literature on mining social security and social welfare data. *Social security data mining (SSDM)* seeks to discover interesting patterns and exceptions in social security and social welfare data.

It aims to handle different business objectives, such as debt prevention. From the data mining task perspective, it involves both traditional data mining methods, such as classification, as well as the need to invent advanced techniques, e.g., complex sequence analysis. Australia is one of the most developed social welfare countries in the world in terms of government policies, infrastructure, the population of benefit recipients, and the advancement of social security techniques and tools. 1) Through a series of projects. We have developed models, customer circumstance changes, declarations, and interactions between customers and government officers. 2) The findings have proven to be very useful for overpayment prevention, recovery, prediction, and deep understanding of customer activities and intervention, the recovery and prevention of overpayments for the government. These substantial practices, Data Mining top ten data mining case studies, offer an opportunity for us to widely review the relevant

work, deeply explore SSDM in conjunction with real-life applications in Australia and present the overview of SSDM. Focusing on a specific SSDM technique, Aim to draw an overall picture of SSDM by sharing our experience, observations, and lessons learned in both reviewing the related work and conducting real-life SSDM tasks. In particular, the main contributions consist of 1) a thorough literature review of social security research in the last half century, and discussion of different categorizations of the related work; 2) a comprehensive framework of SSDM, discussing the main data mining goals, tasks, and principal challenges in mining patterns in social security data; 3) a summary of several case studies, which involve the development of new and effective algorithms and tools to handle social security data. In particular, we highlight the work on mining debt-targeted patterns, such as debt targeted positive and negative sequences, sequential classifiers using both positive and negative sequences, and combined association rules by engaging multiple sources of data; and 4) the extension of discussions about mining general public sector data. The related work on social security research in the past 50 years. Introduce social security business and data characteristics. Outlines a framework for SSDM, Including the main goals, tasks, and challenges in mining social security data. Public-sector data mining based on the lessons learned in conducting SSDM in Australia.

II. REVIEW ON SOCIAL SECURITY/WELFARE RESEARCH

A. Comprehensive Picture

Research on social security and welfare issues started in the mid-20th century. Since then, broad-based issues have been 2datamining.it.uts.edu.au/ssdm added to the investigation and can be categorized into the following main streams.

1) *Political perspective*: One of the main streams of research investigates the problems, issues, factors, and impact of social security and welfare from public policy, social policy, administration, governance, resistance, and practice viewpoints.

2) *Economic perspective*: Another dominant fact and trend is the exploration of issues and the effect of social security models and factors from the standpoint of econometrics, public economics, and political economy. This involves analysis and discussions about economy, earnings, rating, savings, growth, privatization, reform, labor supply, multi entity relationship analysis, and optimal arrangements.

3) *Sociological perspective*: Some researchers are concerned about the social effect of social security policies on society, such as lifecycle, demographic, behavior, aging, retirement, fraud, fairness and affordability, etc.

4) *Regional perspective*: Researchers from different countries introduce the development of social security in their countries, for instance, Canada, India, Latin America, the U.S., Britain, Sweden, China, Italy, Germany, France, and Europe.

5) *Technical perspective*: An emerging trend in social security is the study of technical issues, e.g., infrastructure development, knowledge management, policy and process

modeling, data-driven analysis, and correlation analysis crossing multiple areas .

B. Technical Perspective

From the technical perspective, the main issues that have been addressed in the literature focus on several areas, including problem analysis, process and policy modeling, business oriented analysis, correlation analysis, infrastructure support, and data-driven analysis.

1) *Problem analysis*: From time to time, we find papers discussing or debating the issues of reform, crisis, issues for policies, privatization, uncertainty, optimization, fraud, and effect on economy, society, capital market, human resources, etc.

2) *Process and policy modeling*: Different approaches, e.g., empirical analysis, time-series analysis, quantitative comparative analysis, and equilibrium analysis, have been used and developed to design, simulate, and evaluate policy, pension, benefit, process and their effects, as well as their optimization, choice, and performance rating including accuracy.

III. Related Work

The public sector has also kept “the frontier spirit alive in the computer science community”. In particular, data-driven decision has recently been increasingly recognized as one of the most powerful tools to improve government service objectives. However, mining social security/welfare data is an open, new area in the data mining community. To the best of our knowledge, only two groups 3 have involved SSDM, and a very limited number of relevant publications can be found in the literature. In the following, we discuss the UNC group’s work and address the practices by the UTS group. A case study was conducted on monthly service data and service variations to detect common patterns of welfare services given over time. The study’s authors used a simple sequence analysis method on monthly service administrative databases, which indicates what services were given when, to whom, and for how long. While “common” service procedures can be identified by simply applying a frequent sequence analysis method, it appears that no additional advancement has been made in tackling critical challenges in the data, e.g., mixed transactional data, imbalanced items, and labels. The method only identifies general frequent procedures that are commonsense to business people. No informative and implicit patterns can be identified in this case study. From a business perspective, the identified frequent patterns are not very helpful, since they reflect the actual service arrangements implemented as per policies. Business people want to discover something they do not already know about their business and to develop a deep understanding of why, and how, specific problems face the organization. In our substantial literature review of SSDM, no additional references have been identified that provide substantial insights for mining social security/welfare data. For this reason, this paper presents a comprehensive overview of SSDM, starting from discussing the characteristics of business and data in followed by an SSDM framework

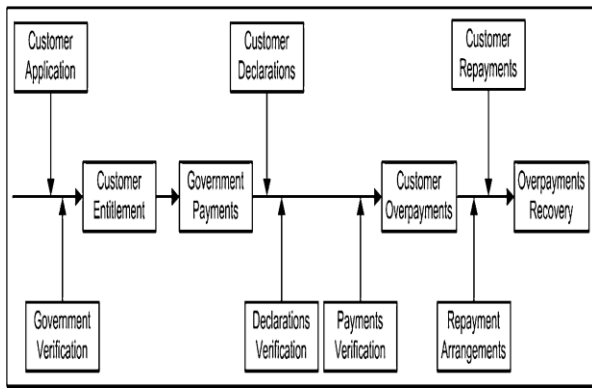


Fig.1. social security business workflow.

IV. CONCLUSION

The occurrence of the global financial crisis, more and more governments have realized the necessity of enhancing social security services objectives and quality. Data mining and machine learning can play a critical role, as we have demonstrated in mining Australian social security data for debt prevention, recovery, customer analysis, etc., during the past few years. However, as the literature review shows, mining social security (and public sector) data are still an open field for business applications in data mining and machine learning. Very few references have been publicized. In this paper, for the first time in the community, we present a picture of studies on social security issues and summarize the key concepts, goals, tasks, and challenges of SSDM, based on our experience and knowledge accumulated through conducting data mining in Australian social security data. We have also highlighted several case studies of mining social security data, including modeling the impact of activity/activity sequences, mining impact-targeted activity patterns, mining positive and negative sequential patterns, conducting impact-targeted sequence classification, and mining combined association rules. We have discussed how the identified patterns are converted into knowledge that can support business people in a more user-friendly way to take decision-making actions. While these case studies aim to present a picture of what can be done in SSDM, many references have been provided so that readers can access information about the specific techniques in more detail. We are currently working on the remaining tasks and challenges that are discussed in this paper, such as, detecting fraud in social security data.

REFERENCE

[1]. H. Berghel, "Identity theft, social security numbers, and the web," *Commun. ACM*, vol. 43, no. 2, pp. 17–21, 2000.

[2]. L. Cao, Y. Ou, and P. S. Yu. (2011). "Coupled behavior analysis with applications," *IEEE Trans. Knowl. Data Eng.*, to be published

[3]. M. Feldstein, "The optimal level of social security benefits," National Bureau of Economic Research, Cambridge, MA, Working Paper 0970, Aug. 1986.

[4]. Y. Zhao, H. Zhang, L. Cao, C. Zhang, and H. Bohlscheid, "Mining both positive and negative impact-oriented sequential rules from transactional data," in *Advances in Knowledge Discovery and Data Mining* (ser. Lecture Notes Computer Science 5476). New York: Springer-Verlag, 2009, pp. 656–663.

[5]. B. D. Bernheim. (1987, May). "Social security benefits: An empirical study of expectations and realizations," National Bureau of Economic Research, Cambridge, MA, Working Paper 2257.

[6]. M. J. Boskin and G. F. Break, *The Crisis in Social Security: Problems and Prospects*. Oakland, CA: Inst. Contemporary Stud., 1977.

[7]. G. Burtless and R. A. Moffitt, "Social security, earnings tests, and age at retirement," *Public Finance Rev.*, vol. 14, no. 1, pp. 3–27, 1986.

[8]. L. Cao, Y. Ou, P.S. Yu, and G. Wei, "Detecting abnormal coupled sequences and sequence changes in group based manipulative trading behaviors," in *Proc. Knowl. Discovery Data Mining*, 2010, pp. 85–94.

[9]. L. Cao, P. S. Yu, C. Zhang, and H. Zhang, *Data Mining for Business Applications*. New York: Springer-Verlag, 2008

[10]. L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E. K. Park, "Flexible frameworks for actionable knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 9, pp. 1299–1312, Sep. 2009.

[11]. C. Coile, P. Diamond, J. Gruber, and A. Jouten, "Delays in claiming social security benefits," *J. Public Econ.*, vol. 84, no. 3, pp. 357–385, 2002.

[12]. M. R. Darby, "The effects of social security on income and the capital stock," UCLA Dept. of Econ., Los Angeles, UCLA Econ. Working Paper 095, Mar. 1978.

[13]. O. S. Mitchell and J. W. Phillips, "Retirement responses to early social security benefit reductions," National Bureau of Economic Research, Cambridge, MA, Working Paper 7963, Oct. 2000.

[14]. *Pattern Analysis and Risk Control of E-Commerce Transactions to Secure Online Payments*, Australian Research Council Linkage Grant, 2007–2009.

[15]. *Detecting Incorrect Income Declaration in Real Time*, UTS-Centre link Contract Research Project, 2010.