

Study of Key Technologies of Insurance Claims Based on Data Mining

SU Wei^{1, a}

¹ Qinghai university of finance and economics college, Qinghai Xining 810001, China

^asuweiqh@163.com

Keywords: Technology, data mining, insurance claim

Abstract. As people's income level rises and the changes of consumer consciousness, the number of insurance product is increasing, the public enthusiasm for buying insurance is higher and higher, as a result the insurance industry is developing rapidly. At the same time, policymakers should notice when gaining huge interest that risks existing in the industry. In the insurance industry, in the study of earlier customer claims, there is little involved in the field of data mining, mostly using traditional statistical methods or purely professional analysis. Although these methods can find some surface features of claims, such as repeated insurance, frequent and high insurance etc., they may have ignored the implied and undiscovered valuable information in huge amounts of data, while data mining is a new method to study from different angles.

The general introduction of data mining

Data mining, a process of cycle, usually involves data preparation, modeling, evaluating and interpretation model, using and consolidating the model steps. Data preparation work includes: the selection of data, the exploration (understanding the data distribution and abnormal data etc.), modification and transformation etc.; building a model is to select algorithm that data mining tools provided and applied to the prepared data, select corresponding parameters and generate the model; evaluation and interpretation model is to compare and evaluate the model, generate a relatively "optimal" model and to explain with business language model; using and consolidating model is to monitor model's performance in practical application, if the model performance is good or not for further investigation and correction, to reflect changes in the business operation rule.

According to the standards of CRISP - DM (cross-industry data mining methodology), data mining application can be divided into the following six steps in the insurance industry: business understanding, data understanding, data preparation, modeling, model assessment and release.

First, commercial understanding: clear goals, namely to identify high risk compensate customer characteristics, and to guide the company's marketing decision-making;

Second, data interpretation: data has a total of 29 variables, variable types have category, Boolean and order categories, different types of variables corresponding to different problems to be solved;

Third, data preparation, on the one hand, the large and complex data preprocessing, excluding missing data, and to distinguish the target variables and explanatory variables; on the other hand, from the descriptive perspective to select high-risk customer statistics;

Fourth, the establishment of model: according to different goal of data mining and data characteristics, different mining algorithm is adopted to establish the model;

Fifth, model evaluation: compare with the result of the model verification, validation, support accuracy verification test to determine the value of the model, such as gain figure and classification matrix and validation;

Table 1: the fields of the clients data

bd_no integer	Single pipeline	sjlp_amt decimal
jz_date date	Clinic date	cy_date date
zjfu_amt decimal	Charge	kh_no integer
sex char	Sex	born_date date
grbe_amt decimal	Coverage	qb_date date
zrqm_date date	Period liability	gr_amt decimal(

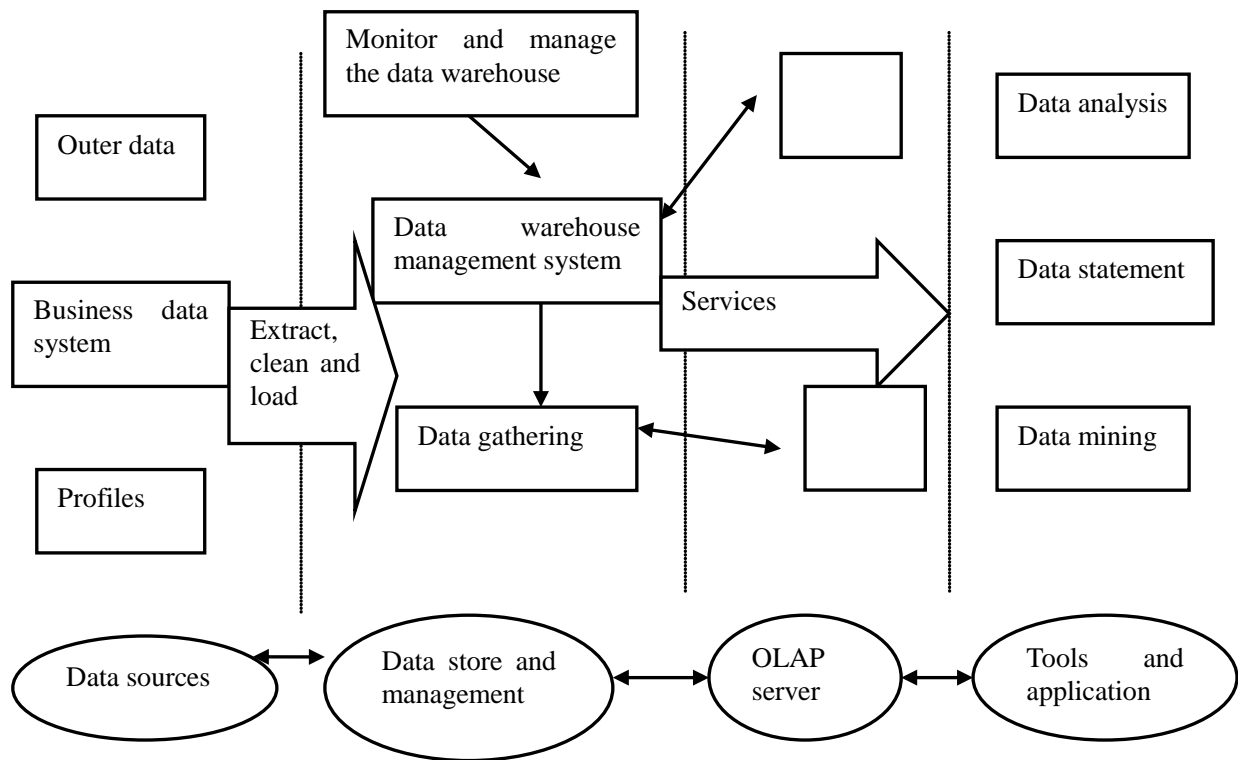


Chart 1: the structure of the data warehouse

Sixth, the model publishing: only to publish the model to policy makers can truly through data mining to reduce the probability of insurance company claims business and cost.

Life, auto, health, accidents, casualties and debt insurance by risk model to manage the price and reserve. Customer claims for medical insurance risk management, mainly through claims the customer data and underwriting customer information, find out the influence ginseng protect unit hospitalization rates, the average rate of outpatient or the cause of the claim amount, danger is planned for later development and provide scientific basis for rates.

Normally, the process including data collection, data description, select (subset) of the source data, data quality assessment and data cleaning, merger and integration build metadata set, load the library data mining and maintenance of library data mining.

Technology study of data warehouse

The data warehouse

Data warehouse system is a kind of decision supporting system, mainly includes data warehouse technology (on-line Analytical Processing technology, short as OLAP), Data Mining technology (DM). Data warehouse technology includes data warehouse modeling, monitoring and management of data warehouse, data quality analysis tools. The structure is shown as the chart 1.

The most fundamental feature of the data warehouse is physically stored data after extracting. The data is not the latest, proprietary, but comes from other databases. Data warehouse data is in the form of physical storage, mainly divided into two kinds: based on the relational database storage form and the multidimensional database storage.

General data warehouse can be divided into the following three types: enterprise data warehouse (EDW), operational database (ODS) and a data mart (Data Mart). EDW emphasizes the dimensional modeling technology, considering the dimensions of the relationships between different entities in the model. Dimensional model mainly includes star model, the snowflake model. In addition to the two main methods of model, bridge connection and bracket also can be used. The following chart 2 is the star model.

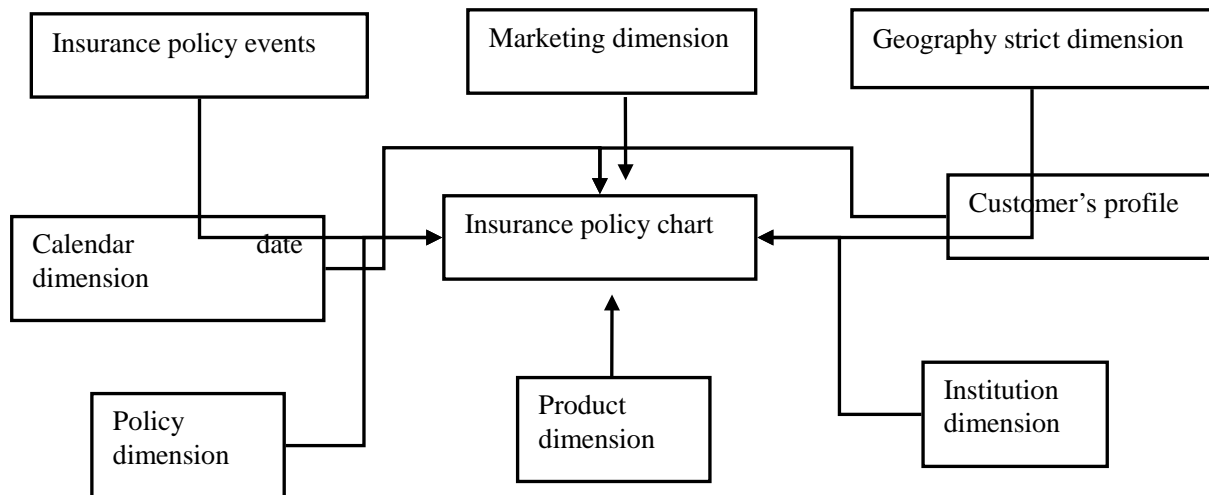


Chart 2: the chart of star model

OLAP

On-line analytical processing (OLAP) is the main application in data warehouse system, it supports the analysis of the complex operation, focuses on the decision support, and provides intuitive query results and is easy to understand. Basic idea of OLAP is to let the decision makers of the enterprise be able to flexibly control the data of the enterprise in the form of multidimensional observing the state of the enterprise and understand the change of the enterprise from various aspects and angles. OLAP system separates from the data source data storage with enough analysis data. When OLAP uses as a way of using their own, the data organization is the same as the data warehouse organization. When the on-line analytical processing (OLAP) combines with data warehouse, OLAP data source is derived from the data warehouse.

Data Mining

Data mining is to analyze the very large observed data set, and the purpose is to find the unknown relationship and to summarize the data that the data owner can understand its valuable in an innovative way. Data mining is based on the established business objectives and the existing problems of the enterprise, to make a study of a large number of business data revealing the hidden rule, and modeling, guided and applied to the actual enterprise operation, shown in the following:

$$\begin{aligned}
 OF &= \max(\text{days}(t)) \\
 &= \max(A \circ W^T) \\
 &= \text{Max} \sum a_i \cdot w_i \cdot \quad (1) \\
 \text{s.t., } a_i &\in [0,1] \text{ dsys}(t) \in (0,1), \\
 w_i &\in [0,1], \sum^n w_i = 1.
 \end{aligned}$$

Technologies of insurance claims

IAA model

IAA application is a set of complete model of the financial services industry. It includes: the demand model; the business model. The design of the components, interfaces, message model; product definition and contract management of general design framework; create a design model of data warehouse and the ability to generate XML and EJB component interfaces. Business relative to the technology is relatively stable, IAA as business model. After ten years of development, more than one hundred insurance companies use it and debug and enrich the content of the new in the world has, and the general structure of adaptability in a changing environment. IAA rules generalization through the contract management process in the product, can modified products without changing the application code more easily, in order to narrow the system to support the new product development time.

ETL technology

ETL is a filling, update the process of data extraction, transformation and the loading in the data warehouse.

Data extraction: extract data from the source data source system that the source system needs.

Data transformation: transform the data get from the source data according to business needs to the required form and clean and process the error, inconsistent data sources.

Data loading: the transformed data loads into the objective data source. The core process of establishing data warehouse system of ETL and its implementation is in accordance with the unified rules of integration and to increase the value of data. It is responsible for the complete transformation of data from the data source to the target data warehouse process, is the important process of the implementation of data warehouse.

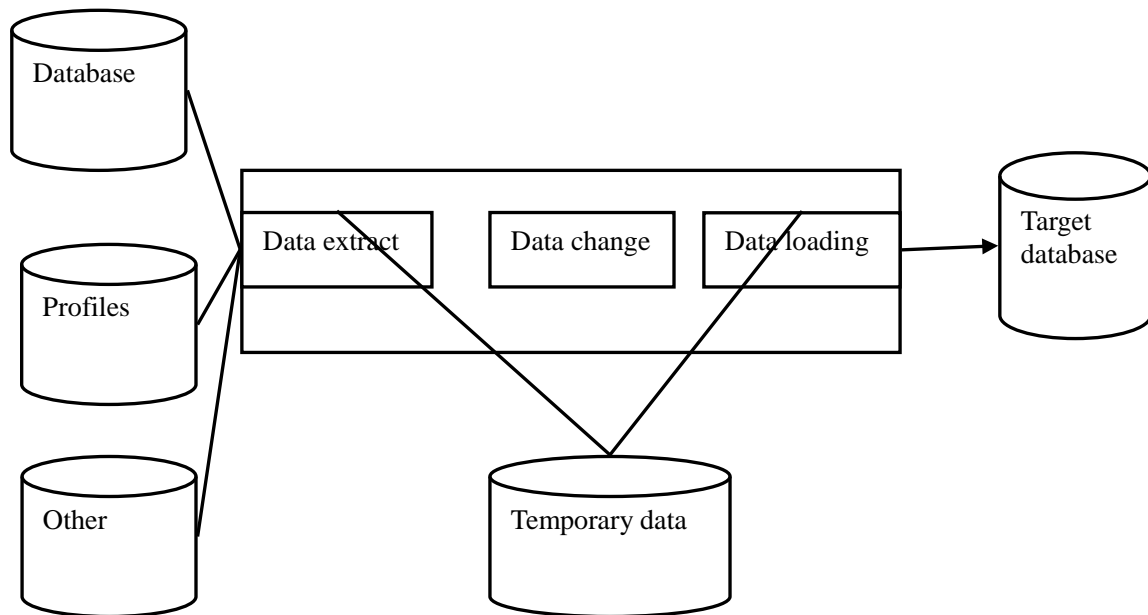


Chart 3: the ETL system structure

The main purpose of ETL is to transform the daily business operation data to the data into a data warehouse corporate global view with the minimum cost (including the impact on day-to-day operations and the demand for skills).

Summary

The medical insurance risk management based on data mining technology gives full play to the strong advantage of data mining technology in the face of huge amounts of data, scientifically helps insurance companies to solve the customer claim risk premium rate making business problem. And with the application of data mining technology in the insurance business, in the future it can also be applied to the insurance customer value management, make the insurance companies be better identified different group of customers, market and product etc., to discriminate different customers and insurance products, insurance market, adopt different customer strategy, expand the insurance market strategy and the strategy of developing new risks, to improve the core competitiveness of insurance companies.

References

- [1] Tian Zhaojin, Dai Wensheng, Xie Bangchang, Data Mining Application in the Insurance Claims[J]. China Statistics,2005,02.
- [2] Chen Tao, Ren Shiquan, The factors influencing medical costs and health insurance business risk control [J]. Finance and economics, 2002(2):87-90.

- [3] Luo Zhiqun, The application of data mining technology in China's mobile communications business research [D]. Business administration department of Hangzhou: Zhejiang university, 2002.
- [4] Liu Mingxia, Ren Shiquan, Risk factors analysis of the relative risk degree of medical treatment insurance [J]. Modern preventive medicine, 2004, (2): 159-161.
- [5] Li Yuquan, Opportunities and challenges of opening the mainland insurance market to the insurance industry in Taiwan [J]. Journal of China insurance, 2005, 03.