International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Graphing Model of Prediction Data for Occupational Incidents in Chemical And Gas Industries

Ganapathy Subramaniam Balasubramanian, Ramaprabha Thangamani

Abstract*: Constant streaming of data for any instances at such high volumes provides insight in various organizations. Analyzing and identifying the pattern from the huge volumes of data has become difficult with its raw form of data. Visualization of information and visual data mining helps to deal with the flood of information. Constant streaming of data for any instances at such high volumes provides insight in various organizations. Analyzing and identifying the pattern from the huge volumes of data has become difficult with its raw form of data. Visualization of information and visual data mining helps to deal with the flood of information. Visual data representation takes the data and its results to all the stakeholders in a meaningful manner which comes out of the data mining process. Recent developments have brought a large number of information visualization techniques to explore the large data sets which can be converted into useful information and knowledge. Observations and inspection data gathered from chemical and gas industries are being piled up on a daily basis as raw data. Continuous analysis is a new term evolving in the industry which continuously performs on the streaming data to have real-time analysis and prediction on-live. In this paper, usage of the various graphing model as per the respective information obtained from the organization have been discussed and justified. It also describes the value addition in making the decisions by representations through graphs and charts for better understanding. Heatmap, Scattergram and customized Radar plots the analyzed data as in the required format to visualize the prediction done for the occupational incidents in chemical and gas industries. As a result of the graphing model, representation provides a higher level of confidence in the findings of the analysis. This fact takes a better visual representation technique and transforms them to provide better results with faster processing and understanding.

Keywords*: Occupational incidents, Graphing Model, Charts and Graphs, Continuous Analysis, Real-time processing.

I. INTRODUCTION

Presentation of the analyzed data is an important step to visualize the information and findings for a better understanding.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
Ganapathy Subramaniam Balasubramanian*, Research Scholar, PG and Research Department of Computer Science and Applications, Vivekanandha College of Arts and Science, TN, India. Email: bganapathy@ceiamerica.com
Dr. T. Ramaprabha, PG and Research Department of Computer Sciences and Applications, Vivekanandha College of Arts and Science, Tiruchengode, TN, India. Email: ramaradha1971@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Meaningful visualization delivers the needed result to the organization to make the decision in a clear way.

Tabulation and graphical methodologies convey the result and its message in a better way. Visual representation of the data cannot be done without understanding the underlying properties of the data types in the domain. Data often deals with values and calculations which would be measurements, observations, calculated numbers, percentages or descriptions of the information. Basically, data are two types: constant and variable. Constant is a value that does not change, while a characteristic, number, or quantity that increases or decreases over time or takes different values in different situations is called variable. Due to unchangeable property, the constant is not used and only variable is used for summary measures and analysis [2].

Visual representation like charts and graphs provides insight into the data and helps to summarize through depiction and figure out the pattern out of it. The purpose of the data analysis is to perform calculations on the raw data using an appropriate algorithm to achieve the results to conclude the solution. A spreadsheet application like Excel paves a good starting point to perform such calculations and to display the analyzed results. The design principles for generating the graphical methodology will depend on the purpose to be used for. Graphical representation keeps the data extremely simple and avoids raw tabular data and/or text. Tabular data presentation should start with simple graphs to see for any outliers or spikes. Definition the verification points helps the analysts to determine the pattern or value of it [21].

II. PROCESS MODEL

The graphing model follows the three-step process: Initiation, Filtration, and Drill-down Data. In the Initiation process, getting an overview of the data is an important functionality. This helps to identify the patterns and focuses on the travel path of the data being plotted. This may end up by finding one or more patterns around the data. To understand the data behind the pattern, analysis has to be performed by drilling down the details. The graphing model may use all three steps for the visual representation of the data. Graphing model processes are useful for analyzing the high-level view of the data and to identify the interesting subsets inside the same. Focusing on the subset of the data through the graphing model provides the interlinks between the items and also to pick the right path within it.

© Copyright: All rights reserved.

Drill down of data to get the details of the data is important to explore further interesting subsets.

The graphing model not only provides the visualization techniques but also defines the relationship between the gaps in the entities.

In information visualization, the data usually consists of a large number of records each consisting of a number of variables or dimensions. Each record corresponds to an observation, measurement, transaction, etc., [12] There are a number of well-known techniques for visualizing such data sets such as x-y plots, line plots, and histograms. These techniques are useful for data exploration but are limited to relatively small and low-dimensional data sets. In the last decade, a large number of novel information visualization techniques have been developed, allowing visualizations of multidimensional data sets without inherent two- or three-dimensional semantics [12, 21].

List of the data type to be visualized and all modern-day visualization technique are given in the below table:

Table. 1. List of data types and visualization techniques

Data type to be visualized	Visualization technique
One dimensional data	Standard 2D/3D display
Two-dimensional data	Geometrically transformed display
Multi-dimensional data	Iconic display
Text/web	Dense pixel display
Hierarchies/Graphs	Stacked display
Algorithms/Software	Interaction and distortion

One of the challenges in reporting and data analysis is to understand the types of graphs to use and why. Understanding the data to be visualized and visualization techniques should put the organization on the right path. By choosing the wrong visual aid or simply leaving it to the default type of data visualization may cause confusion and produces unsatisfactory results through mistaken data interpretation. To decide upon the charts which clarify the result and provide the right path of analysis, the following process methodology helps to identify the better visualization charts for the data obtained.

A. Comparison

A comparison of data should be done among the observed data points or over the timeline of the data points. When it does among the data points, data can be further classified into one variable per data point or two variables per data point. One variable data point can be classified into a few categories or many categories. When comparison needs to be done over the timeline, data should be categorized into cyclical data, non-cyclical data, single or few periods and many periods.

B. Distribution

Distribution of the data crawls over single, two and multiple variables. Over two or more variables deal with three or multiple dimensional charts. Histogram, Scatter and Area define the categories of the charts to be used for visualization in the graphing model.

C. Composition

The composition of the data points relies on the static calculation or the data dynamically changing over time. Static composition works with sharing of the totals, accumulation or subtraction to the totals and component of the components. On the other hand, a dynamic timeline deals with few and many periods of the time which further drills down into relative differences and relative differences combined with absolute differences as per the timeline data points.

D. Relationship

Relationship representation deals with two or more variables in the form of rectangular coordinate data points on a system. The position of the data point can be determined by the value of the given variable. By observing the distribution of the data points, the correlation between the variables can be inferred. Relationship charts are well suitable to show how one variable relates to one or numerous different variables. This is used to represent the positive effect, has no effect, and/or negative effects with respect to another variable.

III. ARCHITECTURE

The architecture of the graphing model used for the visual representation of processed data, as shown in Fig 1., detailed out the components involved for the better outcome. The integration between these components is aligned in such a way to establish a better representation and visualization of the data points. Starting from inspection data through getting the outcome of the predicted results, the graphing model helps to represent the precise results in an anticipated way.

Fig 1. High-level Graphing architecture diagram

Data Streaming is a method of posting a continuous stream of data that can be processed through the algorithms to obtain structural data. Inspection data, sensor data from the gas detection instrument and historical incidents should be streamed through Kafka. The stream is divided into RDDs (Resilient Distributed Datasets) which is a fundamental data structure of Spark. Processing of the data takes place through Spark Jobs. The approach of incremental algorithms can be used to manipulate the history data and real-time data [15].

Heatmap representation of the data can be generated from an algorithm to visualize the results [21]. The data dashboard displays the required heatmap and also keeps the data live through the push notifications. Heat maps are one of the good ways to plot the grouped data. The plotting area of the heat map is divided into equal unique shapes like circle, square or rectangle. The rows and columns represent the data sets in the given data table. Each shape in the heat map is color-coded to denote the respective data point value in the data table.

Replicated values are shown in intense based on the mean, median, standard deviation or geometric mean of the data points on the base the heat map. Color scales can encode continuous data with discrete categories.

Retrieval Number: D2208029420/2020©BEIESP
DOI: 10.35940/ijitee.D2208.029420
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

A scatter plot is a type of chart that consists of multiple data points representing different factors plotted across two axes from the data table [2].

Each data point depicted from the data table in a scatter plot might have multiple observations. Unique colors should be used to identify the variables to see the pattern in the plotted graph. A scatter plot chart allows performing analysis of whether the pattern exists between the given variables by plotting the data points. This also helps to look for outliers or for understanding the distribution of the data among the given variables. To understand the relationship between the two variables, trend lines based on the logarithmic regression, logistic regression, linear regression or binomial regression can also be used based on the factors defined for the desired output.

IV. IMPLEMENTATION

The inspection data is the core data to be processed for deriving the results. Inspection data is directly processed by a calculation engine using the Scoring algorithm [17] to clean up and fill in the data with the cause of the incident for the unknown ones and to define the scoring point based on the daily inspection results. The result of this algorithm along the way updates the view, offering a real-time prediction of the actual outcome occupational incident.

The processed inspection data with its calculated score value have been obtained from organization data is shown below in Figure 2 for the reference. Though the inspection data are collected on the daily or frequent shift basis, bi-weekly sample data is shown here for better representation. Inspection types and timelines are given as the row and column headings respectively.

	1/6/2010	1/20/2010	2/3/2010	2/17/2010
Employee Observation / PPE	82.1	83.1	85.1	77.9
Housing Keeping	31.9	17.2	75.3	30.5
Railings / Covers / Cables	75.0	29.0	31.3	20.3
Fire protection	82.8	88.3	78.5	89.0
First aid / Sanitation	28.2	77.7	66.0	53.7
Tools	32.3	13.3	32.1	31.5
Scaffoldings	26.3	75.0	30.5	28.4
Ladders and Stairs	33.5	26.3	29.7	72.7
Electrical	74.0	34.8	80.8	73.5
Lifts	30.0	16.0	28.2	14.7
Programs / information	34.5	71.8	77.5	33.9
Site / Public & Environmental protection	35.0	67.8	69.5	68.8

Fig 2. Heat map of processed inspection data with its score

Heat map has been defined in such a way that the lowest value to the highest value has been mapped to the red to green color respectively. Datapoint represented are defined between 0% to 100%. On a periodic basis, a batch will be run to calibrate the data to correct the deviations.

SafeOne prediction model [4] has been trained to predict the potential value of the occupational incidents in the industry. Once the model is trained, the test set of data has been applied for verification and validation of the model to write the prediction value. The algorithm helps not only to regression models but also used to build an analytical or a classification system.

Following dashboard displays the radar graph along with Unsafe percentage:

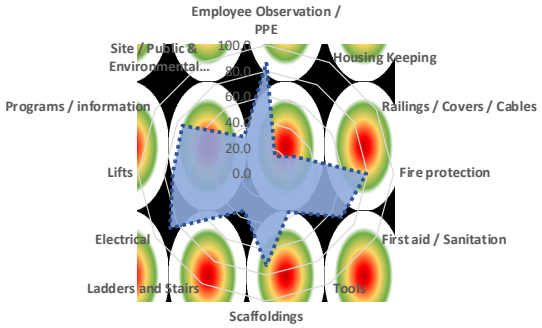

Fig 3. Radar graph to visualize the Prediction model

SafeOne model predicts the "Unsafe" percentage with reference to all the factors considered during the machine learning process. The radar style graph is used to plot the score of each inspection type. Value 0 being the most unsafe zone and value 100 being the safest zone on the radar. Another example prediction which shows lower Unsafe percentage is shown below:

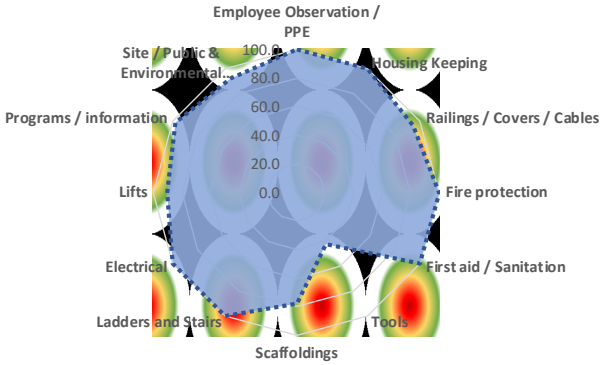

Fig 4. Radar graph with lower Unsafe percentage

ChartJS renders the graph in real-time when the processed data appears from the cloud database. Excerpt of the ChartJS code has been given below:

```
var mySafeOneChart = new Chart(ctx, {
      type: 'radar',
      data: safeOnePredictionData,
      options: chartOptions
});
```

ChartJS options have been used to specify the customizations of the chart for rendering the background highlighter to define the safe to unsafe areas by color-coding. The inspection data scattergram is shown below in Figure 5.

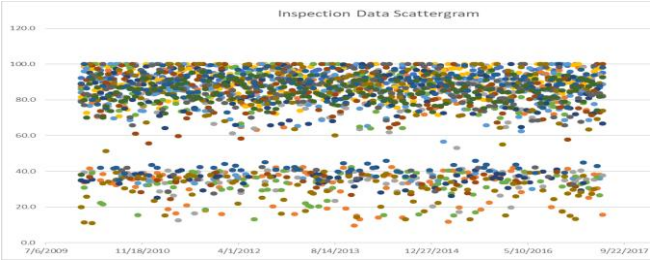

Figure 5. Inspection data scattergram

The position of the inspection data with reference to the percentage value is determined by the value of the variable and observation of every inspection category has been distributed towards the data points which shows the correlation between the inspection category and inspection score obtained through the Scoring algorithm.

© Copyright: All rights reserved.

V. RESULT AND DISCUSSION

Defining the graphing model improves the analysis by means of understanding the results by depicting the calculated prediction data. Angular based dashboard page created using HTML and Angular (JavaScript) framework libraries showcases the data to visualize the results. The outcome of the Prediction model is validated against the known results to verify against the obtained result in the past. The graphing model continuously depicting the real-time calculations through heatmap and radar for every set of data point carries a great value for the organization to look through the depictions and to make the decisions based on the results. Having a continuous graphing model and continuous monitoring improves the quality of the desired output and serves the purpose of the analysis. Having the resulting dashboard to be responsive, it can be viewed on different devices and resolutions.

An error ratio has also been calculated from the results as a part of the prediction model calculation. The idea of visualizing the results in real-time by setting the base platform to smoothly walk through the workflow from raw data to prediction results. The graphing model depends upon the importance of clearly communicating the detailed trend for each result to the organization to make better results. In most cases, the high-level depiction of the data itself serves good enough to observe and understand the current standpoint and also plays a role in decision making.

Overall results achieved through the proposed graphical model fit the requirement to analyze the data of the occupational incident and depicts the predicted data in real-time for better viewability.

VI. CONCLUSION

Graphing methods vary according to the scales of measurements and presentation. When the desired result data in this case, inspection data, is a continuous variable, which is used for the parametric methods. Some of the data which are a discrete variable/qualitative variable, the only non-parametric method should be included to define the calculation for the analysis. Similarly, graphical models are varying according to the scales of measurements. A score set of the inspection data has been obtained through the Scoring algorithm which serves as an input for the prediction model. Result visualization has been depicted using the heatmap, scattergram and the radar graph along with the prediction data defined to determine the safe and unsafe percentage. The overall design of this evaluation of the graphical model explores and implements the fully functional working model of the proposed graphical model architecture in a place that allows visualization of the solution in real-time for better results.

Proper knowledge of data types is necessary to analyze data sets with appropriate statistical methods. This not only enhances the ability of the chemical and gas industries to decide its summary measures but also helps the organization to analyze data sets with proper statistical methods. Statistical graphical model depictions are available to display summaries and findings of data sets. Appropriate identification of the charts to depicts the data sets is important and is very much useful to communicate the summary and findings to the stakeholders of the organization with precise results.

REFERENCES

1. Aloysius, G., & Binu, D. (2013). An approach to product placement in supermarkets using PrefixSpan algorithm. Journal of King Saud University - Computer and Information Sciences, 25(1), 77–87. https://doi.org/10.1016/j.jksuci.2012.07.001
2. Angra, A., & Gardner, S. M. (2018). The graph rubric: Development of a teaching, learning, and research tool. CBE Life Sciences Education, 17(4), 1–18. https://doi.org/10.1187/cbe.18-01-0007
3. AWS. (2016). Building Media & Entertainment Predictive Analytics Solutions on AWS. (December), 31. Retrieved from https://d0.awsstatic.com/whitepapers/Analytics/ME Advanced Analytics on AWS.pdf
4. Balasubramanian, G. S., & Ramaprabha, T. (2019). Prediction model for occupational incidents in chemical and gas industries. International Journal of Recent Technology and Engineering, 8(4), 3836–3840. https://doi.org/10.35940/ijrte.D8212.118419
5. Baranyi, J., & Buss da Silva, N. (2017). The use of predictive models to optimize risk of decisions. International Journal of Food Microbiology, 240, 19–23. https://doi.org/10.1016/j.ijfoodmicro.2016.10.016
6. Bureau of Indian Standards. (2007). Occupational Health and Safety Management Systems. 2000, 1–28. https://doi.org/10.1002/0471435139.hyg049.pub2
7. Davoudi Kakhki, F., Freeman, S. A., & Mosher, G. A. (2019). Evaluating machine learning performance in predicting injury severity in agribusiness industries. Safety Science, 117, 257–262. https://doi.org/10.1016/j.ssci.2019.04.026
8. Development, S. (n.d.). Factors of Predictive Model for Occupational Safety.
9. Helvert, M. van. (2014). Data visualizations in popular Dutch media. Retrieved from Masters of Media website: http://mastersofmedia.hum.uva.nl/2014/04/17/data-visualizations-in-popular-dutch-media/
10. JANICAK, C. A. (1996). Predicting accidents at work. 115–121.
11. Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science, 57, 500–508. https://doi.org/10.1016/j.procs.2015.07.372
12. Keim, D. A. (2002). Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, 8(1), 1–8. https://doi.org/10.1109/2945.981847
13. Khan, F. I., Abbasi, S. ., Mpp, C. De, & European Union. (2006). Techniques and methodologies for risk analysis in chemical process industries. In Journal of Loss Prevention in the Process Industries (Vol. 11). https://doi.org/10.2790/73321
14. Kharagpur, T. (2017). Computational Intelligence, Communications, and Business Analytics. 776(January 2018). https://doi.org/10.1007/978-981-10-6430-2
15. Kroß, J., Brunnert, A., Prehofer, C., Runkler, T. A., & Krcmar, H. (2015). Stream Processing on Demand for Lambda Architectures. In M. Beltrán, W. Knottenbelt, & J. Bradley (Eds.), Computer Performance Engineering (pp. 243–257). Cham: Springer International Publishing.
16. Mooney, C. H., & Roddick, J. F. (2013). Sequential pattern mining -- approaches and algorithms. ACM Computing Surveys, 45(JUNE), 1–39. https://doi.org/10.1145/2431211.2431218
17. Mr. Ganapathy Subramaniam B, D. T. R., & B, M. G. S. (2019). New Scoring Algorithm for predicting the cause of incident occurrences in chemical industries. International Journal of Emerging Technologies and Innovative Research JETIR, (Vol. 6-Issue 5 (May-2019)). Retrieved from http://www.jetir.org/view?paper=JETIR1905D50.
18. Qadah, E., Mock, M., Alevizos, E., & Fuchs, G. (2018). A distributed online learning approach for pattern prediction over movement event streams with apache flink. CEUR Workshop Proceedings, 2083, 109–116.
19. Reurings, M., & Janssen, T. (2006). Accident prediction models for urban and rural carriageways. 81.
20. Sarkar, S., Pateshwari, V., & Maiti, J. (2017). Predictive model for incident occurrences in steel plant in India. 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), (July), 1–5.

© Copyright: All rights reserved.

https://doi.org/10.1109/ICCCNT.2017.8204077
21. Slutsky, D. (2014). The Effective Use of Graphs. Journal of Wrist Surgery, 03(02), 067–068. https://doi.org/10.1055/s-0034-1375704
22. T. Akomolafe, D., & Olutayo, A. (2013). Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. American Journal of Database Theory and Application, 1(3), 26–38. https://doi.org/10.5923/j.database.20120103.01
23. Varsha, A. V, Goerge, G., & Sahajanandan, R. (2019). Lutembacher Syndrome: Dilemma of Doing a Tricuspid Annuloplasty Abstract. Annals of Cardiac Anaesthesia, 22(1), 1–5. https://doi.org/10.4103/aca.ACA
24. Verma, S., & Chaudhari, S. (2017). Safety of Workers in Indian Mines: Study, Analysis, and Prediction. Safety and Health at Work, 8(3), 267–275. https://doi.org/10.1016/j.shaw.2017.01.001

AUTHORS PROFILE

Ganapathy Subramaniam B., from Chennai and born in 1979, research scholar, from Vivekanandha College of Arts and Science affiliated to Periyar University, in the field of Computer Science since Feb 2017. Data mining is his major field of research. He has obtained a Masters of Philosophy in Computer Science from Alagappa University in the year 2006 and completed his Master of Science in Information Technology in 2002. He has earned a bachelor's degree in Commerce from Annamalai University in 2000 and has also obtained a Diploma in Computer Technology from All India Council of Technical Education in 1997. Possess good knowledge in cutting edge technology stack and his field of study is data mining in gas detection domain.

Dr. T. Ramaprabha, M.Sc., M.Phil., Ph.D., works as a Professor of PG and Research Department of Computer Science and Applications in Vivekanandha College of Arts & Sciences for Women (Autonomous), Tiruchengode, Tamil Nadu. She has completed her Ph.D., in the year 2013 at Mother Theresa Women's University, Kodaikanal. She has more than 22 years of teaching and research experience and she has published many papers in conferences and journals. Her research of interest is stereo pair image compression in image processing.

© Copyright: All rights reserved.