

Web News Mining using Back Propagation Neural Network and Clustering using K-Means Algorithm in Big Data

Sukhpal Kaur and Er. Mamoon Rashid

Department of Computer Science Engineering, Chandigarh University, Gharuan, Mohali – 140413, Punjab, India;
sukhs268@gmail.com, mamoon873@gmail.com

Abstract

In today's era, technology increasing day by day that changes the needs and minds of human beings. In this modern world handling data and information is the main problem. Many new technologies are developed for gathering the large data like Big data. In Big data many techniques and technologies are there like K-means. In this paper the main focus of research is to manage huge amount of news like BBC news channel data from the internet. To overcome the problem of managing the web news data, a clustering based K-means and Back Propagation Neural Network algorithm for classification is proposed. In this news are classified and categorized on the basis of text and content to provide the accurate news with less running time.

Keywords: Back Propagation, Big Data, Clustering, Data Mining, Web Mining

1. Introduction

Big data was introduced by Roger Magoulas from O'Reilly media in 2005 to define the large amount of data which traditional data management technique cannot manage and process because of large size and complexity of data. Big data is measured in terabyte and petabyte. Big data is a term which refers to data sets or combination of data sets whose size, complexity and degree of growth make them difficult to be captured, directed and processed by conventional technologies and tools such as relational database and desktop statistics or visualization package, within the time necessary to make them useful.

Data mining is the process of draw out the information from the large data set and converts it into the valuable and understandable form for using. Element of data mining is to get the data from data warehouse and to change form of data store and direct the data in multidimensional database system.

Text mining is the information which is used to store the data. Text mining is the combination of two words "text" and "mining". It means to get the large volume of text to find the most suitable information. It is related to unstructured data. Text is also called text data mining or knowledge discovery.

Web news mining is the collection large volume of public information. Web news is especial type of data available on web. On the web large amount of news are modified and originate. Web news is classifying the information into navigation hints, title, main text, picture, video, advertisements etc. Web news is third most used application of internet. Web news mining is process of to get the most important information very quickly. Web news mining is the feature of text mining. Web news are becoming an important part of the internet user. The benefit of the web news, the user can read the news anywhere, anytime on their mobile phone. But on the internet huge amount of news are available and it is

difficult for the user to read all the news. The benefit of web news mining get the news in short time.

2. Classification of Clustering:

There are various types of clustering algorithm:

2.1 Hierarchical Algorithm

It is based on distance matrix approach for clustering the data. It constructs the cluster step by step. It is used the bottom up and top down approach. Agglomerative nesting and divisive analysis is two techniques are used. Agglomerative nesting is top-down approach and divisive analysis is Bottom-up approach.

2.2 Partitioning Algorithm

In Partitioning Algorithm data is split into various subsets. The cause of splitting the data into various subsets is that examining all probable sub-system is computationally not possible and various greedy heuristic schemas are used in the form of iterative optimization. K-means is partition based algorithm.

2.3 Grid based Algorithms

In grid based algorithms, the objects are collected in the form of grid. The space of object is quantized into the number of cells. This algorithm has fastest processing time.

2.4 Density based Algorithm

This algorithm is based on the density. The basic idea is to continue increasing the given cluster whereas the density in the neighborhood exceeds some threshold. It is based on two approaches: Density reachability and density connectivity. DBSCAN is the example of density based algorithm.

News classification or Text classification is a problematic in computer science. The assignment is to assign a document to one or more classes. This may be done by using algorithm or “manually”. The intellectual organization of forms has mostly been the area of library science, in information knowledge and computer science the algorithmic classification of papers is used. The problems are intersecting, however and there is therefore also multidisciplinary investigation on document classification. The documents can be classified

in texts, imageries, music, etc. Each type of document owns its special classification issue. When not otherwise specified, text cataloguing is implied. Documents may be confidential according to their subjects or according to other attributes (such as text type, author, production year etc). In the rest of this article only subject classification is measured. There are two main attitudes of subject classification of documents i.e. the content based approach and the request based approach. Manually classifying millions of documents require lots of effort and time. Therefore, automatic News Text Classifier is constructing to manage these documents efficiently in less time with minimum efforts. The system will classify input text or document into predefined classes like Entertainment news, Political news, Business news, Health related news etc. We can use these unstructured documents and after classification process, we can convert them into structure information like table. In these tables, we can show the results like patient reports in hospitals are often indexed from numerous aspects using classifications categories of disease, kinds of surgical procedures, insurance compensation codes etc.

3. Related Work

This paper shows the Big data information and characteristics used in world wide. The issues are also mentioned to give idea about the Big data issues in real time¹. Cloud environment is widely used in industry and research aspects; therefore security is an important aspect for organizations running on these cloud environments. Using proposed approaches, cloud environments can be secured for complex business operations². In Big data various issues and challenges are arises like Big data storage, management and processing³. In this paper, proposed a systematic framework of secure sharing of sensitive data on Big data platform, which ensures secure submission and storage of sensitive data based on the heterogeneous proxy re-encryption algorithm⁴. In this paper the authors purposed worker quality evaluation algorithm that is implemented to any difficult crowd sourcing tasks without pre-known results⁵. In this paper, authors solve the clustering by novel distributed implementation of CA that is solved by used a combined data-node parallelization approaches using the MapReduce framework⁶. In this paper authors described i2MapReduce, a MapReduce-based framework for

increase the processing of Big data⁷. In this the authors described that the whole government system can make benefits from using Big data techniques⁸. This paper conclude that by putting up a holistic watch on how to leverage state-of-the-art technique to significantly make better the performance of Hadoop MapReduce jobs⁹. SOLAP is a useful technique to working on long-time spatial data. In this research paper, the authors aggregate the analysis power of SOLAP and the calculating ability of MapReduce¹⁰, proposed Dache, dache is a data-aware cache description strategy, protocol and architecture¹¹. The authors presented the very important optimization chances in the MapReduce framework in premises of decreasing consumption of energy¹². In this research paper, it is described that the unstructured data can be structured and processed with MapReduce component of Hadoop¹³. The paper describes the Big data with 3 Vs, Volume, Velocity and Variety. The paper also identified on Big data processing issues¹⁴. In this paper authors explained that the traditional e-Learning architecture is old and a modern data model that incorporates Hadoop to the existing system is rising in the IT industry¹⁵. In this paper, authors point out the challenges of latency critical Big data computing in financial and propose that the latency demand should be considered in five system levels from the benchmark design down to the hardware¹⁶. This paper inquires sensor-based forecasting in the circumstance of event-organizing places, which introduce an especially unmanageable script because of prominent variations in usage due to the entertained effects¹⁷. In this paper, authors explore a technique to make better the performance of the Hadoop MapReduce framework by optimizing the job and task implementation technique¹⁸. In this paper the authors discussed a data mining considering the organic grape juice and made a predictive models that are able to separate organic grape juice and ordinary grape juice according to components of chemical and elaborated the chemical components which most useful for separating organic grape juice from ordinary grape juice¹⁹. This paper puts up a general idea of data mining and data techniques and data mining in various fields²⁰. Data mining uses classification, prediction, clustering and association techniques to mine the data²¹. Data mining is combination of database and artificial intelligence technology²². This paper focused to extracting the essential knowledge from the text file²³. This paper explained that the text mining techniques and tools are

very close to information professionals because they are related to the information organization, retrieval and dissemination²⁴. This paper is focused to extract the essential knowledge from the text file²⁵. This paper provided a small introduction of text data mining²⁶. In this paper focused on the finding the template from plenty of text files and explore for required template that user wish²⁷. In this paper big research area - Text Mining and technique that are used in text mining are discussed²⁸. This paper described a very small introduction to the text mining and its research status²⁹. In this paper introduced a technique to draw out the primary text of web news. Web news pages clustered and the main texts are determined by applying mutual extracting path³⁰. In this paper, the authors introduced an approach to draw out the main text of web data. Web news pages were clustered and this technique is based on filtered noise of pages³¹. In this research paper the author proposed a Aging Bloom Filter (ABF) to maintain the data about which type of news is well liked currently. Online news is the third most used application of internet³². This paper brings up that to draw out the keyword from the web pages of news, the optimization algorithm Back Propagation Neural Network LM is used³³. In this paper, the authors uses LM optimization algorithm of BP neural network to draw out the keywords from news webpage³⁴. In this paper, the author proposed a web information extraction model WIEByPP and DPPMining is designed to³⁵. In this paper, to find the solution the difficulties of news text extraction, the video text-detection algorithm is composed³⁶. In this paper, the author proposed a video text detection-based algorithm to find the solution the news content extraction problem. This method is used for recognition accuracy and time cost³⁷. In this paper, the authors made an efficient technique for detection of short-term hot topic³⁸. This paper described that unsupervised Back Propagation Neural Network finds out any unknown amount of data³⁹. In this paper the author conclude that any unknown sample of data can be found out by using unsupervised, back propagation neural network, also⁴⁰. In this paper the authors conclude that the Back Propagation algorithm repeats the steps until the error is kept down to the negligible value and passes the error tolerance rate⁴¹. In this paper found that the BPNN is a best tool to perform classification tasks for remote sensing images⁴². According to authors of this, the BPNN is a powerful tool to perform classification tasks for remote sensing images. However,

there is no clear and accurate design specification in the area of remote sensing image classification. The BP learning algorithm is used widely to train such networks⁴³. This paper analyzes the shortcomings of K-mean algorithm and also discusses three dissimilar algorithms that remove the limitations of K-mean algorithm. The first algorithm remove the limitation of specifying the value of K in advance, the second algorithm reduces computational complexity and also removes dead unit problem, third algorithm reduces the time complexity by using two simple data structure to store the information of each iteration which can be used in next iteration⁴⁴. In this paper author discussed that Pattern Recognition can be done using various methods⁴⁵. In these papers three main parts: First is to find the significances of all parameters in MUP are presented explicitly. Second, show how to find out parameters for dissimilar users to change their different memorizing and forgetting capabilities⁴⁶. In this paper, the authors an in-depth experimental analysis approach used in this paper to research MUP's adjustment in the online news services domain. It describes the forgetting, remembering and learning mechanism of users⁴⁷. This paper described the review of six cluster techniques⁴⁸. This paper described that data mining way to draw out data from the large volume sets of data and translate it into an apprehensible form for further use⁴⁹. This paper presents enhanced K-means algorithm which integrates a well ordered method to find the initial centroids. This method specifies the whole process of clustering in $O(n^2)$ time⁵⁰. The time related algorithm is used to create and modify the news stories online automatically and find the short term events and give rank of hotness according to aging theory⁵¹. To classify the web news according to text of the news the Evolving Fuzzy System is used and Evolving Fuzzy System is built human understandable models of different categories⁵². In this paper, the authors introduced the Rapid Miner Linked Open Data extension which is used for better predictive and descriptive models⁵³. Big data is huge volume of data which is consists of 5Vs characteristics. Big data analytics is the process of examining the large amount of data sets⁵⁴. Big data technologies are used for business organization but due to large amount of data it is not easy to control by an organization⁵⁵. In this paper, authors showed the serious decrease of the flapping time of the internal routing protocol by NS2⁵⁶. In this paper the

data aggregation using Hybridized content and computation aware partitioning and aggregation method is presented⁵⁷. In this paper the PAMAR technique are used to avoid the redundancy⁵⁸. In this paper the essential tweets posted in our application by comparing them with various other tweets on the social networking site and get the better result⁵⁹. The aim of this paper is to suggest in the explorative way the methods for fire prevention and suppression using big⁶⁰.

4. BPNN and K-Means

Back Propagation Neural Network is a supervised learning technique in which expected output is already recognize to the system. Back Propagation algorithm performs the iteration of weight adjustment until the error is decreased. In Back Propagation algorithm has three layers input layer, hidden layer and output layer. These layers made of neuron which are linked to each other and make whole system. Weights are assigned on the links which measure the strength of signal. The weight value is calculated according to input signal and the error function back propagated to the input layer. Hidden layer is work for updating the weight function sporadically under its rate decrease or come on acceptable or error tolerance factor.

Figure 1 shows the structure of BPNN. The Back Propagation algorithm has two phases training phase and testing phase. In training phase the training data samples are supply at input layer to train the network with Pre-determine set of classes. In testing phase, the test data which is come which come with input layer used for prediction of applied pattern. The desired output is already known to the network. If the calculated output does not match with expected output the deviation in the result is propagate back to input layer. This process is continued whenever the error is decrease to a negligible amount. The Figure 2 is explaining the flow chart of BPNN.

Steps:

- Randomly choose the initial weights.
- For each training pattern apply the inputs to the network.
- Calculate the output for every neuron from the input layer to the output layer via hidden layer.
- Calculate the error at the outputs.

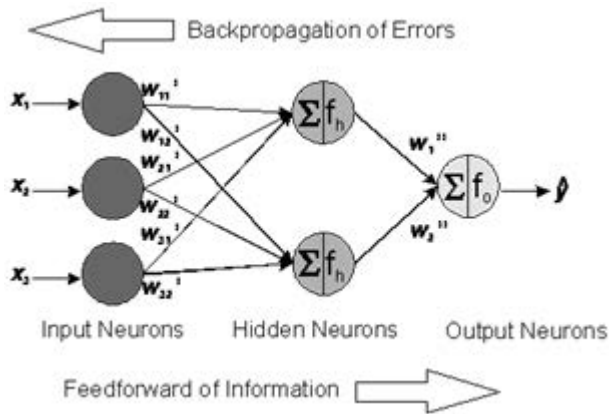


Figure 1. BPNN structure.

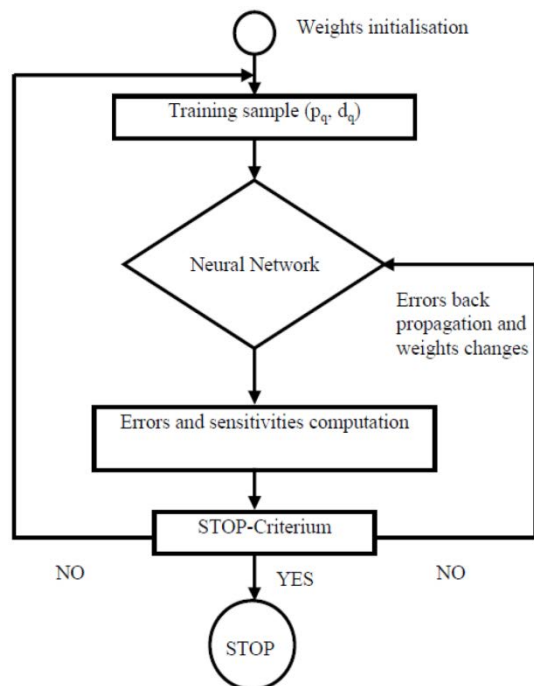


Figure 2. Flow chart of BPNN.

4.1 K-Means

K-means is clustering algorithm which is mostly used in data mining. K-means numerical, non-supervised, iterative and non-determine method. It is simple and very effective method that can make good clustering result. Each cluster is representing by the mean value of objects in K- mean clustering. In K-clustering algorithm, K centroid randomly selected, the value of K is fixed in advance. Each object of data set is associated to the nearest centroid. Similarity is calculated in term of mean value of object in cluster. To find the cluster centriod and distance between data object Euclidean distance is used. K-mean

is non-hierarchical clustering technique. Figure 3 is the flow graph of K- Means that explain the steps of K-Means.

Steps:

- Randomly select K data objects from dataset D as initial cluster centers.
- Repeat;
- Calculate the distance between each data object d_i ($1 \leq i \leq n$) and all k cluster centers c_j ($1 \leq j \leq k$) and assign data object d_i to the nearest cluster.
- For each cluster j ($1 \leq j \leq k$), recalculate the cluster center.
- Until no changing n the center of clusters.

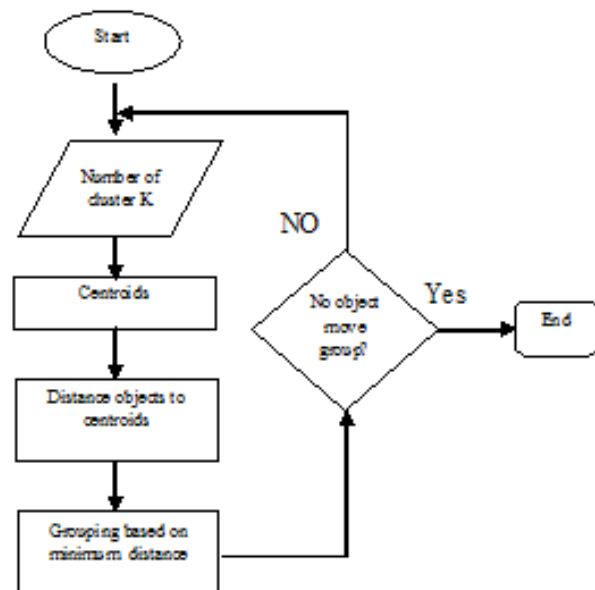


Figure 3. Flow chart of K-means.

4.2 Proposed Work

The aim of this research paper is to propose a model which classifies the news according to their categories based on the contents and text. The authors first classify and studied clustering and clustering techniques. Clustering is the process of grouping the similar data object. Cluster is the grouping of homogenous data. Various clustering and classification techniques are used to cluster the data objects.

In this research work authors proposed an approach to classify the news according to categories like Sports, Movies, Normal and Politics and to makes the clusters of them according to their types and upload the news from BBC world news. This proposed model adds this news

into the cluster based on the content and text of news. In this research work BPNN used to check the performance of the proposed system because sometimes user upload the false news and system accept it as right news and sometimes user upload the correct news but system reject it as false data. This proposed system checks the accuracy of the news. This proposed system saves the time of user.

4.3 Implementation

In this research work MATLAB R2013a on Laptop 64 bit operating system and 2 GB RAM is used. In this proposed system a text classification system is made using K-means and BPNN. In this system, authors classified the news on content basis. In training phase of BPNN, news is divided into four categories i.e. Movies, Sports, Politics and Normal. On the basis of words or content the news are stored. This system can store single word and full news. In text file upload section, upload the news then apply K-means algorithm to make the cluster of news and BPNN algorithm is used to classify the news and check the performance of system. BPNN make the system automated. This proposed system also tests the accuracy of system.

5. Results

K-means shows the cluster of news in different colors. In this red color is showing Movie news, Blue color is showing Normal news, Magenta color is showing Politics news and Black color is showing Sports news and all news cluster has one centroids. Centroids are used to cluster the news according their categories.

Figure 4 shows the cluster of news using K-means. BPNN classify the news. BPNN is checking the accuracy of news. BPPN checks the no. of iterations, time to be completed, performance, Mu, Gradients and Validation of news.

Figure 5 shows the performance of BPNN. In this we have given the training of 50 iteration but it completes in 3 iterations and it takes 10 seconds to complete and it shows the performance, Gradient, Mu and validation checks. It plots the performance, training state and regression of system.

Figure 6 shows the performance of the system. The best validation performance is 1.9917 at Epoch 1 of the system.

Figure 7 show the training state of system in 3 epochs.

It shows the Gradient, Mu and Validation. In this figure the gradient is $6.1874\text{e-}13$ at epoch 3, Mu is $1\text{e-}06$ at epoch 3 and validation checks is 2 at epoch 3.

Figure 8 shows the training, validation, time and testing of the system.

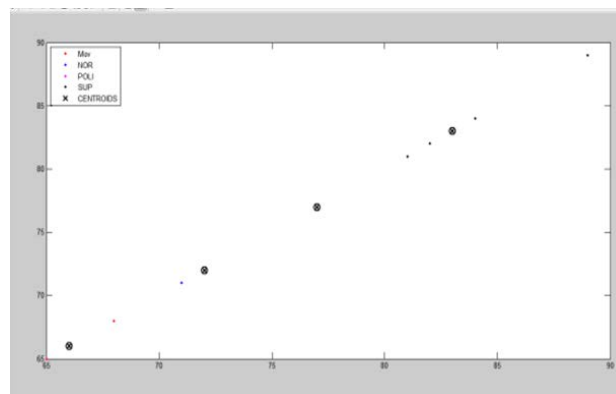


Figure 4. Showing the cluster of news (K-means).

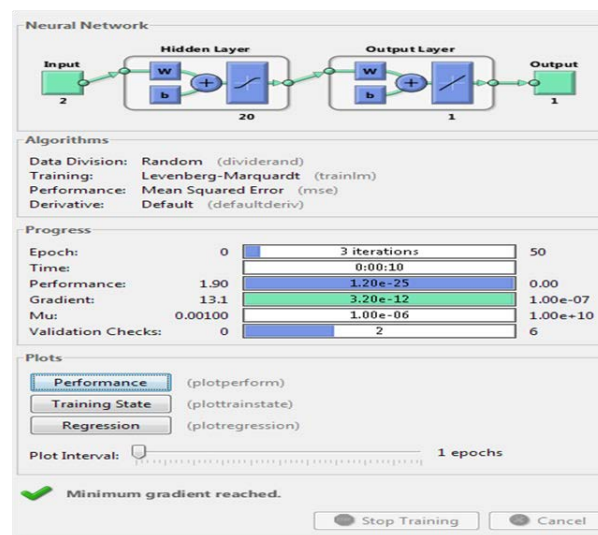


Figure 5. BPNN performances.

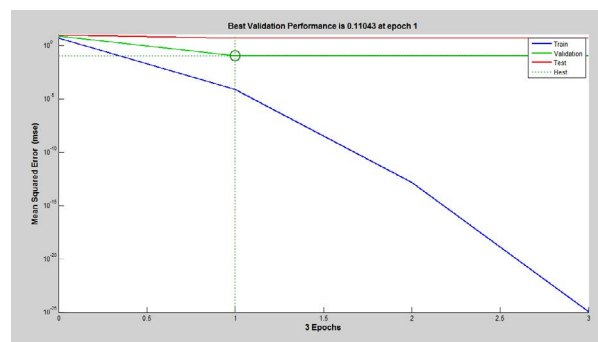


Figure 6. Plot showing the performance of system.

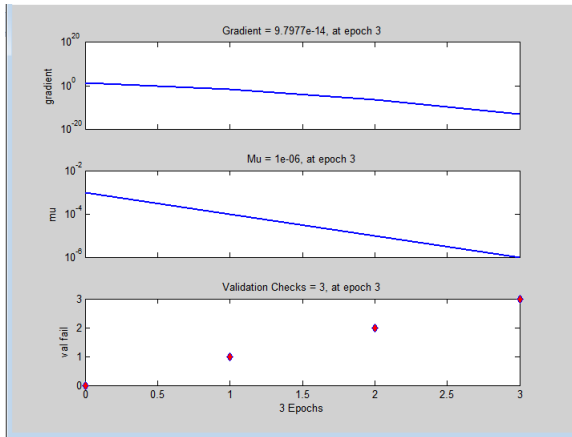


Figure 7. Plot showing the training state of system.

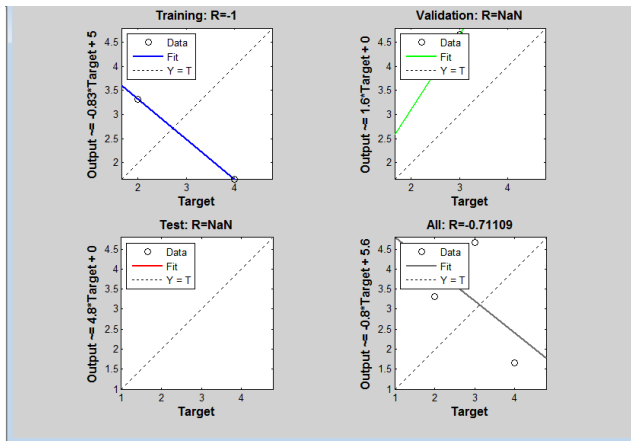


Figure 8. Plot showing the regression of system.

6. Conclusion

Many techniques and technologies are used in Big data like K-means algorithm. There are many problems arise to handle the huge amount of data. For this problem K-means and Back Propagation Neural Network algorithms are proposed for clustering and classification of news from BBC news channel. In this news are categorized as Movies, Sports, Politics and Normal. K-means algorithm is used to make the cluster and BPNN algorithm is used to classification of news that check the false error and rejection rate of system and tests the accuracy.

7. References

1. Saraladevi B, Pazhaniraj N, Victor PP, Basha MSS, Dhavachelvan P. Big data and Hadoop-A Study in Security Perspective. *Procedia Computer Science*. 2015; 50:596–601.
2. Narasimha IV, Sailaja A, Srinivasa Rao R. Security issues associated with Big data in cloud computing. 2014 May; 6:45–56.
3. Kaisler S, Armour K, Espinosa JA, Money W. Big Data: Issues and challenges moving forward.. 47th Hawaii International Conference on System Sciences; 2013. p. 995–1004.
4. Xinhua D, et al. Secure sensitive data sharing on a Big data Platform. *IEEE*. 2015 Feb; 20:72–80.
5. Dang D, Liu Y, Zhang X, Huang S. A crowdsourcing worker quality evaluation algorithm on MapReduce for Big data applications. *IEEE*; 2015. p. 1–11.
6. Hajj N, Rizk Y, Awad M. A MapReduce cortical algorithms implementation for unsupervised learning of Big data. *Elesvier*. 2015; 53:327–34.
7. Zhang Y, Chen S, Wang Q, Yu G. i2MapReduce: Incremental MapReduce for mining evolving Big data. *IEEE*. 2015 Jul; 27(7):1906–19.
8. Archanaa J, Mary Anita EA. A survey of Big data analytics in healthcare and government. *Elesvier*. 2015; 50:408–13.
9. Dittrich J, Quiane-Ruiz JA. Efficient Big data processing in Hadoop MapReduce. 2012; 5(12):2014–5.
10. Li J, Meng L, Wang FZ, Wen Zhang W, Cai Y. A MapReduce-enabled SOLAP cube for large-scale remotely sensed data aggregation. *Elesvier*. 2014 Sep; 70:110–9.
11. Zhao Y, Jie Wu J, Liu C. Dache: A data aware caching for Big data applications using the MapReduce framework. 2014 Feb; 19(1):39–50.
12. Mashayekhy L, Nejad MM, Daniel G, Zhang Q, Shi W. Energy-aware scheduling of MapReduce jobs for Big data applications. *IEEE*. 2015 Oct; 26(10):2720–33.
13. Subramaniaswamy V, Vijayakumar V, Logesh R, Indragandhi V. Unstructured data analysis on Big data using MapReduce. *Elesvier*. 2015; 50:456–65.
14. Bhosale SH, Gadekar DP. A review paper on Big data and Hadoop. 2014 Oct; 4(10):1–7.
15. Logica B, Magdalena R. Using Big data in the academic environment. *Elesvier*. 2015; 33:277–86.
16. Tian X, Han R, Wang L, Lu G, Zhan J. Latency critical Big data computing in finance. *Elesvier*. 2015; 1(1):33–41.
17. Grolinger K, L'Heureux A, Capretza MAM, Seewald L. Energy forecasting for event venues: Big data and prediction accuracy. *Elesvier*. 2016 Jan; 112:222–33.
18. Gu R, et al. SHadoop: Improving MapReduce performance by optimizing job execution mechanism in Hadoop clusters. *Elesvier*. 2014; 74(3):2166–79.
19. Maione C, et al. Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis. *Elesvier*. 2016;49:60–73.
20. Smita, Sharma P. Use of data mining in various field: A survey paper. *IOSR Journal of Computer Engineering*. 2014 May-Jun; 16(3):18–21.
21. Bora SP. Data mining and ware housing. *IEEE*; 2013. p. 1–5.
22. Ahmed AU, Ahmed CF, Samiullah M, Adnan N, Leung CKS. Mining interesting patterns from uncertain databases. *Elesvier*. 2016; 354:60–85.
23. Padhy N, Mishra P, Panigrahi R. The survey of data mining

- applications and feature scope. *IJCSEIT*. 2012 Jun; 2(3):43–58.
24. Verma VK, Ranjan M, Mishra P. Text mining and information professionals. *IEEE*; 2015. p. 133–7.
 25. Agnihotri D, Verma K, Tripathi P. Pattern and cluster mining on text data. *IEEE*; 2014. p. 428–32.
 26. Akilan A. Text mining: Challenges and future. *IEEE*; 2015. p. 1679–84.
 27. Aswini V, Lavanya SK. Pattern discovery for text mining. *IEEE*; 2014. p. 412–6.
 28. Sukanya M, Biruntha S. Techniques on text mining. *IEEE*; 2012. p. 269–71.
 29. Zhang Y, Chen M, Liu L. A review on text mining. *IEEE*; 2015. p. 681–5.
 30. Wang X, Wang WP, Liu B, Wang Z, Wang X. A novel approach to automatically extracting main content of Web News. *IEEE*; 2009. p. 1–4.
 31. Cheng K, Wang H. A web warehouse system for semi-automatically gathering and managing online news. *IEEE*; 2005. p. 343–44.
 32. Kosala R, Blockeel H. Web mining research: A survey. 2000 Jul; 2(1):1–15.
 33. Liu X, Yan X, Yu Z, Qin G, Mo Y. Keyword extraction for web news documents based on LM-BP Neural Network. *IEEE*; 2015. p. 2525–31.
 34. Thorleuchter D, Van den Poel D. Idea mining for web-based weak signal detection. *Elsevier*; 2014 Dec. p. 1–23.
 35. Wu GQ, et al. Web news extraction based on path pattern mining. *IEEE*; 2009. p. 612–7.
 36. Wu YC. Language independent web news extraction system based on text detection framework. *Elsevier*. 2016; 342:132–49.
 37. Yu RG, Zhao MK, Chang P, He MW. Online hot topic detection from web news archive in short terms. *IEEE*; 2014. p. 919–23.
 38. Sangeeta, Aujla GS. Prediction model for influenza epidemic based on twitter data. *International Journal of Advanced Research in Computer and Communication Engineering*. 2014; 3(20):1–4.
 39. Chaudhari JC. Design of artificial Back Propagation Neural Network for drug pattern recognition. *International Journal on Computer Science and Engineering (IJCSE)*; 2010. p. 1–6.
 40. Kaur T. Implementation of Back Propagation algorithm: A Neural Network approach for pattern recognition. *International Journal of Engineering Research and Development*. 2012 Jun; 1(5):30–7.
 41. Jaksa R, Katrak M. Neural Network model of the Back Propagation algorithm. p. 1–6.
 42. Suliman A, Zhang Y. A review on Back Propagation Neural Networks in the application of remote sensing image classification. *Journal of Earth Science and Engineering*. 2015 Jan; 5:52–65.
 43. Yadav J, Sharma M. A review of K-mean algorithm. *IJETT*. 2013 Jul; 4(7):2972–4.
 44. Na S, Xumin L, Yong G. Research on K-means clustering algorithm. *IEEE*; 2010. p. 63–7.
 45. Kishore R, Kaur T. Back Propagation Algorithm: An Artificial Neural Network approach for pattern recognition. *International Journal of Scientific and Engineering Research*. 2012 Jun; 3(6):1–4.
 46. Wang W, Zhao D, Luo H, Wang X. Mining user interests in web logs of an online news service based on memory model. *IEEE*; 2013. p. 151–5.
 47. Li Y, Wu H. A clustering method based on K-means algorithm. *Elsevier*; 2012. p. 1104–9.
 48. Joshi A, Kaur R. A review: Comparative study of various clustering techniques in data mining. *IJARCSSE*. 2013; 3(3):55–7.
 49. Mann AK, Kaur N. Survey paper on clustering techniques. *IJSETR*. 2013 Apr; 2(4):803–6.
 50. Abdul Nazeer KA, Sebastian MP. Improving the accuracy and efficiency of the K-means clustering algorithm. *WCE*. 2009 Jul; 1:1–5.
 51. Jain N, Srivastava V. Data mining techniques: A survey paper. *IJRET*. 2013 Nov; 2(11):116–9.
 52. Iglesias JA, Tiemblo A, Ledezma A, Sanchis A. Web news mining in an evolving framework. *Elsevier*. 2016; 28:90–8.
 53. Ristoski P, Bizer C, Paulheim H. Mining the web of linked data with RapidMiner. *Elsevier*. 2015 Frv; 35(3):142–51.
 54. Ishwarappa, Anuradha J. A brief introduction on Big data 5Vs characteristics and Hadoop Technology. *Elsevier*. 2015; 48:319–24.
 55. Maitrey S, Jha CK. MapReduce: Simplified data analysis of Big data. *Elsevier*. 2015; 57:563–71.
 56. Park HW, Yeo IY, Jang H, Kim NG. Study on the impact of Big data traffic caused by the unstable routing protocol. *Indian Journal of Science and Technology*. 2015 Mar; 8(5).
 57. Karthick N, Kalarani XA. An improved method for handling and extracting useful information from Big data. *Indian Journal of Science and Technology*. 2015 Dec; 8(33):1–7.
 58. Somasekhar G, Karthikeyan K. The pre Big data matching redundancy avoidance algorithm with MapReduce. *Indian Journal of Science and Technology*. 2015 Dec; 8(33):1–7.
 59. Dhamodaran S, Sachin KR, Rahul Kumar R. Big data implementation of natural disaster monitoring and alerting system in real time social network using Hadoop Technology. *Indian Journal of Science and Technology*. 2015 Sep; 8(22):1–4.
 60. Kim BS, Kim DY, Kim KW, Park ST. The improvement plan for fire response time using Big data. *Indian Journal of Science and Technology*. 2015 Sep; 8(23):1–5.