UDC 004.04

DATA MINING MODELS FOR HEALTHCARE

Kuatbayeva A.A[1]., Izteleuov N.E[2]., Kabdoldin A[3]., Abdyzhalilova R.[4]

IITU university, Almaty, Kazakhstan
[1]Ahamala2017@gmail.com, [2]izt.nurzhan@gmail.com,
[3]abylaikabdoldin@gmail.com,[4]xxx.raihan97@gmail.com

Abstract. In the healthcare sector data mining is becoming increasingly popular, because medicine organizations produce and collect large volumes of information day by day. Using data mining and data mining applications can help for each medicine organization and parts in the healthcare industry. For example, clients through data mining can take better services in healthcare, doctors better diagnose patients ' diseases and evolve a course of treatment. Of course, in these examples not the full range of power of data mining in the medicine sector, but main ones. Therefore we have problems like how all this system, also data mining is processed in general and using intelligent data analysis in our country. It is used for storing large amounts of data in different organizations and industries. The objective of this article is to show how the healthcare sector can benefit greatly from data mining and consider possible solution for problems, which described above.

Keywords: Data mining, data mining applications, healthcare sector, medicine industry, machine learning, algorithm.

Introduction

Currently, a large database is very important and the intellectual part is especially valuable. In short, this is all called data mining. It is used for storing large amounts of data in different organizations and industries, as well as for processing and analyzing the same data. This raises the question "Why is this necessary?". The answer to this question is simple, because in the 21st century traditional methods are not relevant. In the time where it requires a minimum of time and effort, as well as money, outdated storage methods are very labor-intensive. Data mining solves these issues. This also applies to the healthcare sector. It is known that new automated systems and data mining allow to evaluate medical and biological indicators of patients ' examination, diagnose diseases and create a treatment algorithm that later has a positive therapeutic effect. Improvements in data technology can improve the efficiency of treatment and diagnostic processes, as well as other parts of the medical industry.

Thus, there are two urgent problems: the problem of how all this system, also data mining is processed. Second problem of using intelligent data analysis in Kazakhstan. In the article possible solutions for these problems.

The purpose of this article is to reveal the possibilities of big data processing. The first part of the article examines the data mining process in the medical sectors and contains answers for the question "How does it work?" and etc. The second and third parts of the article suggest which industries can be used for data mining.

Data mining. methods and algorithm

At the beginning, it is necessary to understand what is meant by data mining. It is also called Knowledge Discovery In Data and studies the process of finding potentially useful, valid, and new knowledge in databases. In a broader sense, Data Mining refers to the concept of data analysis, which assumes that:

data may be inaccurate, incomplete (contain omissions), contradictory, heterogeneous, indirect, and this is why it is necessary to have huge volumes; therefore, understanding the data in specific applications require significant intellectual effort;

data analysis algorithms themselves may have " elements intelligence", in particular the ability to learn from precedents, that is, to draw General conclusions based on particular observations; the development of such algorithms also requires considerable intellectual effort;

the processes of processing raw data into information and information into knowledge can no longer be performed "manually" in the old way and sometimes require non-trivial automation.

Data mining is a multidisciplinary field that combines several branches of science and the main ones are artificial intelligence, statistics, and a database system (Fig. 1).

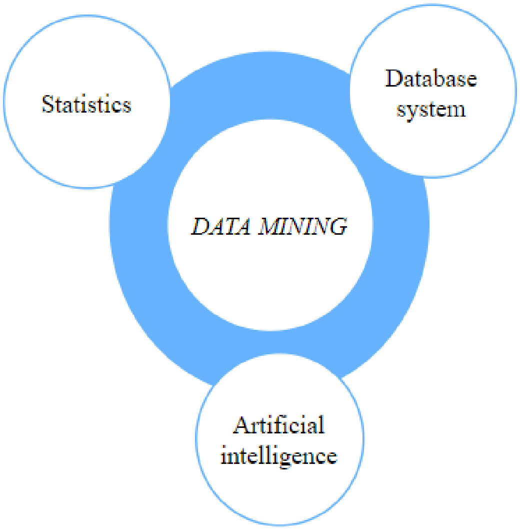

Figure 1 -.Illustration Data Mining that combines several facets of science

Sometimes there is a different character of multidisciplinary approach. This is when Data Mining is considered a combination of computer science, mathematics,and domain expertise. In this case, computer science describes the environment for creating information products, mathematics builds a theoretical basis for solving problems, and the concept of the subject area allows necessary to understand the reality in which there is a problem situation [28].

There are the following typical steps that accompany the solution of data mining problems:

Analysis of the subject area, formulation of research goals and objectives.

Extracting and saving data.

Pre-processing of data:

cleaning;

integration;

transformation.

Meaningful data analysis using Data Mining methods (establishing General patterns or solving more specific, specific problems).

Interpretation of the results obtained by presenting them in a convenient format (visualization and selection of useful data patterns, generating informative graphs and / or tables).

Using new knowledge to make decisions.

Data Mining methods and algorithms include the following: artificial neural networks, decision trees, symbolic rules, the method of support vectors, linear regression, hierarchical and non-hierarchical methods of cluster analysis, various methods of data visualization and etc. If we consider the Support Vector Machine (SVM classifier) method, the main idea here is to map the source vectors to a higher-dimensional space and search for a dividing hyperplane with a maximum gap in this space [29]. The essence of the standard classifier SVM for the case of two classes can be represented using the following expression:

$f(x, W) = \text{sign}(g(x, W))$, где $g(x, W) = < x, W> + b$,

where are the parameters W (weight vector) and b (free coefficient) defined by the training procedure. Solution boundaries classifiers g (x, W) = 0 represent a hyperplane of order L-1 in L-dimensional space (Fig. 2).

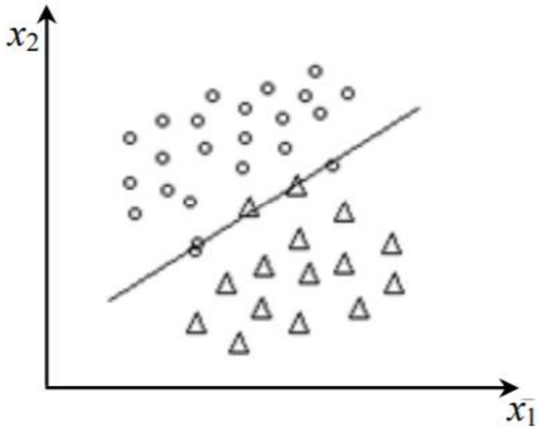

Figure 2 - Illustration of the SVM classifier for two-dimensional space

Solving the classification problem for the case of several classes can be implemented by multiple pairwise classification and combining results (for example, by the majority rule) [29]. The SVM classifier has a fairly high algorithmic complexity, but it also has high computational efficiency. In addition, it is characterized by high accuracy and robustness of results for various statistical characteristics of training data [28].

Data mining in healthcare industry

Healthcare encompasses comprehensive assessment, diagnosis and preventive mechanisms for infections, disabilities and other physical and behavioral impairments in humans [1]. For most nations, the healthcare system is developing at a fast rate. The healthcare field should be seen as a position of rich data because it produces vast volumes of data, including electronic medical documents, administrative reports and other benchmarking findings [2]. However, these safety results remain under-utilized. Data mining, as is recognized, is a non-trivial method of finding real, new, theoretically valuable, and essentially understandable trends in data through mixing, through copious data sets, trends that are too subtle or complicated for humans to recognize. This implies that data mining is able to dig for fresh and useful knowledge from such vast quantities of data. Data mining in health care is primarily used to forecast different illnesses, as well as to support doctors interpret their clinical decisions.

The data mining models can be used in the following healthcare industries: anomaly detection, clustering and classification.

Anomaly Detection

Anomaly analysis is used to find the most important shifts in the data set [3]. Bo Lie et al [4] used three separate forms of anomaly detection, normal support vector data definition, density-induced support vector data definition and Gaussian mixture to test the precision of anomaly detection on unknown dataset of liver disease data collected from UCI. The process is tested using the precision of the AUC. The findings obtained for the healthy data collection were 93.59 per cent on average. Although the total standard deviation from the same data collection is 2.63. The unknown dataset is likely to be present in all databases, anomaly detection will be a reasonable way to address this problem, but as there is only one paper about this approach, we can not say much on the usefulness of the process.

Clustering

Clustering Clustering is a common descriptive method that seeks to define a finite collection of categories or clusters to classify the data [3]. Rui Velosoa[5] used the vector quantization process in the clustering framework to estimate readbacks in intensive medicine. The methods used in the vector quantification process are k-means, k-medoids and x-means. The samples used in this analysis were compiled from the clinical and research findings of the patient. The test for each algorithm is carried out using the Davies-Bouldin Index. The k-means received the strongest outcomes, whilst the x-means received decent results, and the k-medoids got the worse performance. The findings of the study of these studies include a valuable finding in trying to identify the various categories of patients with a greater risk of readmission. A more important comparison of the approach can not be made, since this is the only paper I have addressed in my analysis on vector quantization.

Classification

Classification is the discovery of a predictive learning function that classifies a data item into one of several predefined classes [3]. There are several types of classification: statistical, discriminant analysis, decision tree, swarm intelligence, k-nearest neighbor, logistic regression, Bayesian classifier and support vector.

There is a great deal of scope for data mining technologies in health care. Specifically, they may be categorized as the measurement of medication effectiveness; patient service management; client experience management; and the prevention of fraud and violence. More advanced scientific data mining, such as precision medicine and DNA microarray research, is beyond the reach of this article.

Treatment effectiveness. Data mining software may be developed to determine the effectiveness of medical therapies. Through analyzing and contrasting the triggers, effects and courses of therapy, data mining will include an overview of the courses of action that tend to be effective [6]. Of example, the results of patients groups diagnosed with various medication regimens with the same illness or disorder may be measured to decide which medications perform better and are more cost-effective [7].

Along this line, United HealthCare has used the patient report details to discover opportunities to lower costs and provide quality medicines [8]. Medical profiles have since been created to provide physicians with details about their clinical habits and to equate them with those of other physicians and peer-reviewed professional guidelines.

Certain data mining uses relevant to diagnosis involve comparing the multiple side-effects of medication, gathering specific signs to aid diagnose, determining the most appropriate drug compounds for the diagnosis of sub-populations that react differently from the normal population to such medications, and finding preventive steps that may minimize the incidence of disease [6].

Healthcare management. To help monitor health care, data mining tools may be built to accurately classify and track chronic illness and high-risk patients, plan effective treatments, and minimize hospital visits and claims.

In order to establish effective diagnosis and treatment procedures, for example, the Arkansas Research Network looks at readmittance and resource use and contrasts its findings with existing clinical literature in order to find the right treatment choices, while utilizing facts to justify patient care [7]. The Health Cooperative Group often stratifies consumer demographics through social features and medical problems to determine which communities need the most services, allowing them to establish initiatives to better inform and deter or control such populations [7]. Community Health Alliance has been active in a variety of data analytics projects to improve treatment at reduced prices. Data analysis at the Seton Medical Center is used to shorten patient duration of stay, prevent surgical risks, establish best practices, optimize medical satisfaction, and offer guidance to physicians-all to sustain and increase the standard of health care [9].

Customer relationship management. Although consumer experience management is a critical approach to handling relationships between business organizations-usually banks and retailers and their customers-it is no less relevant in the healthcare sense. Customer experiences can occur via contact centers, medical offices, billing divisions, frustrated environments, and outpatient treatment environments.

As in the case of corporate companies, data mining technologies in the healthcare field may be built to assess the desires, habits of usage and existing and potential needs of consumers to increase their degree of satisfaction [10]. Such methods can also be used to forecast certain items that the health care consumer is willing to buy, whether the individual is willing to comply with recommended medication or whether preventive maintenance is likely to result in a substantial decrease in potential use.

Using data analysis, Patient Capacity Management Corp. has established a Consumer Healthcare Utilization Index that offers an indicator of an individual's willingness to utilize particular clinical facilities, identified by 25 main diagnostic categories, specified diagnostic associated classes or different medical sector areas [11]. The database, focused on millions of healthcare interactions by many million people, is capable of recognizing people who may benefit from the most relevant healthcare programs, enabling patients most in need of particular treatment to seek them, and constantly improving the platforms and messaging used to target effective markets for better safety and long-term patient connections and loyalty. The database was utilized by OSF Saint Joseph Medical Center to deliver the correct information and resources to the most relevant patients at critical times. The end result is more productive and secure contact and improved revenue [11].

Limitations and perspectives

Data processing technologies will be a great value to the healthcare sector. We are not without limits, though.

Healthcare data mining may be constrained by data usability, since raw data mining sources frequently occur in various environments and structures, such as management, hospitals, labs and more. Data must also be obtained and incorporated before data mining can take effect. Although some scholars and analysts have recommended that a data center be constructed before an effort is made to mine data, this may be an expensive and time-consuming undertaking. On a good note, Intermountain HealthCare has effectively developed a data archive from five separate sources — a research data server, an intensive care case-mix network, a testing information system, an ambulatory case-mix system, and a treatment program database — and used it to identify and incorporate effective evidence-based healthcare approaches. Oakley [12] recommended a hierarchical network topology instead of a data center for more effective data mining, and Friedman and Pliskin [13] reported a case study of Maccabi Healthcare Services utilizing established repositories to direct future data mining.

Data processing technologies in the health sector may have enormous scope and usefulness. However, the success of healthcare data mining depends on the availability of clean health data. In this context, it is important for the healthcare industry to understand how data can be best collected, processed, packaged and extracted. Potential avenues involve standardization of clinical terminology and exchange of data through organisations to maximize the value of health data mining applications.

Furthermore, since health details are not restricted to textual details, such as patient reports or hospital documents, it is therefore important to expand the usage of text mining in order to extend the complexity and essence of what health data mining will actually achieve. For particular, it is valuable to be able to combine data and text mining.36 This is often helpful to look at how visual imaging photographs can be incorporated with health data mining applications. Throughout these fields, some improvement has been made [14,15].

Conclusions

Data mining has played a significant role in the healthcare field, especially in the prediction of different types of diseases. Diagnosis is commonly used in illness detection, which is frequently employed in surgical diagnosis. In addition, there is no data mining tool for solving problems with safety data sets. In order to obtain the highest accuracy among the classifiers that are essential for the medical diagnosis with the characteristics of the data being examined, we need to develop a hybrid model that could solve the problems described. Our potential goal is to improve predictions utilizing hybrid models.

References

[1] J.-J. Yang J. Li, J. Mulder, Y. Wang S. Chen, H. Wu Q. Wang, and H. Pan. Emerging information technologies for enhanced healthcare. Comput. Ind., 2015. 69. 3–11.

[2] N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta. Knowledge Management in Healthcare. 2005. 63. 5–18.

[3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI Mag. 1996. 37–54.

[4] B. Liu, Y. Xiao, L. Cao, Z. Hao, and F. Deng. SVDD-based outlier detection on uncertain data. Knowl. Inf. Syst. 2013. 3(34). 597–618.

[5] R. Veloso, F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado. A Clustering Approach for Predicting Readmissions in Intensive Medicine. Procedia Technol., 2014. 16. 1307–1316.

[6] Milley, A. Healthcare and data mining. Health Management Technology, 2000. 21(8), 44-47.

[7] Kincade, K. Data mining: digging for healthcare gold. Insurance & Technology, 1998. 23(2), IM2-IM7.

[8] Young, J. & Pitta, J. Wal-Mart or Western Union? United HealthCare Corp. Forbes, 1997. 160(1). 244.

[9] Dakins, D.R. Center takes data tracking to heart. Health Data Management, 2001. 9(1). 32-36.

[10] Biafore, S. Predictive solutions bring more power to decision makers. Health Management Technology, 1999. 20(10). 12-14.

[11] Paddison, N. Index predicts individual service use. Health Management Technology, 2000. 21(2). 14-17.

[12] Oakley, S. Data mining, distributed networks and the laboratory. Health Management Technology, 1999. 20(5). 26-31.

[13] Friedman, N.L. & Pliskin, N. Demonstrating value-added utilization of existing databases for organizational decision-support. Information Resources Management Journal, 2002. 15(4). 1-15.

[14] Ceusters, W. Medical natural language understanding as a supporting technology for data mining in healthcare. In Medical Data Mining and Knowledge Discovery, Cios, K. J. (Ed.), PhysicaVerlag Heidelberg, New York, 2001. 41-69.

[15] Megalooikonomou, V. & Herskovits, E.H. Mining structurefunction associations in a brain image database. In Medical Data Mining and Knowledge Discovery, Cios, K. J. (Ed.), Physica-Verlag Heidelberg, New York, 2001. 153-180.

[16] Uskenbayeva R., Moldagulova A., Mukazhanov N.K., Creation of Data Classification System for Local Administration. Advances in Intelligent Systems and Computing. 2020

[17] Uskenbayeva R., Sabina R., Aigerim B., Managing Business Process Based on the Tonality of the Output Information, Advances in Intelligent Systems and Computing

[18] Uskenbayeva R.K., Kuandykov A.A., Rakhmetulayeva S.B., Bolshibayeva A.K. Properties of platforms for the transformation and automation of business processes, International Conference on Control, Automation and Systems, 2019.

[19] Moldagulova A.N., Uskenbayeva R.K., Satybaldiyeva R.Z., (...), Kalpeeva Z.B., Bektemyssova, G.U., On identification of hybrid business processes for effective implementation in the form of cloud services. International Conference on Control, Automation and Systems, 2019.

[20] Uskenbayeva R., Kuandykov A., Kalpeyeva Z., Kassymova A. Formalization of Applications for Processing in the e-Commerce System 2019 Proceedings - 21st IEEE Conference on Business Informatics, CBI 2019

[21] Mikhalevic, I.F., Ryjov A.P. Augmented intelligence framework for protecting against cyberattacks Proceedings - 5th International Conference on Engineering and Telecommunication,

EnT-MIPT 2018.

[22] Ryjov A. Personalization and optimization of information retrieval: Adaptive semantic layer approach. Proceedings - 2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016

[23] Kuatbayeva A.A. Modelling situational room for healthcare. World Applied Sciences Journal, 2014.

[24] Kuatbayeva A.A. Fuzzy logic in healthcare situation room modelling. ACM International Conference Proceeding Series, (2014).

[25] Zamyztin A.V. Data mining. Tomsk, 2016. 120p.

[26] Wang L. (ed.). Support vector machines: theory and applications. Springer Science & Business Media, 2005. 177. 434 p.

[27] Hian Cbye Kob and Gerald Tan, Data Mining Applications in Healthcare, 2005

[28] Neesha Jothia, Nur Aini Abdul Rashidb, Wahidah Husainc. Data Mining in Healthcare – A Review, 2015.