3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), 2018

Mining Application on Analyzing Users' Interests from Twitter

Arti Jain[a], Ashutosh Gupta[a], Nikhil Sharma[a], Shubham Joshi[a], Divakar Yadav[b]*

[a]*Department of Computer Science & Engineering, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India*
[b]*Department of Computer Science & Engineering, Madan Mohan Malviya University of Technology, Gorakhpur, Uttar Pradesh, India*

Abstract: In today's world, it is problematic to provide users of social-media with posts that are analyzed from their interest efficiently. Users are unable to see the good quality and variety of posts based on their interest. The mass adoption of smartphones along with an internet connection via wi-fi or cellular network enables to analyse users' interest from Twitter. Twitter is used by a large number of audience to share their posts on a variety of topics as tweets. Then mining users' interests from Twitter can amplify a number of efficacies, such as advertising, trending topics that can be analyzed by interests and recommendation of users' posts. For this purpose, this paper provides an Android application which incorporates Web Services, Jsoup, JSON, Firebase Real-time Database and MVC. The application aids to select the posts which include spectacular images and text that are shown to users as a training set. The personalized posts can later be inferred and analyzed by the users themselves using Suffix Array Data Structure and Artificial Neural Network (ANN). Under ANN, we have used Backpropagation methodology that fires neurons as posts. Kosaraju algorithm and Palette library then help in removing redundant posts while later one also retaining relevant posts with specified hashtags more efficiently and accurately.

Keywords: Android, Backpropagation, Firebase, Kosaraju, Neural Networks, Suffix Array, Twitter.

1. Introduction

With a never-ending growth in the field of mobile technology and swelling usage of smartphones, mobile applications have become a major necessity. In recent years, more and more people are becoming aware of smartphones (Rajurkar & Shirsagar, 2017). These smartphones act as key elements to transform the cultural, social, technological and other diverse aspects of modern civilization. This, in turn, impacts various sectors, namely- business, education, health, psychology etc. For 2020, the number of smartphones users[1] worldwide is forecast to reach 2.87 billion. In today's world, firstly, it is difficult to provide the user with post that are analyzed from his/her interest efficiently. Secondly, users sometimes, are unable to read long posts on some topic of their interests and prefer short posts most of the time. Thirdly, a picture is worth a thousand words (Xie, Pei, Xie, & Xing, 2015). Posts with images have a greater understanding than only with facts. Fourthly, users are unable to search every time for posts based on their taste. There must exist a recommendation system to identify users' interests and by analyzing recommending posts. Mass adoption of smartphones along with an internet connection via wi-fi (Atkinson, Mitchell, Rio, & Matich, 2018) or cellular network enables to analyze users' interests from social media. Social media (Turban, Outland, King, Lee, Liang, & Turban, 2017; Jiang, Qian, Mei, & Fu, 2016; Zarrinkalam, Fani, Bagheri, Kahani, & Du, 2015; Qian, Feng, Zhao, & Mei, 2014) plays a vital role in today's world as it enables users' to freely communicate with each other and share their recent news, ongoing activities or views about different topics or domains. One such important platform for social media is Twitter (Beauchamp, 2017; Kim & Hastak, 2018). Twitter is used by a large number of audience to share their posts on a variety of topics as tweets. These tweets allow users to post a range of tasks that vary from status updates, jokes, news, and events to other formal or informal information. Although tweets are limited to 140 characters but micro-blog format and ease of mobile posting allow users to post and follow their up-to the-minute activities. Twitter is labeled as "The SMS of the Internet" for being among the 10 most-visited websites. As a result, detecting users' interests from Twitter is grasping focused attention for mobile application developers. Tweets incorporate follow-ups of the local reporters, politicians, businesses and private citizens who post information about the community and allow followers to keep in touch with the current events. Community posts often include traffic reports, weather, local news, special events, restaurant reviews and other information that are useful to locals and visitors alike. National and world events, from television shows to natural disasters draws a lot of attention for Twitter users and become the topics of enough posts and discussion to appear as "Twitter Trending Topics". This generates users an idea of what is current, important and popular to Twitter users. In addition, individuals, businesses, and organizations use Twitter to promote themselves, drawing attention to events, raising awareness about their products and services and creating buzz around their brands and help in their customer service and support systems; watching for and responding to mention of their brands. Businesses can reward those who spread a positive message, often with a simple expression of gratitude. Negative coverage can often be circumvented by reaching out to unhappy customers to help resolve their issues. Many Twitter users, even those who tend to read instead of post, have taken advantage of the wide range of business, expertise, opinions, and knowledge that can easily be accessed via Twitter. By simply searching Twitter posts, it is possible to quickly find informal reviews of a product or service. Often, users post a question and quickly receive recommendations,

1 https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide,"Smartphones Users Worldwide 2014-2020", Accessed on: Feb. 11, 2017.

ideas, and offers of assistance in response. For example, "#MannkiBaat hashtag@PMOIndia" is used as a Twitter hashtag. This hashtag implicitly classifies the method that is used by Twitter to organize its tweets. In other words, mining users' interests from social media can amplify a number of efficacies, such as advertising, trending topics etc. that can be analyzed by interests and recommendation of users' posts.

In the current work, our focus is on users' interests and trending topics from Twitter, social media as posts with spectacular images and text that are shown to users as a training set. Then users are privileged to select these posts from the given data set, so that the personalized posts can later be inferred and analyzed by the users themselves. For the purpose of the same, we have integrated various technical specifications varying from Android, JSON, Jsoup, Webs Server, Real-time Database, MVC, Backpropagation to Kosaraju algorithm (refer Section 3). Section 2 describes the related literature work and how this research is differentiated from others. Section 3 discusses various technologies that are used in this work and the need for such technologies. Section 4 depicts the overall architecture of the proposed work. Section 5 shows the experimentation and the results that are obtained so far while accumulating certain additive features such as Timestamping, Suffix Array and Palette library. Section 6 concludes the paper.

2. Related Work

To mine users' interests (Xie, Pei, Xie, & Xing, 2015) researchers have applied user image latency model. They mine users' interests from images but are not efficient in terms of time and require huge data for a desired precision measure. Researchers (Hu, Zhang, Wang, & Li, 2012) have proposed users' interests based on textual and structural information that is obtained from comments. They do not predict if the user is really interested in the recommended post or not. Authors (Zarrinkalam, Fani, Bagheri, Kahani, & Du, 2015) have inferred users' interests and recommended posts according to the concept-based approach and have obtained topics from clustering concepts. Authors (Servia-Rodriguez, Fernandez-Vilas, Diaz-Redondo, & Pazos-Arias, 2013) have evaluated the performance of three different tag clustering algorithms– PAM, Affinity Propagation and UPGMA. They have used an unsupervised measure of the clustering quality i.e. Silhouette Width which estimates the parameters of the cluster analysis. Authors (Kaur, Talluri, & He, 2015) have implemented an Android application that initially asks the user for the search keyword and provides the respective tweets based on the input keyword by analyzing Twitter. Along with the tweets, the user can also get additional information including the username of the person who has posted the tweet, timestamp, count of re-tweets for a particular tweet, location, description and timeline of the user who has posted the tweet but all this is based on the information which the user provides. Researchers (Xue, Zhang, Zhou, Lin, & Li, 2008) have implemented a news recommendation system in the social media by analyzing users' preference. They have taken a number of votes as implicit feedback and the keywords extracted from the comments are ranked based on both quantity and quality of the comments they appear in. Finally, the top-ranked keywords are selected and merged with the keywords representative of the original topics to retrieve the relevant news. A lot of computation is needed to infer relevant news. Author (Beauchamp, 2017) have predicted state-level polls using Twitter but his main focus is on textual data. Authors (Jiang, Qian, Mei, & Fu, 2016) have personalized travel sequence recommendation on social media. Authors (Qian, Feng, Zhao, & Mei, 2014) have proposed personalized recommendation model which focus on user historical rating records but do not bother to recommend real-time items. Authors (Tran, Tran, & Uong, 2010) have proposed a recommendation system for Vietnamese electronic newspaper which uses content-based filtering techniques associated with the users' attention that are determined by inferring a set of common hidden topics from the documents which users have preferred.

However, social media as a crucial medium to convey users' interests has been largely ignored nevertheless. All of these mentioned literature-work although concern about texts, links, clicks, meta-data, social-clues and images; and uses various different techniques to infer the relevant posts or news. They are not much efficient in terms of time and complexity. In some research papers, either a lot of processing and computation is needed or researchers want the user to provide them with certain predefined information which is not be of much requirement in this current work.

In this current research, variety of users' interests and trending topics from Twitter as posts are considered along with the combination of images and text, and have shown to the users as a training set. Then the users are privileged to select these posts from the defined training set so that the personalized posts can later be inferred efficiently and analyzed by the users themselves, with the redundancy of the posts being removed using Palette library and Kosraju algorithm (refer Section 3 and Section 5).

3. Technical Background

Technical specifications, namely- Android, Shared Preferences, Web Server, Jsoup, JSON, Firebase, Model-View-Controller, Backpropagation, and Kosaraju Algorithm that are being used in this research are discussed here.

3.1. Android

Android (Walnycky, Baggili, Marrington, Moore, & Breitinger, 2015) is an open source mobile Operating System which is developed by Google. Android OS is based on the Linux kernel and is designed primarily for touch-screen mobile devices such as smartphones, tablets etc. Android works very well for a wide number of users. In addition, Android applications can be of various types, ranging from entertainment, fitness; office applications to personal applications. In this work, Android-based platform offers a unified approach to the Mining based Application Development Environment and its compatibility with JAVA as programming language offers a more robust and secure solution.

3.2. Shared Preferences

Android provides many ways of storing data for our mining application. One of these ways is called Shared Preferences. Shared Preferences allows saving and retrieving data in the form of a (*key, value*) pair. Here, "key" is defined as an application user, and "value" is defined as users' profile details. A method, <*getSharedPreferences()*> returns a Shared Preferences instance which points to the file that contains the values of the preferences. The Shared Preferences then stores a session for the users' which is valid until the users' sign-out of the Android application.

3.3. Web Server

Web Server through its web services[2] is used to read, store and retrieve users' details. It is used for signup pages (for storing users' data after successful registration) and login page (for secured access after matching users' details with respect to the data stored in the web server). In this work, 000webhost is used as the web server to maintain users' personal information.

3.4. Jsoup

Jsoup[3] is a Java library for working with real-world hypertext languages. It provides convenient API for extracting and manipulating data, using the best of DOM, CSS, and Jquery- like methods. In this research, Jsoup is used to extract mining content i.e. extraction of posts (both text and images) from social media, Twitter.

3.5. JSON

JavaScript Object Notation (JSON)[4] is a lightweight data-interchangeable format which is in very high demand because of its easy handling features. One can read and write JSON with ease and is also suitable for machines to parse and generate. JSON consists of a text format that is completely language independent but uses same conventions as that of the C-family of languages including C, C++, C#, Java, JavaScript, Perl, Python and many others. Due to these fantastic JSON properties in today's social media market, it is considered as an ideal data inter-changer for Android. In this research, JSON is used to send and retrieve users' mining information while communication with the Firebase Real-time Database.

3.6. Firebase Real-time Database

Firebase Real-time Database[5] is a NoSQL database and has different optimization and functionality as compared to the relational database. Here, data is persisted locally, and even while offline, real-time events continue to fire, giving the end users a responsive experience. When the device regains connection, this Real-time Database synchronizes the local data, does changes with the remote updates that have occurred while the client goes offline and so merging any conflicts automatically. This enables to build a real-time scenario that can serve many users without compromising on responsiveness and hence maintains users' mining details in an optimized way.

3.7. Model-View-Controller

Model–View–Controller (MVC)[6] is a software architectural pattern that is used for implementing user interfaces. MVC divides an application into three interconnected parts- Model, View, and Controller. These three interconnected parts separate internal representation of information from the ways that information is presented to or accepted from the users as is depicted in Fig. 1. MVC enhances the readability and clean design of the underlying work.

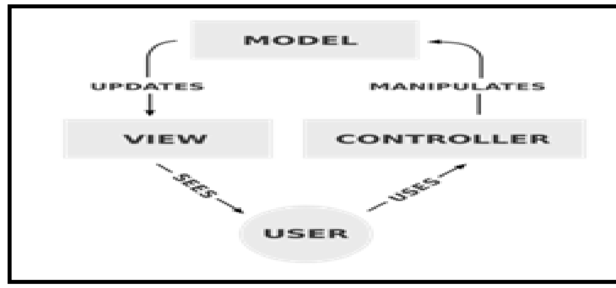

Fig. 1 - MVC Architecture.

- Model: Model represents the data of Controllers and also can be saved and restored when Views are killed.

- View: View (e.g. activity, fragment, service etc.) may be allowed only to read the Model. Since View does not have any business logic of its own so it is never able to modify the Model which then becomes the responsibility of Controller.

- Controller: Controller provides an API to receive commands from the View and process the data, and then it sends the processed data back to the View of the event. We just need to mock an event subscriber to test the event of the Controller. In this way, Controller does not have any awareness of what is View (an Android component or a mocked event subscriber), completely abstracted and so testing becomes much easier.

3.8. Neural Networks- Backpropagation

Backpropagation[7] is an abbreviation for "Backward Propagation of Errors", a common method of training Artificial Neural Networks (ANN) (Walczak, 2018; Shi, Bai, & Yao, 2017). Backpropagation (Gautam, Bhateja, Tiwari, & Satapathy, 2018) is used in conjunction with an optimization method such as Gradient Descent (Manogaran, & Lopez, 2018).

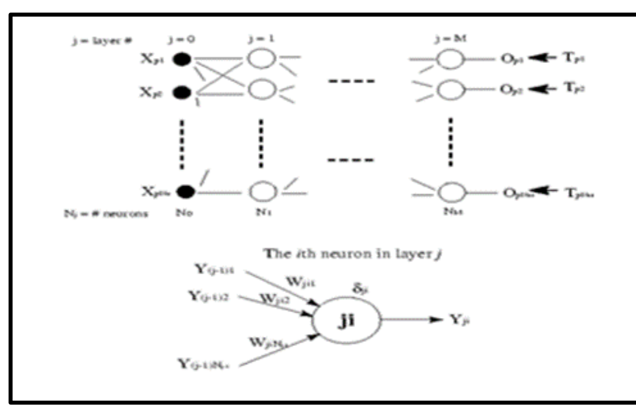

Fig. 2 – Backpropagation Strategy.

[2] https://in.000webhost.com/

[3] https://jsoup.org/

[4] http://www.json.org/

[5] https://firebase.google.com/docs/djjatabase/

[6] https://www.quora.com/Is-there-any-standard-MVC-framework-in-Android-application-development-If-not-is-it-worth-developing-one

[7] https://en.wikipedia.org/wiki/Backpropagation

Backpropagation calculates the gradient of a loss function w.r.to all the weights in the network. A gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function. Backpropagation actually requires a known, desired output for each input value in order to calculate the loss function gradient as is seen in Fig. 2. It is, therefore, usually considered to be a supervised learning method, although it is also used in some unsupervised methods such as auto-encoders. In this paper, backpropagation is used for the proper arrangement of users' posts. Here, Y_{ji} represents an output from the i^{th} neuron in layer j. N_j represents the number of neurons in the j^{th} layer. p^{th} represents the pattern of training sample with No representing dimensional input as $X_{p1}, X_{p2}\ldots, X_{pNo}$. Nm represents dimensional known output response $T_{p1}, T_{p2}\ldots, T_{pNm}$. The actual response to the input pattern by the network is represented as $O_{p1}, O_{p2}\ldots, O_{pNm}$. Net_{ji} represents the sum-of-products of the neuron. W_{jik} represents the connected weights and δ_{ji} represents the error value associated with the i^{th} neuron in the layer j.

3.9. Kosaraju Algorithm

Kosaraju algorithm[8] is Depth First Search (DFS) based algorithm. A directed graph (Roditty & Zwick, 2016) is strongly connected if there is a path between all pairs of vertices. Strongly Connected Component (SCC) (Alshomrani & Iqbal, 2012) of a directed graph is a maximal strongly connected sub-graph. It does DFS two times. DFS of a graph produces a single tree if all vertices are reachable from the DFS starting point. In the next step, it reverses the graph. Here, the algorithm is used for removal of the redundant posts.

```
// The function that discovers and prints all SCCs
  void SCCs()
  {
    Stack st = new Stack();

    // Vertices are marked as not visited (For the first DFS)
    boolean visit[] = new boolean[length];
    for(int j = 0; j < length; j++)
       visit[j] = false;

    // Vertices are filled in stack according to their finishing
    // times
    for (int j = 0; j < length; j++)
       if (visit[j] == false)
          fillOrder(j, visit, st);

    // Reversed graph is created
    Graph graph = getTranspose();

    // Vertices are marked as not visited (For the second
DFS)
    for (int j = 0; j < length; j++)
       visit[j] = false;

    // Now, All vertices are processed in order defined by
Stack
    while (st.empty() == false)
    {
       // Pop a vertex from stack
       int vertex = (int)st.pop();

       // SCC of the popped vertex are printed
       if (visit[vertex] == false)
       {
          graph.DFSUtil(vertex, visit);
          System.out.println();
       }
    }
  }
```

Fig. 3 – Kosaraju Algorithm.

Fig. 3 states the detail description of the Kosaraju algorithm. The algorithm does DFS two times. For the first DFS, the vertices are marked as not visited and are filled in a stack according to their finishing time. Then again vertices are marked as non-visited for the second DFS. Now, all the vertices are processed in an order that is defined by the stack. Vertices are then popped from the stack and SCC of the popped vertex is printed out. For example, there are 3 SCCs as in the Fig. 4. Here, 0, 1, 2 have paths between them. So, it is taken as a single strongly connected component. 3 and 4 formulate the other two strongly connected component separately one after other. In the next step, their directions of pointing to different SCCs are just reversed.

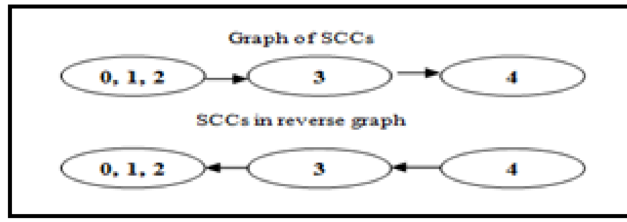

Fig. 4 – Example of SCC.

4. System Architecture

The system architecture is in Fig. 5, variety of users' interests and trending topics are taken from Twitter as posts that incorporate images as well as text through the Android application.

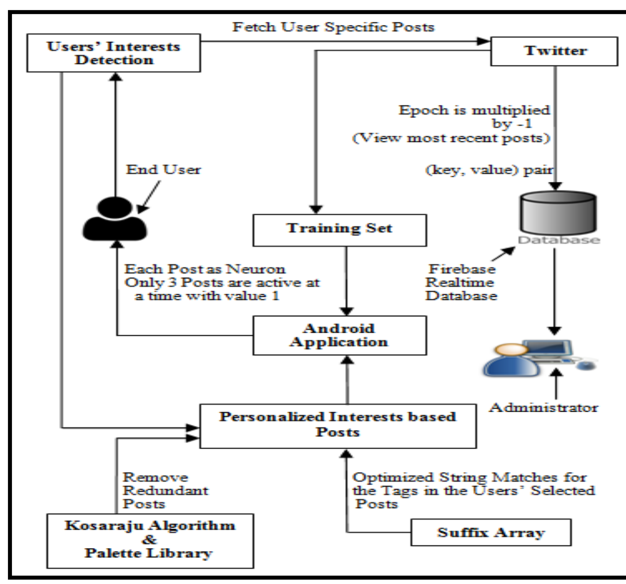

Fig. 5 – System Architecture.

Above posts are given to users as training set which is then handled by Firebase Real-time Database. Users are then privileged to select their posts

[8] http://www.geeksforgeeks.org/strongly-connected-components/

from the defined training set so that the personalized posts can later be inferred and analyzed by these users themselves. Each post epoch is multiplied by *-1* so that the most recent posts come first through the database where all the posts are stored in (*key, value*) pair. Suffix array is used to optimize string matching for the tags in the users' selected posts. Kosaraju algorithm and Palette library are applied to remove redundant posts. Each post act as a neuron and only 3 posts are active at the same time with value *1*.

5. Experimentation and Results

Section 5 details about the experimentation that are conducted for this research work and the results that are obtained while analyzing users' interests from the Twitter.

Primarily, an existing user is navigated to sign in into his/her account or sign up as a new user for the Android application. Throughout the application (*key, value*) pairs are used in such a way that the data is secured by these values and thus ensuring that no other key can access or modify other value posts. Right from storing into the Firebase Real-time Database to having parallel database connections, everything is being done throughout these pair values. Hash sets, on the other hand, have helped to ensure that no duplicate tags are used anywhere in the proposed methodology.

Fig. 6 shows the sample training set that reflects the multimedia or variety of interests or trending topics as posts with images and text from Twitter. These are shown to users so that the user can then be privileged to select multiple posts which are then provided as input for the further analysis.

Fig. 6 – Sample Training Set.

In order to view the most recent posts, time stamping is generated at Firebase as is shown in Fig. 7. The timestamp is considered while multiplication with a factor of *-1* and thus, are sorted in ascending order so that the posts are sorted according to the timeline. Thus, it provides a unique and fastest way to make the post appear to be sorted according to time, making users see only the most recent posts.

Fig. 7 – Firebase Real-Time Database.

For further providing modularity and fast searching in the application for the tags in the posts which the users select, Suffix Array (Xylogiannopoulos, Karampelas, & Alhajj, 2016) based Data Structure (DS) is used. This DS not only provides fast access to elements but also helps in finding the longest prefix match.

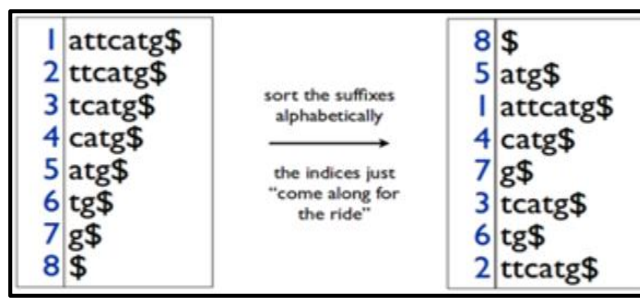

Fig. 8 – Suffix Array Implementation.

As it is seen in Fig. 8 implementing the suffix array DS is quite efficient as it reduces the time complexity to $O(n\log_2 n)$ where n is the number of characters in the tag that exist within the given post which is better than the other string matches algorithm in terms of complexity. So, the optimized string matching is being done for the tags in the posts which the users select.

Now the each and every data is being navigated to the server and is stored as linked-list having all the data as its "Tail" and reference to the link as its "Head" and backward. This structure is what the ANN, Backpropagation actually uses. An important aspect of this ANN is the assurance and display of users' posts with interest accordingly. Here, each neuron has exactly two states i.e. *0 or 1* which depict *False or True* respectively for each post that is fetched as is seen in Fig. 9. In case, if a state *0* happens then two more new neurons are fetched in the path making a chain-reaction which at the end ensures that the user sees only the post in which he or she is interested in. Here, the neuron is an activity on the screen.

```
void test() {
    mList.add(new Items("A", "B", 1));
    mList.add(new Items("AA", "BB", 0));
    mList.add(new Items("AAA", "BA", 0));
    mList.add(new Items("AV", "BC", 1));
    mList.add(new Items("A", "B", 1));
    mList.add(new Items("AA", "BB", 0));
    mList.add(new Items("AAA", "BA", 0));
    mList.add(new Items("AV", "BC", 1));
    mList.add(new Items("A", "B", 1));
    mList.add(new Items("AA", "BB", 0));
    mList.add(new Items("AAA", "BA", 0));
    mList.add(new Items("AV", "BC", 1));
    mList.add(new Items("A", "B", 1));
    mList.add(new Items("AA", "BB", 0));
    mList.add(new Items("AAA", "BA", 0));
    mList.add(new Items("AV", "BC", 1));
    mList.add(new Items("A", "B", 1));
    mList.add(new Items("AA", "BB", 0));
    mList.add(new Items("AAA", "BA", 0));
    mList.add(new Items("AV", "BC", 1));
    mList.add(new Items("A", "B", 1));
```

Fig. 9– Neuron Lifecycle.

The proposed methodology does not use Recycler-Viewer but includes Neuron Lifecycles that are implemented for the neurons as a Vertical View Pager. Each moment only 3 neurons are in the life, and when the user navigates by scrolling up or down the neurons they may get killed and be reborn repeatedly. The result of one such neuron is shown in Fig. 10.

Fig. 10– Personalized Users' Interest Based Post.

Personalized feeds with images and texts are analyzed from different posts which users may select. Similarly, there is 1 post above and 1 post below i.e. 2 other neurons (totaling 3 neurons). These 3 neurons or posts are active at the same time. Certain posts now have the problem of redundancy. For this purpose, Kosaraju algorithm is implemented to remove such redundant posts. As it may be possible that many online marketers tweet similar post at the same time in order to make that post viral within seconds. This may help them to reach a wide group of an audience within a short span. So, strongly connected directed graphs are constructed by considering the similar posts or redundant posts within a given time and are shown to the users as a single post with no redundancy at all. It is observed that the Kosaraju algorithm although removes redundant post but certain relevant posts are also being deleted which must be necessarily shown to the users. In order to take care of this aspect, Palette library is used.

In the Palette library, different colors for a particular post are added as in Fig. 11. Each image comprises of a lot of variants in color e.g. Red color contribution is 30% and density of Red is depicted as 36%.

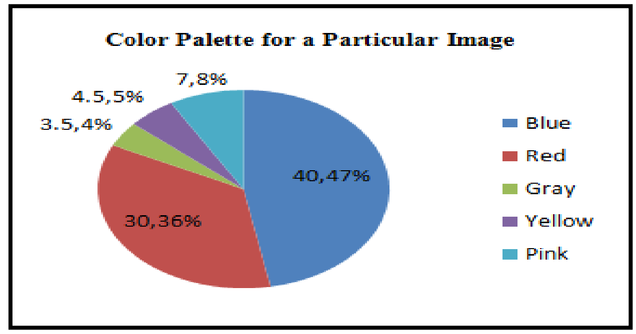

Fig. 11– Colour Palette for an Image.

Also, with the usage of the Palette library, different profiles are extracted from the posts having images that are clustered according to same color group as it is seen in Fig. 12. Thus, it helps in removing redundant posts and retaining relevant posts more efficiently and accurately.

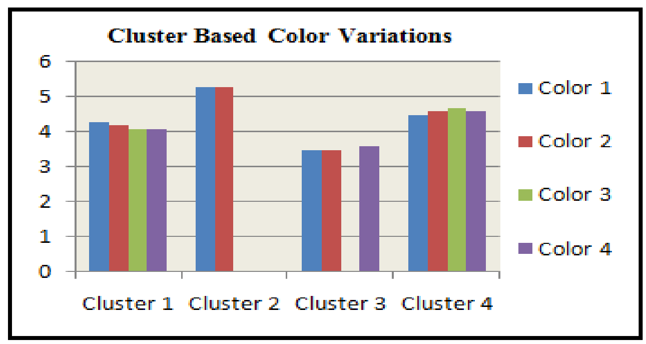

Fig. 12– Clustered Posts.

6. Conclusion

Android-based mining application on analyzing users' interests from Twitter is effective to recommend the interested topic's posts. These posts accompany both spectacular images and text that are successfully being analyzed by users' interests and inferring personalized interests'-based posts efficiently. Posts are coming in the time stamping order when we multiply each post epoch value by a factor of -1. Each post, activity acts as a neuron, where at time 2 more neurons are active, one above and below the current neuron, totaling 3 posts at a time that are fired as neurons using Backpropagation. Suffix array data structure is quite efficient as it reduces the time complexity to $O(n\log_2 n)$ where n is the number of characters in the tag that exist within the given post which is better than the other string matches algorithm in terms of complexity. So, the optimized string matching is being done for the tags in the posts which the users select. Redundancy among posts is not visible as it is removed using Kosaraju algorithm but certain legitimate posts are also being deleted which must be necessarily shown to the users. Palette Library is used where each post comprises of lots of color variants and color densities and these different profiles are extracted from the posts having relevant hashtags that are clustered according to the same color group and shown to users as a single post. Thus, all the legitimate posts that are not redundant are visible to the users and successfully inferring users' interest.

REFERENCES

Alshomrani, S., & Iqbal, G. (2012). Analysis of Strongly Connected Components (SCC) Using Dynamic Graph Representation. *International Journal of Computer Science Issues*, 9(4).

Atkinson, J. S., Mitchell, J. E., Rio, M., & Matich, G. (2018). Your WiFi is leaking: What do your mobile apps gossip about you? *Future Generation Computer Systems*, *80*, 546-557.

Beauchamp, N. (2017). Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61(2), 490-503.

Gautam, A., Bhateja, V., Tiwari, A., & Satapathy, S. C. (2018). An improved mammogram classification approach using back propagation neural network. In *Data Engineering and Intelligent Computing* (pp. 369-376). Springer, Singapore.

Hu, C., Zhang, C., Wang, T., & Li, Q. (2012, January). An adaptive recommendation system in social media. In *International Conference on System Science (HICSS), 2012 45th Hawaii* (pp. 1759-1767). IEEE.

Jiang, S., Qian, X., Mei, T., & Fu, Y. (2016). Personalized travel sequence recommendation on multi-source big social media. *IEEE Transactions on Big Data*, *2*(1), 43-56.

Kaur, H., Talluri, M., & He, J. S. (2015, June). Get Twitter information: A collaborative Android application for big data analysis. *In International Conference on Collaboration Technologies and Systems (CTS),* 2015 (pp. 483-484). IEEE.

Kim, J., & Hastak, M. (2018). Social network analysis. *International Journal of Information Management: The Journal for Information Professionals*, 38(1), 86-96.

Manogaran, G., & Lopez, D. (2018). Health data analytics using scalable logistic regression with stochastic gradient descent. *International Journal of Advanced Intelligence Paradigms*, *10*(1-2), 118-132.

Qian, X., Feng, H., Zhao, G., & Mei, T. (2014). Personalized recommendation combining user interest and social circle. IEEE transactions on knowledge and data engineering, 26(7), 1763-1777.

Rajurkar, N., & Shirsagar, P. (2017). Impact of smartphones on society. *International Journal of Research in Science & Engineering*, *3*(2), 143-150.

Roditty, L., & Zwick, U. (2016). A fully dynamic reachability algorithm for directed graphs with an almost linear update time. *SIAM Journal on Computing*, *45*(3), 712-733.

Servia-Rodriguez, S., Fernandez-Vilas, A., Diaz-Redondo, R. P., & Pazos-Arias, J. J. (2013, September). Comparing tag clustering algorithms for mining Twitter users' interests. In *International Conference on Social Computing (SocialCom)*, 2013 (pp. 679-684). IEEE.

Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2298-2304.

Tran, M. V., Tran, X. T., & Uong, H. L. (2010, December). User interest analysis with hidden topic in news recommendation system. In *International Conference on Asian Language Processing (IALP) (pp. 211-214). IEEE.*

Turban, E., Outland, J., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2017). *Electronic Commerce 2018: A Managerial and Social Networks Perspective*. Springer.

Walczak, S. (2018). Artificial neural networks. In *Encyclopaedia of Information Science and Technology, Fourth Edition* (pp. 120-131). IGI Global.

Walnycky, D., Baggili, I., Marrington, A., Moore, J., & Breitinger, F. (2015). Network and device forensic analysis of android social-messaging applications. *Digital Investigation*, *14*, S77-S84.

Xie, P., Pei, Y., Xie, Y., & Xing, E. P. (2015, January). Mining User Interests from Personal Photos. In *Association for the Advancement of Artificial Intelligence* (pp. 1896-1902).

Xue, Y., Zhang, C., Zhou, C., Lin, X., & Li, Q. (2008, December). An effective news recommendation in social media based on users' preference. In International Workshop on Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. (vol. 1, pp. 627-631). IEEE.

Xylogiannopoulos, K. F., Karampelas, P., & Alhajj, R. (2016). Repeated patterns detection in big data using classification and parallelism on LERP Reduced Suffix Arrays. *Applied Intelligence*, *45*(3), 567-597.

Zarrinkalam, F., Fani, H., Bagheri, E., Kahani, M., & Du, W. (2015, December). Semantics-enabled user interest detection from twitter. In *International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (vol. 1, pp. 469-476). IEEE.