International Journal of Computing Communication and Information System(IJCCIS)
Vol.7. No.1 –2015, Pp. 15-21
©gopalax Journals, Singapore
available at : www.ijcns.com
ISSN: 0976–1349

SOCIAL SECURITY AND SOCIAL WELFARE DATA MINING: AN OVERVIEW

T.Balasubramanian
Department of Computer Science
Sri Vidya Mandir Arts and Science College
Uthangarai, Krishnagiri (Dt).
Tamilnadu, India
balaeswar123@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 data mining (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. A summary of SSDM case studies is also presented with substantial citations that direct readers to more specific techniques and practices about SSDM.

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

Keywords: Data mining, Technical Perspective with data mining

I.INTRODUCTION

ACHINE learning and data mining are increasingly used Min 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. 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)

MACHINE learning and data mining are increasingly used Min 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. 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.

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. Since 2004, we have been engaged in conducting data mining for the Australian Commonwealth Government[1] through a series of projects. We have developed models, algorithms, and systems to indentify key drivers, factors, patterns, and exceptions, indicating high risk of customers, customer circumstance changes, declarations, and interactions between customers and government officers.[2]

In this paper, rather than focusing on a specific SSDM technique, we aim to draw an overall picture of SSDM by sharing our experience, observations, and lessons learned in both re-viewing the related work and conducting real-life SSDM tasks. This is the first paper in this field, to the best of our knowledge, that provides a comprehensive literature review of over 100 references and a substantial framework of SSDM. 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 catego-rizations 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 clas-sifiers 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.

This paper is organized as follows. Section II summarizes the related work on social security research in the past 50 years. In Section III, we briefly introduce social security business and data characteristics. Section IV outlines a framework for SSDM, including the main goals, tasks, and challenges in mining social security data. In Section V, we briefly introduce our real-life assignments in conducting SSDM in Australia by illustrating five case studies and discuss the development of actionable knowledge for business needs. Section VI discusses public-sector data mining based on the lessons learned in conducting SSDM in Australia. We conclude this paper in Section VII.

II .REVIEW ON SOCIAL SECURITY/WELFARE RESEARCH

II.I . Comprehensive Picture

Research on social security and welfare issues started in the mid-20th century. Since then, broad-based issues have been 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.

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.
3) *Sociological perspective:* Some researchers are concerned about the social effect of social security policies on society, such as lifecycle .

4) *Regional perspective:* Researchers from different countries introduce the development of social security in their countries, for instance.

5) *Technical perspective:* An emerging trend in social security is the study of technical issues.

II.II 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 com-parative 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.

II.III Related Work of Social Security Data Mining

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 in Section V-A.

In a case study was conducted on monthly service data and service variations to detect com-mon 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 fre-quent 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 pa-per presents a comprehensive overview of SSDM, starting from discussing the characteristics of business and data in Section III, followed by an SSDM framework in Section IV.

II.IV Retrospection on Mining Social Security Data

Our substantial literature review work and practices in Australia (see Section V) reveal the following observations about the existing research on mining social security/welfare data.

1) It is a very open area in terms of applying and conducting pattern/anomaly discovery on social security data (i.e., SSDM). In the very limited work available from the literature, no such systematic work has been done in terms of drawing an overall picture of SSDM from either the business or technical side, nor in addressing the challenges and opportunities in SSDM.

According to our experience in conducting SSDM in Australia, social security/welfare business and data consist of comprehensive characteristics and complexities specific to the data mining community which are not comparable with those in many domains. This is reflected through the nature of mixing politics, economy, society, organizational and business processes, and the Internet, as well [3]One at the University of North Carolina (UNCgroup)(http://ssw.unc.edu/ma/index.html) and the other is our group (UTS group) (http://datamining.it.uts.edu.au/ssdm).

III. SOCIAL SECURITY SERVICES AND DATA

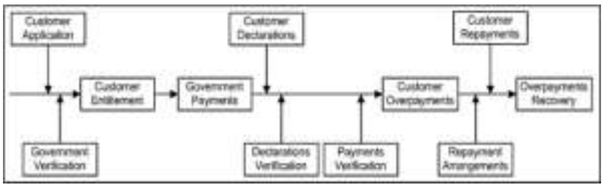
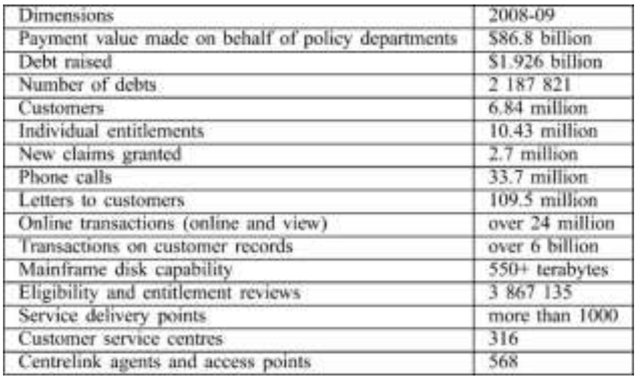


Fig. 3.1:Social security business workflow.

is checked by the government. Payments are arranged on the premise of customer entitlement and policies. The customer is required to declare any changes that may affect payment entitlement. Once a customer declaration is lodged, it will be verified by the government. As a result, customer payments are further verified and adjusted if necessary. In some cases, overpayments to the customer may occur for reasons such as incorrect declaration. The government will seek to recover the debt, and the customer will be requested to pay back such overpayments through repayment arrangements made between the government and the customer. More information about social welfare business can be found on the respective government websites[4] and in reports.

Taking the Australian social welfare business as an example, as a one-stop-shop, the Australian Commonwealth Department of Human Services (Centrelink[5]) delivers a range of Common-wealth services to the Australian community and is responsible for the distribution of around $86.8 billion, i.e., 30% of the Commonwealth's outlay, to about one-third of Australians. In day-to-day interactions between the government and customers, Centre link accumulates a large amount of interaction data. For instance, Centre link provides 361 000 face-to-face services each working day and processes 6.6 billion transactions against customer records each year. It has been shown that the number of such interactions increases every year. The government has progressively recognized the importance of analyzing these interactions to obtain a deep understanding of customers and organization–customer relationships, to actively manage customers, to improve government service quality and objectives, and to inform policy design An issue of particular interest is the identification of the drivers that cause noncompliance in organization–customer interactions. Noncompliance drivers may result from many aspects or staff errors. The 2007–2008 audit report by the Australian National Audit Office (ANAO) drew attention to the importance of deeply understanding customers, and of ad-dressing the behavior and behavioral changes in rising debt from the perspective of the customer, government administration, client group, and community.

TABLE 1.CENTRELINK BUSINESS DIMENSIONS 2008–2009



Dimensions	2008-09
Payment value made on behalf of policy departments	$86.8 billion
Debt raised	$1.926 billion
Number of debts	2 187 821
Customers	6.84 million
Individual entitlements	10.43 million
New claims granted	2.7 million
Phone calls	33.7 million
Letters to customers	109.5 million
Online transactions (online and view)	over 24 million
Transactions on customer records	over 6 billion
Mainframe disk capability	550+ terabytes
Eligibility and entitlement reviews	3 867 135
Service delivery points	more than 1000
Customer service centres	316
Centrelink agents and access points	568

increases dramatically every day. Table 1 provides an overview of some Centre link business dimensions related to data.

As Table 1 shows, the huge amount of social security data accumulated by Centre link consist of very useful information recorded from customer service centers, agents and access points, the Internet, interviews and reviews for all services, customers, staff and agents, and debt. The $1926 million in debt raised in 2008–2009 compares with $1831 million in 2007– 2008. Such data can be classified into the following categories:

1) Customer demographic and circumstance data, recording information about a customer and his/her circumstances, circumstance changes, etc; for instance, home address and the history of address change;

2) Benefit/allowance data, regarding the information about specific benefit/allowance design and applicability in alignment with customer eligibility, and management processes;

3) Customer pathway data, reflecting the history and relevant details of a customer's use of government services, such as the number of services, when, and from which service centers the services have been applied for, and granted;

4) Activity data, providing activity records information about who (maybe multiple operators) processes what types of activities (say change of address) from where (say

customer service centers) and for what reasons (say the ac-tion of receipt of source documents) at what time (date and time), as well as the resultant outcomes (say raising or recovering debt);

5) Facility usage data, regarding the resources used by or for customers, e.g., phone calls and online services;

6) Service policy data, information about policies, the applications of policies to customers with particular circum-stances;

7) Service transactional data, day-to-day information recorded regarding the use of services, such as new registrations, new claims, debt review, etc.;

8) Service performance data, concerning service quality and performance, such as overpayments and their distribution.

IV. FRAMEWORK OF SOCIAL SECURITY DATA MINING

A. Basic Framework

Like any other domain, data mining applications in social security are driven by business objectives and underlying data. Based on the introduction of social security business and data in Section III, Fig. 2 presents a high-level SSDM framework. It consists of three layers: the data layer, the business objective layer, and the data mining goal layer.

The business objective layer includes the main aims and expectations for the implementation of social security services For instance, Fig. 4.1 lists the main objectives , including customer service enhancement (to instantly provide high-quality services to those with particular needs), payment correctness enhancement (e.g., to pay the right amount to those who are eligible), business integrity enhancement (e.g., to improve the consistency and accuracy and to speed up processing), debt management and prevention (e.g., to recover and prevent debt instantly), outlays cause identification (e.g., to identify outlays incurred by staff error), income transparency improvement (e.g., to improve customer earnings reporting and to detect gray income automatically), performance enhancement (e.g., to reduce customer waiting time in service centers or call centers), service delivery enhancement (e.g., to strip out unnecessary contacts and provide easier and more efficient pathways to services), service/risk profiling (e.g., to identify customers most at risk of incorrect payments and to identify opportunities to reduce the debt more efficiently), customer need satisfaction (e.g., to identify customers with special or more urgent needs than others), accountability assurance (e.g., to identify areas of significant. financial or operational risk and to pinpoint more effective arrangements to manage risks), fraud and noncompliance detection (e.g., to identify international or staff fraud and noncompliance), process optimization (e.g., to streamline processes for easier service access and delivery), and key performance indicator (KPI) enhancement (i.e., to identify where and how the key performance indicators can be enhanced).

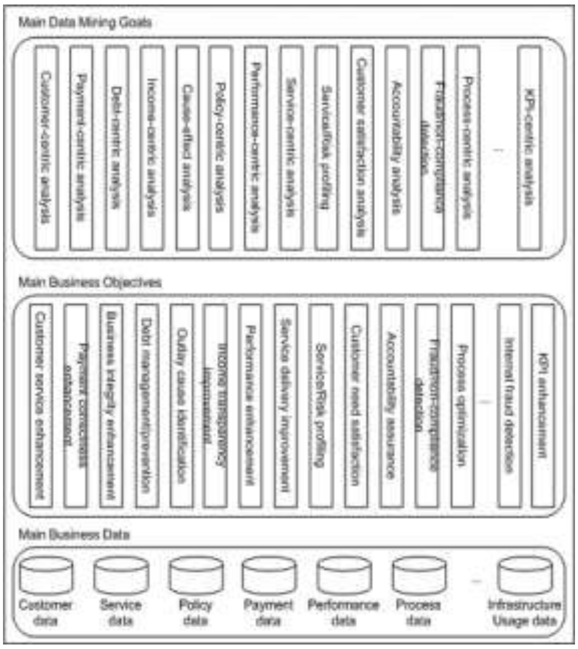


Fig. 4.1 : SSDM framework.

To support the aforementioned major business objectives, the government invests in efficient information infrastructure. As a result, data are acquired and constantly updated at every place and time in the business operation. The data layer summarizes the main data resources. It consists of customer data (customer demographic and circumstance information), service data (service usage and procedural information), policy data (government policy and the applications of policy), payment data (customer payment information), performance data (ser-vice performance and operational performance), process data (business process and change applied to customers), infrastructure usage data (the use of IT resources and services), etc.

While every effort has been made to rectify problems, it has been disclosed that the government is facing longstanding, as well as emerging, problems in achieving and improving the main business objectives [119]. The accumulation of business data provide a unique and essential premise to disclose hidden and implicit channels, indicators, and solutions for these issues, as shown by the data mining pilots in Centre link (see Section V for more information). The data mining layer lists the main goals in mining social security data to enhance business objectives; for instance, customer-centric analysis, payment-centric analysis, debt-centric analysis, income-centric analysis, cause–effect analysis, policy-centric analysis,performance.

V. CASE STUDIES OF SOCIAL SECURITY DATA MINING PRACTICES

A. Australian Practices

Australia is one of the most advanced countries in terms of in-venting SSDM to enhance decision support systems. Currently, the main techniques and methods for decision support consist of data matching, service profiling, business intelligence, and data analytics.

1) Data matching refers to the process and tools to match social security data against other sources of data, such as from immigration, customs, taxation, and banking systems.

2) Service profiling profiles customers associated with different services.
3) Business intelligence usually refers to data warehousing and reporting, including *ad-hoc* and online analytical processing-based analysis.

4) Data analytics covers a broad scope from descriptive analysis to data-driven pattern and anomaly detection and analysis.

VI. CONCLUSION

1) While many existing algorithms and models seem to be suitable for SSDM and public-sector data mining, it is necessary to carefully check that they fit perfectly into the public-sector data characteristics. Variations may be necessary. A typical situation is that many service patterns are identified, but most if not all of them simply reflect policy and process arrangements, which is not helpful for government service decision making.

2) While different methods can be used and are helpful, ser-vice profiling and domain-driven decision rules are most convenient and effective for public service decision making.

Although public-sector data mining shares character is-and challenges of SSDM, based on our experience and know tics and needs with general business applications, public- edge accumulated through conducting data mining in Australian sector data provide great opportunities to update existing techniques as well as to develop new approaches, in aspects such as customer–government interactions, mixed information from political, economic, sociological, and technical perspectives, from mining single data sources to multiple sources, and from focusing on single business lines to crossing business in relevant service departments. For instance, negative sequence analysis, becomes very useful in detecting those customers who social security data.

We have also highlighted several case studies of mining social security data, With 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 demon-started 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, 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.

REFERENCES

[1] H. Aaron, “Demographic effects on the equity of social security benefits,” in *The Economics of Public Services*, M. Feldstein and R. Inman, Eds., London: Macmillan, 2007.

[2] B. Agarwal, "Social security and the family: Coping with seasonality and calamity in rural India," *J. Peasant Stud.*, vol. 17, pp. 341–412, 1990.

[3] L. Alexander and T. Jabine, "Access to social security microdata files for research and statistical purposes," *Soc. Secur. Bull.*, vol. 41, no. 8, pp. 3–17, 1978.

[4] A. J. Auerbach and L. J. Kotlikoff. (1984). "An examination of empir-ical tests of social security and savings," National Bureau of Economic Research, Cambridge, MA, Working Paper 730. [Online]. Available: http://www.nber.org/papers/w0730.

[5] A. J. Auerbach and L. J. Kotlikoff. (1985, Oct.). "Simulating alternative social security responses to the demographic transition," National Bureau of Economic Research, Cambridge, MA, Working Paper 1308. [Online]. Available: http://ideas.repec.org/p/nbr/nberwo/1308.html.

[6] D. Baker and M. Weisbrot, *Social Security: The Phony Crisis*. Chicago, IL: Univ. of Chicago Press, 2000.

[7] 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. [Online]. Available: http://ideas.repec.org/p/nbr/nberwo/2257.html.

[8] R. J. Barro and C. Sahasakul, "Average marginal tax rates from social security and the individual income tax," *J. Bus.*, vol. 59, no. 4, pp. 555– 566, 1986.

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

[10] D. Blanchet and L.-P. Pele. (1997, Oct.). "Social security and retirement in France," National Bureau of Economic Research, Cambridge, MA, Working Paper 6214. [Online]. Available: http://ideas.repec.org/p/nbr/ nberwo/6214.html

[11] D. Bloom, D. Canning, R. Mansfield, and M. J. Moore. (2006). "Demo-graphic change, social security systems, and savings," National Bureau of Economic Research, Cambridge, MA, Working Paper 12621. [Online]. Available: http://econpapers.repec.org/RePEc:nbr:nberwo:12621

[12] R. W. Boadway and D. E. Wildasin, "A median voter model of social security," *Int. Econom. Rev.*, vol. 30, no. 2, pp. 307–328, 1989.

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

[14] 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.

[15] L. Cao, "In-depth behavior understanding and use: The behavior infor-matics approach," *Inf. Sci.*, 180, no. 17, pp. 3067–3085, 2010.

[16] L. Cao *et al.*, *Social Security Data Mining for Public Services*. [Online]. Available: http://datamining.it.uts.edu.au/icdm10/index.php/case-study

[17] L. Cao, D. Luo, and C. Zhang, "Knowledge actionability: Satisfying technical and business interestingness," *Int. J. Bus. Intell. Data Mining*, vol. 2, no. 4, pp. 496–514, 2007.