

Data Mining on Social Security and Social Welfare

Sphoorti Yaliwal¹, Pooja Yaliwal², Namita Mirjankar³U.G. Student, Dept. of Computer Application, KLES's BCA P.C.Jabin Science College, Hubballi, India¹U.G. Student, Dept. of Computer Application, KLES's BCA P.C.Jabin Science College, Hubballi, India²Lecturer, Dept. of Computer Application, KLES's BCA P.C.Jabin Science College, Hubballi, India³

ABSTRACT: The significance of social security and welfare business has been recognized in more and more countries. It has an impact on huge share of the population and influences government service policies and people's life quality. Typical welfare countries, like Australia and Canada, have accumulated an enormous quantity of social security and welfare data. Rising business problems like deceitful outlays, and client service and performance enhancements challenge existing policies, including techniques and systems as well as data matching and business intelligence reportage systems. The necessity for a deep understanding of client and client-government interactions through advanced information analytics has been progressively recognized by the community at massive. But so far, however, no significant work on the mining of social security and welfare information has been stated. To the best of our information, this paper attracts a comprehensive overall image and summarizes the corresponding strategies and illustrations to investigate social security and social welfare data, predominantly, social security data mining (SSDM), based on an intensive evaluation of a huge number of associated references from the past half century. Typically, we introduce associate degree SSDM framework, as well as business and research problems, security and social welfare data and services, as well as challenges, goals, and tasks in mining security and social welfare data.

KEYWORDS: Data Mining, social security, social welfare, social security data mining (SSDM)

I. INTRODUCTION

Data mining is the method of studying data from various perspectives and abridges it into useful facts. Technically, data mining is the system of locating correlations or patterns amid dozens of fields in huge relational databases. Because of the advancements in many fields like statistics, artificial intelligence, pattern recognition and computation algorithms, data mining has reached a new height. Data mining has a large number of applications but this paper considers the importance of the data mining on social security and social welfare data. Because all the members of society have the right to social security, nowadays the significance of social security and welfare business has been increasingly recognized in number of countries. Most of our personal data are transformed from paper documents to digital form. Australia is one of the most developed social welfare countries in the world in terms of government policies, infrastructure, the populace of benefit recipients, and the advancement of social security techniques and tools. Therefore, in this paper, the data is collected from the social welfare services in Australia.

Mining social security and social welfare data is demanding. Social security data are exceptionally unpredictable, including all the significant issues that are examined in the data quality and engineering field, such as sparseness, dynamics, and distribution. There are many factors that contribute to challenges faced by data mining of social security data, like particular business objectives in social security, government objectives, communications between clients and government officers, client behavioural statistics, and general difficulties in taking care of big business information etc.

Numerous analysis has been conducted from political, economic, sociological, and regional perspectives, when compared with a much slower effort made on the technical perspective. The primary concern involves problem evaluation, process and policy modelling, business evaluation, correlation evaluation, infrastructure development, and emerging data-driven evaluation. The dominant truth and trend of policy and economy oriented research are, a very limited studies can be discovered within the literature on mining social security and social welfare data. Social Security

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

Data Mining (SSDM) attempts to discover interesting patterns and exception in social security and social welfare data. Its goals are to handle exceptional business objectives, along with debt prevention. From the data mining task perspective, it includes traditional data mining plan, for example, classification, and the need to invent advanced methods, e.g., complex sequence analysis.

Australia is one of the most evolved social welfare countries in terms of infrastructure, government policies, the populace of benefit recipients, and the advancement of social security tools and techniques. Many series of projects have been carried out on the data mining for the Australian Commonwealth Government since 2004. Systems, models and algorithms to discover key drivers, factors, exceptions and patterns, indicating high danger of customers, customer situation adjustments, declarations, and interactions between government officers and customers have been developed. The findings have confirmed to be very useful for recovery, prediction, overpayment prevention and deep understanding of customer activities and intervention, the prevention and recovery of overpayments for the government. In this paper instead of focusing on a particular SSDM technique, it aims to draw an overall picture of SSDM through sharing our observations, experience and lessons learned in both reviewing the related work and carrying out real-life SSDM functions. In particular, the main contributions consist of 1) a radical literature overview of social security research in the past half century, and discussion of various categorizations of the associated work; 2) a complete framework of SSDM, discussing the main data mining tasks, important challenges in mining patterns in social security data and goals; 3) a summary of several case research, which includes the development of latest and powerful algorithms and tools to handle social security data. Mainly, we focus on the work on mining debt-targeted patterns, such as debt focused positive and negative sequences; and 4) the extension of discussions about mining standard public area data.

This paper summarizes the related work on social security research in the past 50 years and establishes social security business and data characteristics. It also outlines a framework for SSDM, including the challenges, main goals and tasks in mining social security data. It also discusses public sector data mining constructed on the lessons accomplished in managing SSDM in Australia.

II. REVIEW ON SOCIAL SECURITY AND SOCIAL WELFARE RESEARCH

The review includes the analysis of data collection. The data mining is an important task in the knowledge discovery process. To perform the data mining, the data is collected from different interpretation. In which some are more important than others are. The technical interpretation data collection has an important role in this era. After carrying out the data mining, some estimation is made.

A. Comprehensive Concept

Research on social security and social welfare problems was commenced in the mid-20th century. Since then, broad based problems were delivered to the research and can be classified into the following essential streams. There are different views for data collection. By collecting data in this interpretation, it is easy for data selection.

i. Political interpretation:

One of the main streams of research investigates the impact, factors, and problem of social security and welfare from public policy, social policy, management, and governance, resistance, and exercise viewpoints.

ii. Economic interpretation:

Another dominant truth and trend is the exploration of issues and the effect of social protection fashions and elements from the point of view of econometrics, public economics, and political financial system. This involves evaluation and discussions about economy, earnings, rating, financial savings, growth, privatization, reform, labor supply, multi entity relationship analysis, and premier arrangements.

iii. Sociological interpretation:

A couple of specialists are worried about the social effect of social security regulations on our society, for example, lifecycle, explanatory, conduct, maturing, retirement, misrepresentation, reasonableness and moderateness etc.

iv. Regional interpretation:

Researchers from social security exclusive nations introduce the improvement of in their countries, for instance, Canada, India, Latin America, the United States, Britain, Sweden, China, Italy, Germany, France, and Europe.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

B. Technical interpretation

A rising pattern in social security is the investigation of specialized issues, e.g., Improving infrastructure, knowledge administration, modeling of strategy and procedure, analyzing data driven patterns, and connection assessment crossing numerous zones.

From the technical interpretation, the principle issues that have been addressed inside the literature attention on numerous regions, such as trouble analysis, process and policy modelling, business oriented evaluation, correlation analysis, infrastructure guide, and statistics-driven analysis.

i. Problem evaluation:

Every so often, we find papers discussing or debating the issues of reform, disaster, and troubles for policies, privatization, uncertainty, optimization, fraud, and effect on economic system, society, capital marketplace, human resources, and many others.

ii. Method and policy modelling

Distinct strategies, e.g., empirical analysis, time-series analysis, quantitative comparative analysis, and equilibrium analysis, had been used and developed to layout, simulate, and examine coverage, pension, benefit, procedure and their consequences, as well as their optimization, choice, and performance score such as accuracy.

iii. Enterprise-oriented analysis:

Key commercial enterprise problems, which includes income and earnings, fee, benefit claiming, behaviour, retirement, hazard, saving, etc., are studied from political, monetary, sociological, and technical views.

iv. Correlation evaluation:

The relationship among social security and different economic systems have been studied; as an example, the relationship with fitness affairs, Medicare, taxation, stock and market, in addition to with political shape, financial development, labour pressure, and human capital.

v. Infrastructure support:

Discussions on building IT systems, supporting the analysis of social safety statistics, simulating and optimizing tactics, policy, overall performance fees, and many others have been conducted.

vi. Data-driven evaluation:

Currently, the cost of information and facts-driven choices has been increasingly more diagnosed. various analysis methods are being developed to research “what do social protection information show”, e.g., to pick out drivers, enablers, carrier styles, linkages, demographic, behaviour, change, statistics measures and composites, fraud, and chance adjustment from non-random choice, stochastic forecast, sequential analysis, time-series, equilibrium analysis, records mining, knowledge management, micro estimation, and e-government factors.

III. SOCIAL SECURITY SERVICES AND DATA

A. Social Security business

There is a wide variety of social security services furnished by each authority to assist people to become independent and to support the ones in need. In addition, each service carried out through different steps in order that large amount of statistics is created with each service. Fig. 1[15] illustrates a purpose–impact relationship among customers and government in the social welfare business. A customer lodges an allowance application, that's checked by the government. Payments are organized on the premise of client entitlement and policies. The customer is needed to confirm any changes which can affect payment entitlement. Once a customer statement is lodged, it will likely be verified by means of the authorities. As a result, client bills are further verified and changed if required. In some instances, overpayments to the customer can also arise for motives such as wrong declaration. The authorities who are looking to recover the debt request the customer to pay back such overpayments through compensation arrangements made among authorities and the customer. Detailed information about social welfare business can be found on the various government web sites and in reports

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

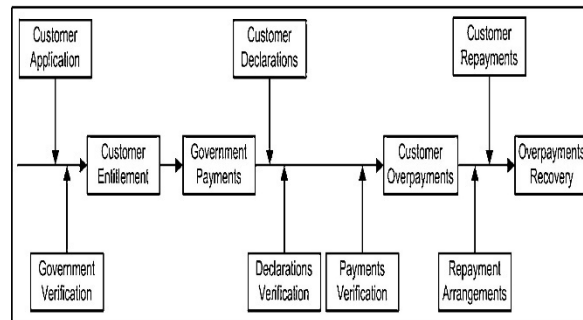


Fig. 1. Social security business workflow.

For the data mining process, the statistics accumulated from Centrelink. Centrelink is software in the branch of Human Services that gives you a range of payments and services for retirees, the unemployed, families, carers, parents, men with disabilities, Indigenous Australians, and those from numerous cultural and linguistic backgrounds. In this paper the discussion about an easy business workflow is executed. Due to imposing everyday social security services, a massive amount of data has been accumulated which will increase dramatically every day.

B. Social Security Data

The social security and social welfare region carries a massive amount of information. So it's far very hard to perform on the data. Table 1 [15] presents an overview of some Centrelink enterprise dimensions associated with data. As Table 1 shows, the huge quantity of social security information accrued via Centrelink encompass very beneficial statistics recorded from customer service centres, marketers and access points, the internet, interviews and critiques 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 information may be classified into the following eleven classes: 1) Client demographic and condition data, 2) Benefit/allowance data, 3) consumer pathway statistics, 4) Activity data, 5) Facility usage information, 6) Service policy data, 7) Service transactional data, 8) Service performance data, 9) Interaction data, 10) Operation data, 11) Operational performance.

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

TABLE 1 CENTRELINK Business Dimensions 2008–2009

IV. FRAMEWORK OF SOCIAL SECURITY DATA MINING

The main motivation behind data mining applications in social security is business targets and hidden data. After the thorough study of classification of data and its attributes, the creation of the fundamental frame work of social security data mining makes more effective. Dissecting the social security data mining objectives and difficulties the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

undertakings for the framework was made. In Fig.2[15], a high-level SSDM framework is shown. The main three layers of the fundamental frame work of social security data mining are: the data layer, the business objective layer, and the data mining goal layer.

The data layer abridges the primary data resources. It comprises of client data (client demographic and situation data), administration information (administration use and procedural data) and so on. The business objective layer incorporates the primary points and desires for the execution of social security services. Case in point, a client administration upgrade (to right away give top notch administrations to those with specific needs) and so on. The data mining goal layer records the fundamental objectives in mining social security data to upgrade business goals; for example, client driven analysis, payment-centric investigation, debt-centric analysis, income- centric analysis etc.

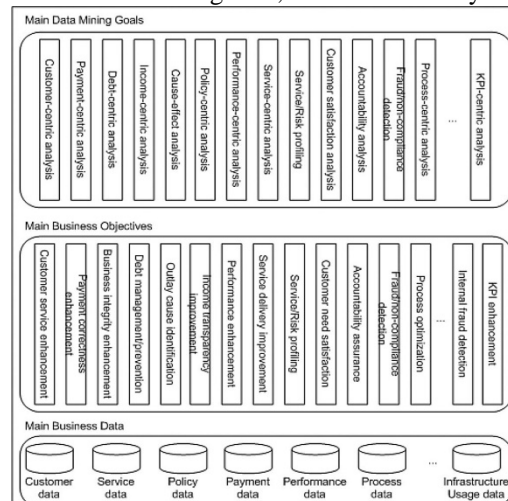


Fig. 2. SSDM framework.

The investigation of the way of the information uncovers that the social security data mining confronts a great deal of difficulties. The key difficulties faces are concerned with the principle methodologies of social security processing and pattern analysis as well as the delivery of proper knowledge. To bolster the data mining objectives numerous undertakings, should be taken in SSDM. The classification of the SSDM undertakings as indicated by the principle elements in social disability/welfare business and they are data processing, change detection, debt analysis and so on.

V. CASE STUDIES OF SOCIAL SECURITY DATA MINING PRACTICES

A. Australian Practices

Australia is a standout amongst the most progressive nations as far as inventing SSDM to improve decision support systems. As of now, the principle strategies and techniques for decision support systems comprise of information coordination, administration profiling, business knowledge, and information analytics.

1. Information coordination refers to the procedures and tools to match social security data against other sources of data. The data may be from immigration, customs, taxation, and banking systems.
2. Administration profiling is used for profiling clients related to different administrative areas.
3. Business knowledge for the most part alludes to data warehousing and reporting, including specially appointed and online scientific handling based investigation
4. Information analytics cover an expansive extension from descriptive analysis to information-driven pattern and abnormality recognition and investigation.

B. Different Methods Of Data Mining

A progression of researches and business ventures are directed with Centrelink in data mining. The most two critical issues related to Centrelink are prevention of debt and recovery from debt. These two critical issues are taken into account while data mining is performed. There are diverse systems and strategies for data mining. The important

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

five strategies of data mining on social security and welfare data are: Modelling effect of activity/activity sequence, mining impact target designs, Mining Positive and Negative consecutive patterns, Mining Attribute Combined examples and Identifying High Impact Behaviour for Intervention.

VI. CONCLUSION

All over the globe, almost all the countries have realized the importance of the social security and its applications, as the primary issue faced by the world at present is financial crisis. When we study the Australian social security thoroughly, we understood that data mining has a very vital role to play for the prevention and recovery of debt as well as for the client analysis so forth, amid the previous couple of years. Be that as it may, data mining related to government managed savings and open area information is still an open field for business applications in data mining and machine learning. With technical point of view in mind, few researches were carried out. So there is an expansive degree for researchers. These researches have helped many governments in making vital decisions regarding social security and administrations. Such sociology exploration could be precious for better welfare program organization, program assessment, and strategy investigation. In addition to this, we have been enlightened to observe that with the help of data mining, we can thoroughly analyse reliability as well as security of the social security systems.

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. (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>.
- [5] D. Baker and M. Weisbrot, *Social Security: The Phony Crisis*. Chicago, IL: Univ. of Chicago Press, 2000.
- [6] 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>.
- [7] 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.
- [8] H. Berghel, "Identity theft, social security numbers, and the web," *Com-mun. ACM*, vol. 43, no. 2, pp. 17–21, 2000.
- [9] 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>
- [10] 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.
- [11] 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.
- [12] L. Cao, "In-depth behavior understanding and use: The behavior infor-matics approach," *Inf. Sci.*, 180, no. 17, pp. 3067–3085, 2010.
- [13] L. Cao et al., *Social Security Data Mining for Public Services*. [Online]. Available: <http://datamining.it.uts.edu.au/icdm10/index.php/cas-e-study>
- [14] 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.
- [15] Longbing Cao, Senior Member, IEEE; *Social Security and Social Welfare Data Mining: An Overview*
- [16] Neethu Joseph, Aswathy Wilson; *Data mining on social security and social welfare data*
- [17] <http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/>
- [18] <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>