

Data mining on social security and social welfare data

Neethu Joseph, Aswathy Wilson

Abstract— Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the immediate need for turning such data into useful information and knowledge. The government faces new and intense pressure to collect and use personal data. The need for a deep understanding of public and public-government interactions through advanced data analytics has been increasingly recognized by the community at large. Mining Social security/welfare data is challenging. The challenges arise from business, data, and the mining of the data. Social Security Data Mining (SSDM) seeks to discover interesting patterns and exceptions in social security and social welfare data. The SSDM framework including business and research issues, social security/welfare services and data, and the different methods for SSDM framework.

Index Terms — Data Mining, social security data mining (SSDM), social welfare data mining

1 INTRODUCTION

Data mining is the process of analysing data from different perspectives and summarizing it into useful information. The issues in data mining are noisy data, missing values, static data, sparse data, dynamic data, relevance, interestingness, heterogeneity, algorithm efficiency, size and complexity of data. These types of problems often occur in large amounts of data.

Data mining have a large number of applications. Advancements in Statistics, Machine Learning, Artificial Intelligence, Pattern recognition and Computation capabilities have given present days data mining functionality a new height.

The main application areas of data mining are Bio-informatics and cure for diseases, fight against terrorism, web and semantic web, business trends, fraud detection and management, scientific applications etc. [1, 2] There is many application areas. But in this seminar consider the importance of the data mining on social security and social welfare data. Because everyone as a member of society has the right to social security. Now a days the importance of social security and welfare business has been increasingly recognized in more and more countries. In our country itself, most of our personal data

converted from paper documents to digital form. [1] 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. So in this paper, the data collected from the social welfare services in Australia.

2 TARGET AUDIENCE

The seminar is targeted at practitioners and researchers who are interested in the data mining, specifically in the area of social security and welfare.

3 SEMINAR OUTLINE

The seminar is structured into four sections. First the review on social security/welfare research then moving from Social security services and data to Framework of social security data mining. And last the different methods used for data mining.

3.1 Review On Social Security/Welfare Research

In the review on social security and social welfare data mining includes the analysis of data collection. The data mining is an important task in the knowledge discovery process [3]. To performing the data mining the data is collected from different perspectives. In which some are more important than others. The technical perspective data collection has an important role in this era. After performing the data mining some observations are made.

There is different perspectives for data collection. By collecting data in this perspectives it is easy for data

-
- Neethu Joseph is currently pursuing masters degree program in computer science and engineering in Calicut University, India, PH-9496171555. E-mail: neethujosephc@yahoo.co.in
 - Aswathy Wilson is Asst. Professor in computer science and engineering department in Jyothi Engineering College Thrissur, India, PH-9656175020. E-mail: aswathy@jecc.ac.in
(This information is optional; change it according to your need.)

selection. The different perspectives are Political Perspective, Economic Perspective, Sociological Perspective, Regional Perspective and Technical Perspective.

There is different perspectives for data collection. By collecting data in this perspectives it is easy for data selection. The different perspectives are Political Perspective, Economic Perspective, Sociological Perspective, Regional Perspective and Technical Perspective.

Studies on social security issues started in the middle of the 20th century. Till now most of the researches are done under the political, economic, sociological and regional perspectives. Only few researches are done in the technical perspective. In this seminar the technical perspective is taken into account. In the technical perspectives the main issues are problem analysis, process and policy modeling, business oriented analysis, correlation analysis, infrastructure support, and data-driven analysis.

3.2 Social Security Services And Data

There is a large number of social security services are provided by each government. And each services carried out through different steps. So huge amount of data is created with each service. For the data mining process the data collected from Centerlink. Centrelink is a program in the Department of Human Services that delivers a range of payments and services for retirees, the unemployed, families, carers, parents, people with disabilities, Indigenous Australians, and people from diverse cultural and linguistic backgrounds.

In this paper the discussion about a simple business workflow is performed. As a result of implementing day-to-day social security services, a huge amount of data has been accumulated which increases dramatically every day.

The social security and social welfare area contains a huge amount of data. So it is very difficult to perform on the data. If the data collection is based on some classification, then the analysing of data is little easy. The data have mainly eleven classifications.

The social security data have many characteristics. Each of them are important in different views. The knowledge about the characteristics of data, make the data mining process more efficient. There is eleven characteristics for social security data.

3.3 Framework Of Social Security Data Mining

The data mining applications in social security are driven by business objectives and underlying data. The creation of the basic framework of Social Security Data mining after studying the data classifications and characteristics makes more efficient. The basic framework

consists of three layers. Analysing the social security data mining goals and challenges the tasks for the framework created.

In this paper a basic framework of Social Security Data mining is provided that consists of three layers: the data layer, the business objective layer, and the data mining goal layer. 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) etc. The business objective layer includes the main aims and expectations for the implementation of social security services. For instance, a customer service enhancement (to instantly provide high-quality services to those with particular needs) etc. The data mining goal 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 etc.

The main data mining goals in the social security area into the following five classes, according to the understanding and practices of key entities, problems, and challenges in social welfare business by data mining: 1) over-payment-centric analysis; 2) customer-centric analysis; 3) policy-centric analysis; 4) process-centric analysis; and 5) fraud-centric analysis. Each of these classes contains different goals.

The study of the nature of the data reveals that the social Security Data mining faces a lot of challenges. The key challenges consider in terms of the main procedures of social security data processing, pattern analysis, and knowledge delivery.

To support the data mining goals many tasks need to be performed in SSDM. We categorize the SSDM tasks according to the main entities in social security/welfare business and to address the SSDM goals data processing, activity analysis, customer risk analysis, earnings analysis, change detection, debt analysis and fraud detection.

3.4 Different Methods Of Data Mining

A series of researches and commercial projects are conducted with Centerlink in data mining. The debt prevention and debt recovery are two of the most important issues in Centerlink. So based on these two concepts the data mining is performed. There is different techniques and methods for data mining. The data mining on social security and welfare data prefer five methods. That are Modeling impact of activity/activity sequence [3], Mining impact target patterns [3, 4], Mining Positive and Negative sequential patterns [1], Mining

4 CONCLUSION

Data mining play a critical role in social security for debt prevention, recovery and customer analysis. Mining social security and public sector data are still an open field for business applications in data mining and machine learning. There is few researches are carried in technical perspective. So there is a large scope for researchers. The researches in the social security and social welfare data helps the public and government to take important decisions about social security services. Such social science research could be priceless for better welfare program administration, program evaluation, and policy analysis.

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. The reliability and security of social security systems could be investigated and enhanced using data mining.

REFERENCES

- [1] L. C. H. B. Y. O. Y. Zhao, H. Zhang and C. Zhang, "Data mining applications in social security," New York: Springer, pp. 81-96, 2009.
- [2] D. K. 1Professor and G. J. U. o. S. . T. H. D. B. R. S.
- [2] D. o. C. S. . E. G. J. U. o. S. . T. H. Dean, Faculty of Engineering & Technology, "Rise of data mining: Current and future application."
- [3] Y. Z. L. Cao and C. Zhang, "Mining impact-targeted activity patterns in imbalanced data," IEEE Transactions, vol. 20, pp. 1053-1066, Aug. 2008.
- [4] K. F. H.-C. M. Kum, D. Duncan and W. Wang, "Social welfare program administration and evaluation and policy analysis using knowledge discovery and data mining (kdd) on administrative data," Nat. Conf. Digital Govern. Res, pp. 1-6, 2003.
- [5] Y. Z. D. L. L. Cao, H. Zhang and C. Zhang, "Com- bined mining: Discovering informative knowledge in complex data," IEEE Transactions, vol. 41, pp. 699-712, june,2011
- [6] L. Cao, "In-depth behavior understanding and use: The behavior informatics approach," Inf. Sci., 180, no. 17, p.3067-3085, 2010.