Dynamic Representation and Visualization in Networking for Big Data Using Ai

Sayyed Aziz Ahmed[1], Dr. Manoj Eknath Patil[2]

Research Scholar[1], Research Guide[2]

[1,2]Department of Computer Science & Engineering, Dr.A.P.J.Abdul Kalam University, Indore(M.P)

minaltkolhe@gmail.com[1] , mepatil@gmail.com[2]

Article Info
Page Number: 9814 - 9822
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: *15 September 2022*
Revised: *25 October 2022*
Accepted: *14 November 2022*
Publication: *21 December 2022*

Abstract

It's hard to figure out what's going on with a lot of network data because both the amount of data and the way it changes over time and space are getting bigger. When nodes are added to or taken out of a network, both the topology of the network as a whole and the properties of its links are dynamic and can change. Using the usual methods for data mining, it is either impossible or too expensive to process large, dynamic networks. Neither of these is the best choice. To do this, we came up with a technology called DNAV, which stands for "dynamic network analysis and visualisation." The AI based algorithm is mostly based on a dynamic graph, which is made up of links that are separated by time. Each edge segment in the network shows how the graph elements in the network, like communication hubs, have changed over time. The proposed static view of dynamic networks can make the process of visually analysing these networks much easier because. To study and analysis malware data classification and Thevisualisation has data filtering modules like time selection, hops settings, entity selection, and edge weight thresholds to make it easier to see large networks. to suggest using AI to make dynamic representations and visualizations of big data that is networked.

Keywords: Artificial Intelligence ,Visualization Networking , big data

Introduction

As we move into the age of "big data," the data are becoming more complex and dynamic. On mobile social networks, the number of connected devices and users is expected to keep growing at an exponential rate. This will only make the relationships between them stronger. Anyone can connect to a network or cut off their connection at any time. Over time, it's possible that the qualities of a network's links and its other parts will change. Dynamic networks include wireless and mobile communication networks, social networks, the Internet of Things (IoT), sensor networks, and other types of networks. The topology of dynamic networks [1] changes over time because the links between nodes are only kept up sometimes. The spatial-temporal aspects of large networks and the many event links between nodes make it harder to figure out how dynamic networks work. Even though data mining techniques [2] have been made for analysing dynamic networks, it is often impossible to take into account all of the possible outcomes when working with large networks because of how complicated they are. Data mining and machine learning can take a long time and have a lot of mistakes, so there are times when it might be better to look at dynamic networks visually. This is because it makes it easier for people to look for patterns than those algorithms. Small multiples and animation are both common ways to show how networks change over time. But because people's minds are limited, it can be hard to keep track of changes over time when using these methods. Even though there are a lot of different visual approaches, like multi-dimensional visualisations [4–6], it is often hard to read the visualisation and keep the mental map up to date [7]. In the age of "big data," one of the most difficult things to do is to keep researching effective and scalable ways to look at huge, changing networks. Despite these problems, we made a simple visual analytics tool that academics and people who run dynamic networks can use to look for patterns in the networks they run. It's important to know that some changes in dynamic networks are thought to be predictable while others are not (i.e., anomalies). One of our main goals is to help human investigators see patterns and notice out-of-the-ordinary changes in systems that are complicated and always changing. The dynamic graph view, which is an important part of the Dynamic Network Analysis and Visualization (DNAV) tool, is the focus of this article. We use the graph because most people are already familiar with how it's put together (a node-link diagram). In dynamic graphs, which are different from static graphs, time and location information is put into the links between nodes using a new method. This is how static graphs work. If we let users choose the properties of edges, the links in our dynamic graphs can be broken up into time-bound pieces. When more than one status value can be used to describe a certain amount of time, the segments can be broken down even further. Dynamic edge attributes change over time, and these changes are shown by a range of colours. By breaking links into segments, we can show a dynamic network in a single, static view. This means that users don't have to remember how to move from one frame to the next, as they do with animation schemes. Our dynamic graphs now have a number of interactive features, like a time selection bar that lets researchers limit their analysis to a certain time frame. This change was made to improve how well the graphs can be scaled. Analysts can also set the maximum number of network hops that can be taken from any particular node in the system. When big networks are broken down into the smaller graphs that make up the whole network, it's much easier to see the network as a whole. Scalability problems are also solved by putting in place edge filtering that is based on edge weights (e.g., number of connections). Even though the public has access to about four million records of communication, we put the DNAV tool through a lot of tests. With the help of the suggested dynamic graph display, strange changes in the way the network communicates can be found in time and place.

Related work

With the help of more powerful high-throughput technologies and new biological data, it is hoped that scientists will be able to learn more about the complicated systems that make cells work. Gene expression measurements can be used to figure out links between genes in a cell, which could lead to the discovery of important interactions that upset the cell's normal state.

Liu, A et al[1]The solid foundations of data visualisation technology that were built during the time of "big data analytics" can be used as a base for fitness solutions based on AI that can build on this base (AI). Because of this, the people who wrote this study made a VEVF, which is a good way to visualise movies that uses both artificial intelligence and big data analytics. The authors of this work use machine learning to sort sports videos into different groups. This method looks at the video footage by extracting both temporal and spatial features. We start with a network of convolutional neural networks. On top of that, we add a layer of temporal pooling. This is where we start.

Zhuang, Yt. Et al[2]Our study focuses on how artificial intelligence (AI) can be used to big data, thus we look at the most recent advances in AI theory and practise. In order to create AI that is explainable, powerful, and generalizable, we draw the following findings regarding how data-driven machine learning can be integrated with human knowledge (common priors or implicit intuitions): from elementary arithmetic to sophisticated neural reasoning; from data-driven to logic-rule-based data-driven models

Torre-Bastida, et al.[3]In addition to analysing the published literature, bio-inspired algorithms are used to find developing trends and open problems in big data that haven't been solved yet but can be solved by these algorithms. These problems haven't been fixed yet, and they haven't been fixed yet either. The second thing that has been added to this work is an explanation of how bio-inspired algorithms need to be changed before they can be used with Big Data. In this situation, a lot of different data sources need to be fused together before they can be processed and mined, and bio-inspired algorithms need to be changed so they can work. This separation makes it possible to compare and contrast how well existing methods work in a wide range of problem spaces and research areas. The goal is to find new ways to study and develop. In the end, the survey makes suggestions for how future research should go and lists problems that still exist in this field of study, even though people have tried to solve them before.

Himeur, Y., et al.[4] If God wills it, the second part of this article will explain how AI and big data analytics are used in the real world. There are three examples given to show how AI and big data analytics can be used in BAMSs. These examples have to do with finding energy problems in homes and businesses and making sure that energy and performance are at their best in sports arenas. In the end, this article looks at some possible directions and makes some helpful suggestions for making BAMSs in smart buildings work better and be more reliable.

Johnson, M., et al[5]The goal of this study is to find trends, gaps, and opportunities in the BD&AI workforce by using bibliometric analysis and data from job postings to draw conclusions. The research team started their study by using papers about BD&AI that had been published in high-impact journals to do bibliometric analysis and build word co-occurrence diagrams. Their goal was to find out how changes in technology were affecting different business sectors.

Proposed Methodology

Scalability and changeability are two areas where visual analytics can be hard. Table 2 shows the current state of research for both static data and dynamic data. The data size is used to organise the information. Most of the time, the solutions that were made for type A or type B problems don't work well for large datasets that are always changing [9].

The "four Vs" of big data present problems, but approaches that focus on visualisation can turn these problems into opportunities [2].The Amount of Data: The methods were made so that very large data sets could be handled and conclusions could be drawn from them.Also, the procedures are flexible enough that they can be used to add any number of different types of data that may be needed.Because of these technologies, businesses can now switch from batch processing to real-time stream processing, which speeds up the processing.The methods not only help users make infographics and heatmaps that look good, but they also help them make money by getting useful information from huge amounts of data. Value: The methods help people make money by finding patterns in huge amounts of data.Big data is notoriously hard to visualise because the data comes in many different forms (structured, semi-structured, and unstructured). A huge amount of information needs to be looked at quickly. When working with a lot of data, it can be hard and take a lot of time to make a completely new tool for data visualisation that has efficient indexing. By combining cloud computing with modern graphical user interfaces, it is possible to make it easier to manage big data on a large scale [3].

Graphs, tables, texts, and trees, among other types of metadata, can cause problems for visualisation systems. Large data sets often have formats that are not organised. It is important to move visualisation closer to the data so that relevant information can be pulled from large amounts of data more quickly. Visualization software is best used on-site. Because the data set is so big, it might be hard to visualise, which means that many processes will need to run in parallel. With parallel visualisation approaches, the challenge is to break up a task into smaller parts that can be worked on at the same time [10, 11].In this age of big data, one of the most important parts of the discovery process is the use of good data visualisation. One way to deal with the problems caused by big data's high complexity and dimensionality is to use different dimensionality reduction strategies. There's no guarantee that they'll always be helpful, though. When there are more dimensions used in a visualisation, it is more likely that interesting patterns, correlations, or outliers will be found [11].

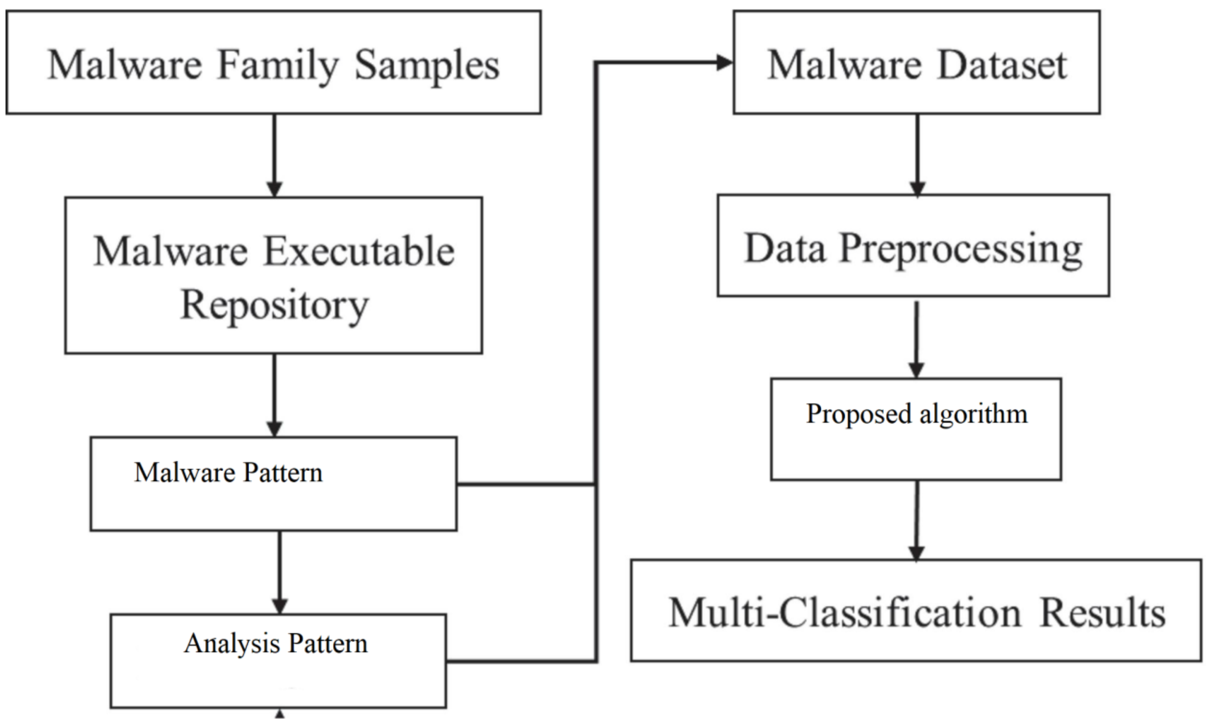

Figure 1Flow chart of proposed approach

Challenges of Big Data Visualization

Scalability and changeability are two areas where visual analytics can be hard. Table 2 shows the current state of research for both static data and dynamic data. The data size is used to organise the information. Most of the time, the solutions that were made for type A or type B problems don't work well for large datasets that are always changing [9].

The "four Vs" of big data present problems, but approaches that focus on visualisation can turn these problems into opportunities [2].

The Amount of Data: The methods were made so that very large data sets could be handled and conclusions could be drawn from them.

Also, the procedures are flexible enough that they can be used to add any number of different types of data that may be needed.

Because of these technologies, businesses can now switch from batch processing to real-time stream processing, which speeds up the processing.

The methods not only help users make infographics and heatmaps that look good, but they also help them make money by getting useful information from huge amounts of data. Value: The methods help people make money by finding patterns in huge amounts of data.

Big data is notoriously hard to visualise because the data comes in many different forms (structured, semi-structured, and unstructured). A huge amount of information needs to be looked at quickly. When working with a lot of data, it can be hard and take a lot of time to make a completely new tool for data visualisation that has efficient indexing. By combining cloud computing with modern graphical user interfaces, big data management can be made more scalable [3].

Graphs, tables, texts, and trees, among other types of metadata, can make it hard for visualisation systems to do their jobs. Most of the time, there is no structure to how a lot of data is stored. It is important to move visualisation closer to the data so that relevant information can be pulled from large amounts of data more quickly. Visualization software is best used on-site. Because the data set is so big, it might be hard to visualise, which means that many processes will need to run in parallel. With parallel visualisation approaches, the challenge is to break up a task into smaller parts that can be worked on at the same time [10, 11].

In this age of big data, one of the most important parts of the discovery process is the use of good data visualisation. One way to deal with the problems caused by big data's high complexity and dimensionality is to use different dimensionality reduction strategies. There's no guarantee that they'll always be helpful, though. When there are more dimensions used in a visualisation, it is more likely that interesting patterns, correlations, or outliers will be found [11].

When trying to visualise a lot of data, you might run into issues like the ones below [12]:

Most of the objects in the dataset are too close to each other, which makes it hard to see anything. Users will no longer see them as two separate things.

Even though it can be helpful, reducing how much data is immediately visible can cause some data to be lost.

Data visualisation techniques can only do so much because of the aspect ratio and resolution of the device and because of how people see the world.

Because viewers can't keep up with the constantly changing visuals, they can only passively take in the information that is being shown.

Because static visualisation makes it hard to see, the high performance requirement for this method is not very high.

When visualising a lot of data, it can be hard to make the data easy to understand and easy to interact with. When data is reduced by sampling or filtering, it can hide interesting patterns or outliers, but when all of the data is shown at once, it can be too much for people to understand. When there is a lot of latency, which is caused by querying large data stores, it can be hard to talk in a natural way [13].Data visualisation in Big Data applications is hard to do because Big Data is so big and has so many dimensions. Most of the Big Data visualisation tools available today are not scalable, don't work well, or take too long to respond. When doing visual analytics, uncertainty can make it hard to make effective visualisations that take uncertainty into account [5].In [14], the difficulties and possible solutions of visualising large amounts of data were talked about:

One way to fix the speed problem is to use hardware that is only used for that purpose. There is the chance to use parallel processing that works well and more storage space. Another option is to put the data in random access memory. Several computers have been set up in a grid so that this goal can be reached.Second, getting control of the data: If you want to find a solution to this problem, you need to have the right kind of domain expertise on hand.Pay attention to the quality of the data by using data governance or information management to make sure the records are correct.When presenting useful findings, it may be helpful to group data into a bird's-eye view from which smaller subsets of data can be seen and the data itself can be presented in an effective way.Handling outliers One way to deal with data irregularities called "outliers" is to either leave them out of the collected data or make a new chart that only shows the outliers.

Proposed AI based model

A voting system is a common way to combine the results of several simple classifiers. After each base classifier has had a chance to make a guess about each dataset instance, the voting method is used to decide which class each data point belongs to. Based on the vote of the base model Cj, the equation makes a guess about the class label of each instance x. This is like what happens when people vote by majority or plurality.

$$\hat{y} = mode\{C_1(x), C_2(x) \ldots . C_m(x)\}$$

Consider the following hypothetical situation: there are three classifiers, C1, C2, and C3, and each of them puts an instance of the dataset into one of two binary classes (either 1 or 0). If at least two different classifiers agree that a particular instance belongs in class 1, that class will be called 1.

When we vote with a weighted majority, on the other hand, we can give each classifier Cj a certain level of importance. wj is a symbol that stands for this meaning. By naming an instance in this way,

$$\hat{y} = \arg\max \sum_{j=1}^{m} w_j f_A(C_j(x) = i)$$

we give ourselves the ability to use equations.

where L is the list of names for categories that are unique and A is the characteristic function (Cj(x) = I L). In the example above, if each classifier had a weight of 0.6, 0.2, and 0.2, respectively, the final prediction would be 1.

Results analysis

This section talks about the experimental environment because it is important to large amounts of data. The proposed architecture uses open-source parts, such as Apache Spark for effective data preparation, HDFS for distributed storage, and Python for data-intensive analysis.. Every test is run on a multinode cluster with a master node, two worker nodes, and the Ubuntu 18.04.2 operating system. This cluster is powered by Apache Spark and has a master node and two worker nodes. The main server is run by an Intel(R) Core(TM) i7-5500U processor that runs at 2.4 GHz. It also has 1 TB of RAM and 16 GB of DDR3 1600 MHz RAM. Each worker node has an Intel(R) Core(TM) i5-8250U processor that runs at 1.6 GHz, 8 GB of DDR4 memory, and a 1 TB hard drive. shows a picture of the equipment that was used in the experiment. In this section, we'll talk about what happened when the suggested method was used in Apache Spark. The best base classifier and the usual ensemble techniques are used, and then the results of using the recommended ensemble methods on a unified set of features are compared to those of using the usual ensemble techniques and the best base classifier. During our evaluation, we made sure to keep track of the model's accuracy, recall rates, and F1 scores. In the sections that follow, we'll talk about each of these things in more depth: The suggested malware detection ensemble approaches are tested on a set of 150000 files, of which 20000 are dangerous and 92,122 are safe. The next step is to make a combined feature vector that has 1000 dimensions. This will require taking into account both static and dynamic attributes, such as file metadata, file size, packer detection, and information about sections. It's the job of this feature vector to give a raw value, which it does very well.

First, let's look at the results of an experiment that was done in an environment with a lot of data. The experiment was about how to classify datasets. To start, 10-fold cross validation is used to build five basic models, and then an equation is used to figure out how well each model works. After that, the TPR, FPR, FNR, precision, accuracy (percent), F-measure, and MCC are used to evaluate the results of the base classifiers' classification work.

Table 1 Classification results using base classifiers

Machine Learning Model	TPR	FPR	FNR	Precision	F-Measure	Accuracy (%)	MCC
LR	0.733	0.144	0.144	0.733	0.732	83.1	0.574
KNN	0.810	0.065	0.080	0.822	0.816	91.5	0.743

Random Forest	0.833	0.039	0.053	0.844	0.838	93.7	0.788
SVM	0.798	0.104	0.112	0.783	0.784	88.2	0.677
Proposed model	0.967	0.014	0.020	0.974	0.970	97.2	0.850

Conclusion

The images could be still or they could move around. Interactable visualisations of data are more useful than static data tools because they help people find new insights. Using interactive visualisations, it is possible to get useful information from large amounts of data. Using interactive brushes and making links between different visualisation techniques and network or web-based technologies could make scientific work easier. Web-based visualisation makes it easy to get to dynamic data quickly and keep graphics up to date. Some of the older ways of visualising data have not been made nearly as useful as they could be to deal with huge amounts of data. There is a big gap in the number of cutting-edge methods and tools that can be used to visualise big data across a wide range of big data applications. In this study, we focus on the most recent developments in Big Data visualisation and do a spark of the visualisation software packages that can now be used for Big Data visualisation. This will lead to the creation of new tools and methods for making huge amounts of data easier to understand. The close integration of Big Data analytics and visualisation could help applications that use Big Data. Immersive virtual reality is a modern and effective way to deal with the problem of how to manage high dimensionality and abstraction in a useful way. This will give a big boost to making sense of big data through visualisation.

Reference

[1]. Liu, A., Mahapatra, R.P. & Mayuri, A.V.R. Hybrid design for sports data visualization using AI and big data analytics. *Complex Intell. Syst.* (2021). https://doi.org/10.1007/s40747-021-00557-w

[2]. Zhuang, Yt., Wu, F., Chen, C. et al. Challenges and opportunities: from big data to knowledge in AI 2.0. Frontiers Inf Technol Electronic Eng 18, 3–14 (2017). https://doi.org/10.1631/FITEE.1601883

[3]. Torre-Bastida, A.I., Díaz-de-Arcaya, J., Osaba, E. *et al.* Bio-inspired computation for big data fusion, storage, processing, learning and visualization: state of the art and future directions. *Neural Comput&Applic* (2021). https://doi.org/10.1007/s00521-021-06332-9.

[4]. Himeur, Y., Elnour, M., Fadli, F. *et al.* AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. *ArtifIntell Rev* (2022). https://doi.org/10.1007/s10462-022-10286-2

[5]. Johnson, M., Jain, R., Brennan-Tonetta, P. et al. Impact of Big Data and Artificial Intelligence on Industry: Developing a Workforce Roadmap for a Data Driven Economy. Glob J Flex Syst Manag 22, 197–217 (2021). https://doi.org/10.1007/s40171-021-00272-y

[6]. Rao, T.R., Mitra, P., Bhatt, R. *et al.*The big data system, components, tools, and technologies: a survey. *Knowl Inf Syst***60**, 1165–1245 (2019). https://doi.org/10.1007/s10115-018-1248-0.

[7]. Naghib, A., Jafari Navimipour, N., Hosseinzadeh, M. *et al.*A comprehensive and systematic literature review on the big data management techniques in the internet of things. *Wireless Netw* (2022). https://doi.org/10.1007/s11276-022-03177-5
[8]. Shih, WC., Yang, CT., Jiang, CT. *et al.* Implementation and visualization of a netflow log data lake system for cyberattack detection using distributed deep learning. *J Supercomput* (2022). https://doi.org/10.1007/s11227-022-04802-y
[9]. Mohamed, A., Najafabadi, M.K., Wah, Y.B. *et al.* The state of the art and taxonomy of big data analytics: view from new big data framework. *ArtifIntell Rev***53**, 989–1037 (2020). https://doi.org/10.1007/s10462-019-09685-9
[10]. Wang, J., Zhao, Y., Balamurugan, P. *et al.* Managerial decision support system using an integrated model of AI and big data analytics. *Ann Oper Res* (2022). https://doi.org/10.1007/s10479-021-04359-8
[11]. Wu, Z., Xue, R. & Shao, M. Knowledge graph analysis and visualization of AI technology applied in COVID-19. *Environ Sci Pollut Res***29**, 26396–26408 (2022). https://doi.org/10.1007/s11356-021-17800-z
[12]. Liu, H., Zhang, Y., Lian, K. *et al.*Health care data analysis and visualization using interactive data exploration for sportsperson. *Sci. China Inf. Sci.***65**, 162101 (2022). https://doi.org/10.1007/s11432-021-3412-9
[13]. Al-Shomar, A.M., Al-Qurish, M. &Aljedaani, W. A novel framework for remote management of social media big data analytics. *Soc. Netw. Anal. Min.***12**, 172 (2022). https://doi.org/10.1007/s13278-022-00996-4
[14]. Yin, Xd. Dynamic data driven big data cooperative control scheme with virtual visualization for mobile multimedia communication. *Cluster Comput***22** (Suppl 1), 1541–1548 (2019). https://doi.org/10.1007/s10586-018-2175-2
[15]. Tripathi, M.K., Nath, A., Singh, T.P. *et al.* Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. *Mol Divers***25**, 1439–1460 (2021). https://doi.org/10.1007/s11030-021-10256-w
[16]. Sen, S., Agarwal, S., Chakraborty, P. *et al.* Astronomical big data processing using machine learning: A comprehensive review. *Exp Astron***53**, 1–43 (2022). https://doi.org/10.1007/s10686-021-09827-4
[17]. Ram Kumar, Manoj Eknath Patil ," Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies", Turkish Journal of Computer and Mathematics Education ,Vol.13 No.3(2022), 987-993.
[18]. Vibhor Mahajan, Ashutosh Dwivedi, SairajKulkarni,Md Abdullah Ali, Ram Kumar Solanki," Face Mask Detection Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022