International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1291-1295 www.ijaem.net ISSN: 2395-5252

Novel Mechanism for the Regional Health Disease Trend Analysis Model

Honey Jain, B.E IT, Pune
Tanmayee Tushar Parbat, B.E IT, Pune
Rohan Benhal, BBA IT, Pune

Submitted: 10-08-2021 Revised: 22-08-2021 Accepted: 25-08-2021

ABSTRACT

The most significant component used to identify and file health insurance prices is medical trend. Trend analysis is used by insurance companies to anticipate future expenditures and premiums. Medical trends are used by governments in the rate-setting process. In this study, we look at four statistical methods for finding trend factors: average ratio, linear regression, exponential regression, and time series analysis. On the basis of leave one out analysis, an effective strategy for spotting outliers is described.

Keywords:Exponential regression, regional health disease, Trend Analysis Model

I. INTRODUCTION

Medical trend analysis is mostly used to estimate future medical costs or claims. Forecasting can be used by insurance firms to calculate future health insurance premiums. Administrators can use the data to decide if the premiums charged by health insurance are appropriate. It can be used by governments to keep track of health-care systems. For trend analysis, a variety of methodologies have been developed. These methods aren't just for medical trend analysis; they may be applied to a variety of fields, including climate change research [3], quality control of healthcare advances, cost of care trends, medical intervention evaluation, and so on. Linear regression is a well-known statistical technique for forecasting and prediction. [4] summarises the use of linear regression in trend analysis. The assumption in linear regression is that the detection and estimation have a linear relationship. The least square approach or the smallest absolute deviation method can be used to compute the Ls relationship. The first technique is more popular and has the advantage of being simple to apply because it solves a linear system, whereas the second method is more resilient in circumstances when the data set contains outliers. For trend analysis, there are several extensions to the linear regression method that can be employed. Generalized linear models include approaches such as logistic regression and exponential regression. Instead of the original variables, these techniques presume that the altered response and explanatory factors have a linear connection. If the values of the variable are projected to expand exponentially, it is more realistic to simulate the incremental pattern of the variable under examination (i.e., at a stable rate from period to period). The Ls approach has been applied to the study of climate and weather data [3, 5], as well as maternal and child health insurance. When Rosenberg looked at the trend in child health insurance, he used another regression model, Poisson regression. The response variable is assumed to have a Poisson distribution. Poisson regression, as opposed to standard least squares regression, has the advantage of being able to account for both the fluctuatLon throughout time and the variability at each time point.

Associated work

L. Anying and colleagues [1] In recent years, Chinese society has become increasingly informatized, and all parts of social life are becoming unable to function without it. The medical profession information provides consumers seeking medical counsel with previously unknown convenience, and this study focuses on the regional medical disease prediction approach. The original data suppliers offer data, we gather and collect the data, execute data mining and analysis in a big data computing environment, and finally produce the disease trends forecast report with the use of a big data processing platform.

M. S. Islam, et al. [2] used a zero-inflated Poisson regression model to see if meteorological variables are connected to the number of COPD admissions. This might help you better grasp the factors that influence the prevalence rates of various diseases. Our preliminary findings suggest the need for improved approaches to investigate regional inequities in our healthcare system as a result of weather change.

Using a continuous density HMM that incorporated the intrinsic nature of disease transmission for alert level prediction, X. Zhou et al [3] predicted epidemic alert levels. Respective models are being developed to track both the national and regional levels of epidemic alert in the United States. The proposed method can deliver real-time surveillance results weeks before the CDC releases its reports. Although the focus of this study is on monitoring infectious disease in the United States, we believe that a similar approach might be used to monitor epidemics in poorer nations as well.

W. Lin and colleagues [4] Three significant predictors of thallium scan service usage are the lowest temperature, average moisture, and the number of visiting outpatients in the cardiovascular division. For hospitals and healthcare organisations, this information can be used as a strategic reference in the management of equipment and human resources.

A. Bansal and colleagues [5] Corona virus, commonly known as COVID 19, is a serious epidemic that is spreading over the world. Italy and China have been identified as two of the key epicentres where the epidemic took full effect. As a result of COVID-19, the greatest death rates in the globe are recorded here. The United States, as one of the top countries, has also been among the countries with an increasing number of COVID 19 instances. The ARIMA model, or auto regressive integrated moving average model, is employed in this work to forecast the pandemic trajectory through time (i.e., April 2020).

P. Shankar et al [6] investigate Machine Learning and Natural Language Processing techniques for spotting agricultural discussions on Twitter, then analysing, categorising, and grouping them so that a stakeholder can use them to identify crop disease incidences, patterns, and trends at the regional scale. Current systems that search for agricultural diseases using keywords do not necessarily return agriculturally related tweets, and those that do potentially include a wide range of topics.

II. PROPOSED METHODOLOGY

To calculate an accurate outcome, a solid data collection must be obtained before the forecasting process can begin. Insurance companies or the government can both provide useful information. The past and anticipated experiences should, ideally, come from the same source. Furthermore, if aggregate trend analysis is undertaken, the most appropriate data should come from groups with the same policy or groups with similar policies. The length of the event is the most important factor to consider when considering time period. It is customary to look back 36 to 48 months. Shorter intervals offer a number of drawbacks. First, there are fewer data points in a shorter time span, resulting in inaccuracy and more unpredictability. Second, seasonal tendencies may appear to be long-term because a short period cannot demonstrate behaviour that is longer than the seasonal period being studied. Long-term data, despite the law of large numbers, which implies that more data points equal more exact results, has problems. On the one hand, long-term data can accurately represent recent data, but with the occurrence of severe economic downturns, long-term data cannot accurately represent recent data, causing more error in the forecasting result. Earned premiums, incurred claims, and memberships should all be included in basic trend analysis for health review purposes. The procedure of obtaining health insurance is extremely hard. A full study necessitates policy specifics as well as examination of a variety of elements. However, because to commercial confidence Lal considerations, insurance companies are usually hesitant to divulge too many details. Medical trend analysis, on the other hand, is used in the health rate review process to give a reference point for determining the acceptability of a submitted premium increase request, not to duplicate the insurers' work. As a result, the data is believed to be if claims and premiums may be estimated on a per member per month (PMPM) basis for basic trend analysis. e memberships are not directly employed in the analysis, but they are useful in interpreting the analysis's reliability. The number of members covered by the company is reflected in the term memberships. Small membership coverage, such as an enrolment of less than 5000, will typically have higher fluctuations, raising variability and causing in analysis. To avoid lowering reliability, it will require more complex statistical methods in analysis. In medical insurance rate, two types of data are most commonly used. The policy experience, which includes the number of enrolments, premiums, and claims for each month, is included in the monthly data. Each calendar year's policy experience is summarised in the annual data. When the data is yearly, the annual trend factor is calculated using trend analysis. The monthly trend is provided by the trend analysis when the data is monthly. The following

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1291-1295 www.ijaem.net ISSN: 2395-5252

transformation could be used to report the annualised trend: As previously stated, some experience periods for medical data are too extensive to provide current significant trend information. When a corporation files a premium increase request, some data collection processes are not completed completely. It's also possible that the data sets were created by chance. The data will contain enough errors to prevent an appropriate trend rate estimation calculation. As a result, finding a simple approach to detect these outliers is an important part of trend research. The average ratio approach and the time series method, on the other hand, do not allow for the removal of outliers because they require consecutive data points. Computing the regression curve and studying the residuals using a number of methods outlined in [14] is a statistically plausible method for detecting outliers in the regression method. However, this method has an unavoidable flaw: outlier-related mistakes are built into the data's simulated behaviour. Using a regression curve that is defective owing to outliers may result in inaccurate outlier detection. We present a leave one out detection strategy for detecting outliers in a data collection. We perform the regression without that data point, re-predict the value at that time point with the regression curve, compute the difference between the prediction and the original value, and express it as a multiple of the regression's standard error. The multiple is a positive number when the original value is more than the prediction, and a negative number when the original value is less than the prediction. If the multiple falls outside of the range 1.96 to 1.96, the point is considered an outlier with 95 percent confidence. When there are numerous outliers, we only delete the most severe one to prevent a less severe one from being wrongly recognised due to outlier error. The Ls data-cleaning method is repeated until no outlier le remain. The advantage of this strategy is that each regression may be run without the prior data set's outliers. The accuracy of the final regression and the resulting model will improve as the number of outliers is reduced.

For the experiment, the proposed method uses the Python 3.7 tool. The tests are carried out on an Intel Core i7 processor with a CPU usage rate of 2 GHz and 48 GB of RAM. The classifier is given the training data, and the trained data is compared to the testing data. Sixty percent of the data is used for training, whereas forty percent is used for testing. The prediction challenge is set to one-step prediction for a network structure with 16 convolution layers. To speed up the training process, a mini-batch learning technique with 13 samples in each batch is used. The confusion matrix for heart disease training data specifies the proposed prediction's high accuracy. Precision, recall, F1-score, and accuracy are all improved with the proposed strategy.

Accuracy, precision, F-measure, and recall are some of the performance measures that are calculated.

• Accuracy: Precision is also known as Positive Predictive Value, which is defined as the ratio of relevant to retrieved instances. As seen, the precision is expressed.

Recall: The number of relevant documents returned by a search is divided by the total number of relevant documents available.

• F1-score: F-measures are statistical variability that performs Random Error Representation.

Where TP, TN, FP, and FN indicate True Positive, True Negative, False Positive, and False Negative accordingly, accuracy is defined as the ratio of accurately predicted to total number of observations. The performance measures, such as Accuracy, precision, F-measure, Recall are calculated as follows

• Precision: Precision is also known as Positive Predictive Value which is the ratio of relevant instances to the retrieved instances. The precision is expressed as shown

Recall: Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents.

$$Recall = \frac{TP}{TP+FN}$$

• F1-score: F-measure measures are statistical variability that performs Representation of random errors.

$$F-measure = \frac{2TP}{(2TP+FP+FN)}$$

Accuracy is defined as the ratio of correctly predicted to the total number of observations

Where, TP, TN, FP and FN are represented as True Positive, True Negative, False Positive, and False Negative respectively.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

III. CONCLUSION AND FUTURE

Big Data analytics are critical in predicting heart attacks and customising cardiovascular disease treatment. Big data analysis shows some significant prospects to forecast future health status from health factors and provide the best outcomes for better decision making in heart disease prediction. However, more traffic data contaminated the algorithms, causing ambiguity and difficulty in predicting heart disease. To address this problem, big data is employed to ensure that medical services are delivered on time, and accurate diagnosis is used to examine the patient's history.

REFERENCE

[1]. L. Anying, C. Ke, S. He and L. Yu, "The Industry Data Analysis Processing Model Design: The Regional Health Disease Trend Analysis Model," 2014 International Conference on Cloud Computing and Big Data, 2014, pp. 130-133, doi: 10.1109/CCBD.2014.11.

[2]. M. S. Islam, T. Chaussalet, N. Ozkan and E. Demir, "An approach to exploring the effect of weather variations on chronic disease incidence rate and potential changes in future health systems," 2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS), 2010, pp. 190-196, doi: 10.1109/CBMS.2010.6042639.

[3]. X. Zhou, Q. Li, Z. Zhu, H. Zhao, H. Tang and Y. Feng, "Monitoring Epidemic Alert Levels by Analyzing Internet Search Volume," in IEEE Transactions on Biomedical Engineering, vol. 60, no. 2, pp. 446-452, Feb. 2013, doi: 10.1109/TBME.2012.2228264.

[4]. W. Lin, C. Lin and C. Wu, "Using time series analysis to predict affecting factors of thallium myocardial perfusion scan service usage frequency," 2018 IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 43-46, doi: 10.1109/ICASI.2018.8394282.

[5]. A. Bansal, A. Bhardwaj and A. Sharma, "Forecasting the Trend of Covid-19 Epidemic," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 406-409, doi: 10.1109/PDGC50313.2020.9315795.

[6]. P. Shankar, C. Bitter and M. Liwicki, "Digital Crop Health Monitoring by Analyzing Social Media Streams," 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), 2020, pp. 87-94, doi: 10.1109/AI4G50087.2020.9310985.

[7]. G. Lan, C. Lee, J. Wu, H. Li, S. Ho and V. S. Tseng, "A Framework for Personalized Health Trend Analysis," 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks, 2009, pp. 405-409, doi: 10.1109/I-SPAN.2009.126.

[8]. Rajawat A.S., Rawat R., Barhanpurkar K., Shaw R.N., Ghosh A. (2021) Vulnerability Analysis at Industrial Internet of Things Platform on Dark Web Network Using Computational Intelligence. In: Bansal J.C., Paprzycki M., Bianchini M., Das S. (eds) Computationally Intelligent Systems and their Applications. Studies in Computational Intelligence, vol 950. Springer, Singapore. https://doi.org/10.1007/978-981-16-0407-2_4

[9]. M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015, pp. 520-525, doi: 10.1109/ABLAZE.2015.7154917.

[10]. M. Tsuji, "Analysis of the long-run effect of e-Health intervention on chronic diseases: A DID-PSM approach," 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), 2015, pp. 511-516, doi: 10.1109/HealthCom.2015.7454556.

[11]. J. P. Claiborne, C. Wellbeloved-Stone and R. S. Valdez, "Recent trends in diabetes-related consumer health information technology research," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 4283-4288, doi: 10.1109/EMBC.2019.8857052.

[12]. J. Alsakran and L. Alnemer, "Visual Analysis of COVID-19 Trends," 2020 Seventh International Conference on Information Technology Trends (ITT), 2020, pp. 196-201, doi: 10.1109/ITT51279.2020.9320891.

[13]. Z. Yan and Y. Liu, "The multidirectional information fusion mobile health management technology," 2013 First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech), 2013, pp. 1-4, doi: 10.1109/Ubi-HealthTech.2013.6708068.

[14]. A. Bashir, A. W. Malik, A. U. Rahman, S. Iqbal, P. R. Cleary and A. Ikram, "MedCloud: Cloud-Based Disease Surveillance and Information Management System," in IEEE Access, vol. 8, pp. 81271-

81282, 2020, doi: 10.1109/ACCESS.2020.2990967.

[15]. P. Majety and V. Umamaheshwari, "An electronic system to recognize heart diseases based on heart sounds: A stochasatic alogorithm implemented on DSK6713," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016, pp. 1617-1621, doi: 10.1109/RTEICT.2016.7808106.

[16]. A. S. Billis, E. I. Papageorgiou, C. A. Frantzidis, M. S. Tsatali, A. C. Tsolaki and P. D. Bamidis, "A Decision-Support Framework for Promoting Independent Living and Ageing Well," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1, pp. 199-209, Jan. 2015, doi: 10.1109/JBHI.2014.2336757.

[17]. M. Nii, M. Momimoto, S. Kobashi, N. Kamiura, Y. Hata and K. Sorachi, "Medical Checkup and Image Data Analysis for Preventing Life Style Diseases: A Research Survey of Japan Society for the Promotion of Science with Grant-in-Aid for Scientific Research (A) (Grant number 25240038)," 2015 7th International Conference on Emerging Trends in Engineering & Technology (ICETET), 2015, pp. 117-122, doi: 10.1109/ICETET.2015.38.

[18]. R. Pinto et al., "Assessing the Impact of Public Health Campaigns Through Epidemiological, Communication and Education Indicators," 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), 2020, pp. 147-149, doi: 10.1109/ICALT49669.2020.00050.

[19]. Rajawat A.S., Rawat R., Barhanpurkar K., Shaw R.N., Ghosh A. (2021) Sleep Apnea Detection Using Contact-Based and Non-Contact-Based Using Deep Learning Methods. In: Bansal J.C., Paprzycki M., Bianchini M., Das S. (eds) Computationally Intelligent Systems and their Applications. Studies in Computational Intelligence, vol 950. Springer, Singapore. https://doi.org/10.1007/978-981-16-0407-2_7

[20]. N. Akshay and G. V. Krishna, "Design & implementation of real time bio-signal acquisition system for quality health care services for the population of rural India," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016, pp. 1315-1319, doi: 10.1109/RTEICT.2016.7808044.

[21]. J. Li, "Construction of Big Data Epidemic Forecast and Propagation Model and Analysis of Risk Visualization Trend," 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI), 2020, pp. 21-25, doi: 10.1109/ICAACI50733.2020.00009.

[22]. Rajawat A.S., Barhanpurkar K., Goyal S.B., Bedi P., Shaw R.N., Ghosh A. (2022) Efficient Deep Learning for Reforming Authentic Content Searching on Big Data. In: Bianchini M., Piuri V., Das S., Shaw R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_26

[23]. E. T. Ogidan, K. Dimililer and Y. Kirsal-Ever, "Machine Learning Applications in Disease Surveillance," 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019, pp. 1-4, doi: 10.1109/ISMSIT.2019.8932927.

[24]. D. B. Mehta and N. C. Varnagar, "Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 1158-1162, doi: 10.1109/ICOEI.2019.8862544.