A Novel Tweet Recommendation Framework for Twitter

Kamaljit Kaur, Kanwalvir Singh Dhindsa

Abstract*: In order to keep them updated users follow various Twitter accounts to get the latest information. As their social network increases it becomes challenging for them to find the relevant content from the massive collection of information. A Twitter user needs to scan a lot of less relevant posts to find the interesting tweets. Important updates may get lost if user is not able to read all the messages. So there is need that the most relevant updates are shown to the user first. Traditionally, the most retweeted tweets are considered popular and are brought forward. In order to improve the attractiveness of the incoming tweets we propose a personalized tweet ranking method based on the trending topics in the user network. A hashtag ranking model is developed to map the tweets into a ranked list of hashtags. The tweets corresponding to those hashtags are then ranked based on the linear weighted model that considers features related to tweet, author of tweet and the user. Finally, conducting a pilot user study we analyze the effectiveness of the proposed framework.

Keywords: Favorites, Hashtags, Retweets, Tweets, Twitter Timeline

I. INTRODUCTION

Twitter is no longer a social network but has grown out into a real-time personalized news service. The purpose of the Twitter Timeline is to help its user stay informed about what is going around the world. Earlier the tweets from every followed person were shown in chronological order with the most recent tweets on the top of timeline on opening the application. As the user base and its follow list grew, the chronological feeds limitation became clear. Various changes were introduced in an attempt to improve the twitter experience and make the timeline more engaging. In 2014 Twitter included recommended tweets, accounts and topics on the user timeline and for the first time exposed them to the content from people they did not follow. In 2015 Twitter introduced the feature 'while you were away' that gives the recap of some of the top tweets that the user might have missed from the user accounts they follow. 'In case you missed it' has replaced 'while you were away' with the motivation to users who log into network less often. In 2016, Twitter introduced Algorithmic Timeline that means tweets would no longer appear in the order they were posted but program will decide which tweets will appear on the timeline when you open the application. It ensures that more tweets will be shown from the people with whom more interactions are done. The main reason behind the introduction of the algorithmically selected and ranked tweets was drawing in new users and more importantly making the existing users more active. When the user logs in the system, the algorithmic timeline algorithm studies all the tweets from the accounts been followed by the user and gives each tweet a relevance score based on different factors relating to tweet, author of the tweet and the user whose timeline is ranked. Out of all the tweets from the user timeline the tweets are such ranked that it will be of the most interest to the user. The set of highest scoring tweets are shown at the top of timeline with the remainder shown directly below. Another important feature in Twitter is hashtags that are keyword or a phrase used to describe a topic or a theme. Twitter users put hashtags in their tweets to categorize them in a way that makes it easy for other users to find and follow tweets about a specific topic. Trending Topics are hashtags that are being widely discussed and are very popular. To the best of the knowledge, there is no existing work on ranking tweets based on these trending topics. Twitter provides its user with a list of top ten trending topics but these trends are general topics which are popular based on the user location and are not personalized. In this paper, we extend our existing model that extracts the trending hashtags from the user network [25]. The content of the user and the user network which reflects the user interest is considered in order to generate what are the popular topics of discussion and is further used to organize the user timeline. We propose a hashtag ranking model to map the tweets into a ranked list of hashtags. The tweets corresponding to those hashtags are then ranked based on the linear weighted model that considers features related to tweet, author of tweet and the user.

II. RELATED WORK

Due to information overload, one of the problems faced by Twitter users is to scan large number of less relevant posts in order to find the interesting content. Effective user recommendation approach through which the user receives the information from the right users and is not lost in the uninteresting content is one solution.

Revised Manuscript Received on August 09, 2019.

***** Correspondence Author

Kamaljit Kaur* Department of Computer Science, Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India. Email: kamaljit_cse@sggswu.org

Dr. Kanwalvir Singh Dhindsa. Department of CSE & IT, Baba Banda Singh Bahadar Engineering College, Fatehgarh Sahib, Punjab, India . Email: kdhindsa@gmail.com

But this leads to a tradeoff as a user with large number of friends can experience information overload due to high volume of messages from large number of connections which makes it hard for the user to find the useful information whereas users with small number of social connections might miss important messages of their interest which do not reach them. In order to improve the attractiveness of incoming tweets, many personalized tweet ranking methods have been developed. Most of the work in this area is leveraged on the use of the retweet behavior in order to show the popular tweets first. The process of propagating the tweet from another user is called retweeting which is an indicator that user who has decided to share the tweet in his network considers that its content contains useful information. Using the features related to the tweet, tweet content, author of the tweet and the user the incoming tweets are ranked to bring those tweets forward which are more likely to be retweeted. The author has developed a tweet ranking model based on the relevance of the tweet according to the user interest profile. The user profile consists of interest concept vector that represents the topics that are of interest to the user and the user affinity vector that considers the interaction between the users. Lu, Lam and Zhang [24] in their work ranked tweets based on user profile constructed from their tweets and the relevance was measured between tweet and user interest. Wikipedia concepts are used to represent the user interest profile. Explicit Semantic Analysis algorithm is used to extract the concepts from the tweets and the user profile is expanded using random walk on Wikipedia concept. Uysal and Croft [22] rank the possible users of a tweet based on their likelihood of retweeting that tweet. Using Decision Tree Based Classifier first the tweet is classified as retweetable or not for the user based on author, tweet, content and user features. Amiri and Shobi [23] use link prediction strategy to recommend similar tweets using twitter information like tweets, likes, retweet information and tweets semantics that explicitly reflect user interest. Du et al [25] emphasizes on capturing user interest over time as key aspect for personalized tweet recommendation. Their work provides two recommendation models named as Session-based Temporal Graph model and Singular Value Decomposition model to learn user preference and recommend personalized tweet. Further the efficiency of these techniques are improved through the parallel implementation using Hadoop Map-Reduce framework.

III. PROPOSED FRAMEWORK

Twitter users are interested in finding currently popular information on the network. Twitter displays a list of immediately popular keywords as hashtags on the user's homepage to help users discover the emerging topics in Twitter, and these keywords are referred to as trends. This work is extension to our previously developed framework towards personalizing this trend list which is important when the user wants to see what is trending in his network of users rather than the generic location based topics as recommended by Twitter. The proposed framework emphasizes on the user's social network structure to recommend the tweets that are currently been popular. Content-based user profiling approach is employed as the users tweet about things that interest them. Tweets of their friends can further provide insight into the user's interest. The followees are not considered as source of profiling information as they may or may be not of the interest to the user. By providing the registered twitter username as input; the user and the friends of the user are found in the real time from the Twitter stream. For the given username the corresponding userid is retrieved and is in turn used to find all the users been followed. For each of the user followed the list of userid is retrieved and stored for further use. After the userid of the user and all the people followed by the user are found then the Tweet corpus is generated that includes the tweets posted by the user and his friends. After the tweet corpus is generated the remaining steps are shown in Figure 1. For each of the user tweet the tokens are generated and then from the list of generated tokens the corresponding hashtag entities are found and these tags are saved in the database.

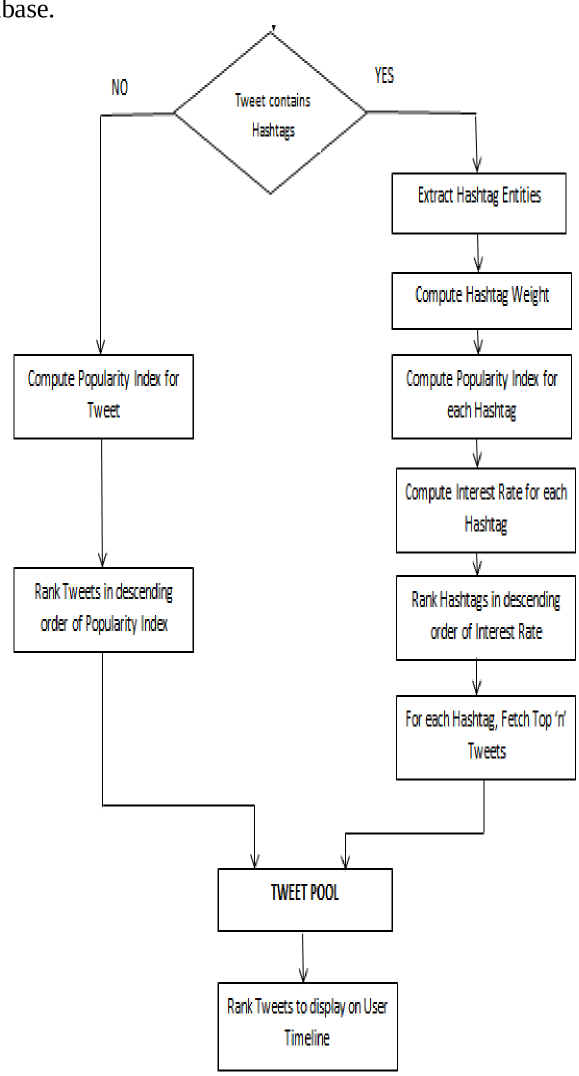

Figure 1: Ranking Twitter Timeline Using TrendNet Tweet Rank Algorithm

This process is repeated for all the users and all the hash tags retrieved from the user tweets are saved into the database. Simple frequency count is used as a hash tag weighting function so that the profile vector could be transformed into frequency count of various hash tag entities been used in the tweets of target user and its friends. For each user, the saved hashtags are retrieved from the database and the frequency count is calculated. If the frequency count of the hashtag is greater than two then it is set to maximum frequency which is assumed as two. This is done in order to avoid users to promote their hashtags as trends by repeating the same hash tag entities multiple times in their tweets. This process is repeated for all the users in the network. At the end the sum total of frequency of each occurred hashtag is computed for all the users taken into consideration.

Every user in social media is unique and it is important for the social media to identify the particular user interests. In this step the computation of popularity count of hashtag is performed using two popular Twitter indicators of Favorite and Retweet. For each tweet (which is known as Status in twitter terminology) its corresponding Status Text, Status ID, User ID, User Name, Favorite Count, Retweet Count and Created At timestamp is saved in the Tweet database.

From this database for each tweet its corresponding indicators of Favorite and Retweet is fetched and used to find the popularity count for each generated hashtag entities. The Popularity count is the sum total of the Favorite Count and Retweet Count. Interest Rate for each hashtag is computed by multiplying the frequency count and popularity count. Then this list of hashtags in arranged in the descending order of their Interest Rate Count. For each hashtag, the top-n tweets are fetched. Similarly for the tweets that do not contain hashtags, its popularity index is computed using the Favorite and the Retweet indicators. The tweets are arranged in the descending order of their popularity index. These collected tweets are then ranked based on the tweet ranking algorithm.

Algorithm: Hashtag Ranking [25]

1. Computing Hashtag Count and Hashtag Limit:

For each user j:

For each hashtag h_s generated by that user:

h_{s_cnt} = get_count (h_s)

If $(h_{s\_cnt} > H_{limit})$ $then\ set$ $(h_{s\_cnt} = H_{limit})$

$hc_s = h_{s\_cnt}$

2. Computing Hashtag Weight:

For each hashtag h_s over all the users belonging to User Network compute hashtag weight as:

$$HW_s = \sum_{j=1}^{n+1} (hc_{sj})$$

3. Computing Hashtag Popularity:

For each hashtag h_s:

For each tweet t_y related to hashtag h_s:

$$HP_s = \sum_{s=1}^{z} TR_{sy} + TFC_{sy}$$

where, TR_{sy} is Retweet Count and TFC_{sy} is Favorite Count of the tweet in which that hashtag appears

4. Computing Hashtag Interest Rate:

$$HIR_s = \sum_{s=1}^{z} \left(\prod_{s} HP_s \ \ HW_s \right)$$

Tweet Ranking Algorithm

To compute the ranking score of tweets; the linear weighted recommendation model is developed that considers the following six features. The first feature considers the social popularity of the tweet author in the microblogging system. The ratio indicates that more the number of users following the tweet author, more popular the user is.

$$w_p(u_{TA}) = \frac{|followers(u_{TA})|}{|followees(u_{TA})|}$$

The second feature considers whether the author of the tweet in which the hashtag is occurring is also following the target user u_T. More weightage is given if there is a bidirectional relation between both the users as it has higher possibility that the target user will be interested in the content.

$$w_f(u_T, u_{TA}) = \begin{cases} 1, & if\ relation\ between\ u_{TA}\ and\ u_T\ is\ bidirectional \\ 0, & if\ relation\ between\ u_{TA}\ and\ u_T\ is\ unidirectional \end{cases}$$

The third feature calculates the number of friends shared between the target user and the tweet author.

$$w_c(u_T, u_{TA}) = \left| followees(u_{TA}) \bigcap followees(u_T) \right|$$

More common friends a candidate user shares with the target user; it is more likely that they share similar tastes.

The fourth feature considers the user profile keywords. More weightage is given if either the hashtag matches with any of the profile keywords or there is any match between the profile keywords of the candidate user with that of the target user.

$$U_{Profile}(u) = \{w_1, w_2, \ldots, w_n\}$$

$$w_{hp}(u_{profile}(u_T, u_{TA})) = \begin{cases} 1, & if\ match \\ 0, & if\ no\ match \end{cases}$$

The fifth feature considers whether the tweet in which the hashtag is occurring is mentioning the target user. If the tweet is mentioning the user then higher weight is assigned to its corresponding hashtag.

$$w_m(u_T, u_{TA}) = \begin{cases} 1, & if\ tweet\ of\ u_{TA}\ mentions\ u_T \\ 0, & if\ tweet\ of\ u_{TA}\ donot\ mention\ u_T \end{cases}$$

The sixth and the last feature consider the type of tweet. Less weightage is given to hashtag for the tweet which is a reply tweet than the original message.

$$w_t(t_{TA}) = \begin{cases} 0, & if\ tweet\ is\ original\ message \\ -1, & if\ tweet\ is\ a\ reply\ tweet \end{cases}$$

The ranking model is the linear weighted model of all the six features discussed above. The ranking model can be described by the equation below:

$$W_s = w_p(u_{TA}) + w_f(u_T, u_{TA}) + w_c(u_T, u_{TA}) + w_{hp}(u_{profile}(u_T, u_{TA})) + w_m(u_T, u_{TA}) + w_t(t_{TA})$$

Based on the tweet weights the order of the tweets is calculated and recommended to the user.

IV. RESULTS AND DISCUSSIONS

Trends in Twitter are used to represent the currently popular topics in discussion. To the best of our knowledge, using personalized trends to rank user timeline is the first step in this direction. The proposed work can be divided into two parts where at first we retrieve the personalized hashtags from the twitter corpus [25] and then tweets are ranked. Based on the popularity index of the tweets and the hashtag the tweet pool is created.

The ranking score of the tweets is calculated using the linear weighted model which is then used to re-rank the tweets to be shown on the user timeline. The Table1 shows the tweets generated by the proposed framework for the author's twitter user account. The most favorite and retweeted tweets are ranked at higher order in the user timeline along with other discussed user and tweet features.

Table 1: Tweet Timeline Using TrendNet Tweet Rank Algorithm

User	Tweets	Favorite Count	Retweet Count
Donald J. Trump	France just put a digital tax on our great American technology companies. If anybody taxes them, it should be their…	153795	38410
Donald J. Trump	Wow! Big VICTORY on the Wall. The United States Supreme Court overturns lower court injunction, allows Southern Bor…	140774	37299
Narendra Modi	India is very proud of @HimaDas8's phenomenal achievements over the last few days. Everyone is absolutely delighted…	124522	15562
Narendra Modi	Indian at heart, Indian in spirit! What would make every Indian overjoyed is the fact that #Chandrayaan2 is a ful…	118894	18801
Donald J. Trump	Apple will not be given Tariff waiver, or relief, for Mac Pro parts that are made in China. Make them in the USA, no Tariffs!	93434	20898
Donald J. Trump	The WTO is BROKEN when the world's RICHEST countries claim to be developing countries to avoid WTO rules and get sp…	83949	22343
Narendra Modi	Special moments that will be etched in the annals of our glorious history! The launch of #Chandrayaan2 illustrate…	86014	13661
Donald J. Trump	@FoxNews is at it again. So different from what they used to be during the 2016 Primaries, & before - Proud Warrio…	75371	15289
Capt.Amari nder Singh	Have promoted Vir Chakra awardee Satpal Singh of the #KargilWar from senior constable to ASI. He deserves special t…	10119	2002
Akshay Kumar	Teamwork makes the dream work! Watch the team work towards their dream, #MissionMangal in #DilMeinMarsHai!Song out…	8882	972
Congress	RT @RahulGandhi: From its first day, the Cong-JDS alliance in Karnataka was a target for vested interests, both within & outside, who saw t…	0	8873
Capt.Amari nder Singh	4 years ago on this very day, India lost its #MissileMan. My heartfelt tribute to the former President…	5405	638
President of India	Watch LIVE as President Kovind pays homage at the Chinar Corps War Memorial in Srinagar on Kargil Vijay Diwas. Due…	4814	741
NASA	Our #Artemis program will return astronauts to the Moon by 2024.	4245	1025
DILJIT DOSANJH	#ArjunPatiala IN CINEMAS TODAY @kritisanon @varunsharma90	2514	123
Manish Sisodia	Remembering the great #APJAbdulKalam on his 4th death anniversary. The president who wanted to be remembered as a t…	1048	228
Markandey Katju	Dr #APJAbdulKalam's life is a role model for all. My tributes to the great son of India on his death anniversary.	186	31
Sukhbir Singh Badal	I join the nation in fondly remembering India's #MissileMan and the People's President, Dr APJ Abdul Kalam. His sim…	75	30
Cloudera	Cloudera Now is the world's premier online #bigdata event designed for architects, #analysts, #data scientists,…	12	7
Cloudera	Are you ready for THE #BigData community event? Register soon for #ClouderaNow!	8	7

V. CONCLUSION

A novel approach towards tweet recommendation is presented in the paper that first detects the popular keywords and the hashtags that are popular in the user network. The aim is to find which are the important topics of discussion among the user friends rather than the general trends as provided by the Twitter. Using these topics a tweet pool is generated which are then ranked using linear weighted ranking features. Based on the pilot user study on the author account it shows the effectiveness of the proposed algorithm with more personalized tweets in user timeline. Further the research can be extended by taking into consideration additional twitter features into account.

REFERENCES

1. Lu, C., Lam, W. and Zhang, Y., 2012, July. Twitter user modeling and tweets recommendation based on wikipedia concept graph. In Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence (pp. 33-38).
2. Pla Karidi, Danae & Stavrakas, Yannis & Vassiliou, Yannis. 2016. A Personalized Tweet Recommendation Approach Based on Concept Graphs. 253-260. 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0056.
3. Sharma, Chhavi & Bedi, Punam. (2018). Community based Hashtag Recommender System (CHRS) for twitter. Journal of Intelligent & Fuzzy Systems.
4. Ding, Z., Zhang, Q. and Huang, X., 2012. Automatic hashtag recommendation for microblogs using topic-specific translation model. *Proceedings of COLING 2012: Posters*, pp.265-274.

5. Ran, C., Shen, W. and Wang, J., 2018, April. An Attention Factor Graph Model for Tweet Entity Linking. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*(pp. 1135-1144). International World Wide Web Conferences Steering Committee.
6. Ma, J., Feng, C., Shi, G., Shi, X. and Huang, H., 2018. Temporal enhanced sentence-level attention model for hashtag recommendation. *CAAI Transactions on Intelligence Technology*, *3*(2), pp.95-100.
7. Gong, Y., Zhang, Q. and Huang, X., 2018. Hashtag recommendation for multimodal microblog posts. *Neurocomputing*, *272*, pp.170-177.
8. Kou, F.F., Du, J.P., Yang, C.X., Shi, Y.S., Cui, W.Q., Liang, M.Y. and Geng, Y., 2018. Hashtag Recommendation Based on Multi-Features of Microblogs. *Journal of Computer Science and Technology*, *33*(4), pp.711-726.
9. Ben-Lhachemi, N. and Nfaoui, E.H., 2018. Using Tweets Embeddings For Hashtag Recommendation in Twitter.*Procedia Computer Science*, *127*, pp.7-15.
10. Ben-Lhachemi, N. and Nfaoui, E.H., 2018, April. Hashtag Recommendation Using Word Sequences' Embeddings. In*International Conference on Big Data, Cloud and Applications*(pp. 131-143). Springer, Cham.
11. Otsuka, E., Wallace, S.A. and Chiu, D., 2016. A hashtag recommendation system for twitter data streams.*Computational social networks*, *3*(1), p.3.
12. She, J. and Chen, L., 2014, April. Tomoha: Topic model-based hashtag recommendation on twitter. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 371-372). ACM.
13. Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B. and Van de Walle, R., 2013, May. Using topic models for twitter hashtag recommendation. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 593-596). ACM.
14. Ran, C., Shen, W. and Wang, J., 2018, April. An Attention Factor Graph Model for Tweet Entity Linking. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*(pp. 1135-1144). International World Wide Web Conferences Steering Committee.
15. Ma, J., Feng, C., Shi, G., Shi, X. and Huang, H., 2018. Temporal enhanced sentence-level attention model for hashtag recommendation. *CAAI Transactions on Intelligence Technology*, *3*(2), pp.95-100.
16. Gong, Y., Zhang, Q. and Huang, X., 2018. Hashtag recommendation for multimodal microblog posts.*Neurocomputing*, *272*, pp.170-177.
17. Kou, F.F., Du, J.P., Yang, C.X., Shi, Y.S., Cui, W.Q., Liang, M.Y. and Geng, Y., 2018. Hashtag Recommendation Based on Multi-Features of Microblogs. *Journal of Computer Science and Technology*, *33*(4), pp.711-726.
18. Terán, L., Mensah, A.O. and Estorelli, A., 2018. A literature review for recommender systems techniques used in microblogs. *Expert Systems with Applications*, *103*, pp.63-73.
19. Orellana-Rodriguez, C. and Keane, M.T., 2018. Attention to news and its dissemination on Twitter: A survey. *Computer Science Review*, *29*, pp.74-94.
20. Zhang, A., Zheng, M. and Pang, B., 2018. Structural diversity effect on hashtag adoption in Twitter. *Physica A: Statistical Mechanics and its Applications*, *493*, pp.267-275.
21. Lee, H., Abdar, M. and Yen, N.Y., 2018. Event-based trend factor analysis based on hashtag correlation and temporal information mining. *Applied Soft Computing*
22. Uysal, I. and Croft, W.B., 2011, October. User oriented tweet ranking: a filtering approach to microblogs. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 2261-2264). ACM.
23. Amiri, M.Z. and Shobi, A., 2017. A link prediction strategy for personalized tweet recommendation through doc2vec approach. *Research in economics and management*, *2*(4), pp.63-76.
24. Du, G., Sun, J., Huang, X. and Yu, J. Mining User Interests for Personalized Tweet Recommendation on Map-Reduce Framework. In Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 1, pages 201-208
25. Kaur K, Dhindsa K.S, 2019 TrendNet Algorithm: Towards Personalization of Twitter Trends. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, pages 1555-1560
26. Kaur K, Dhindsa K.S, 2019, Study of Existing Design Architectures and Frameworks in Twitter. In Proceedings of the IEEE International Conference on Automation, Computational and Technology Management (ICACTM)

AUTHORS PROFILE

Kamaljit Kaur is an Assistant Professor in Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, Fatehgarh Sahib. She is pursuing her Ph.D. in Computer Science from Punjab Technical University, Kapurthala, Punjab, and MTech and BTech in Computer Science from Punjab Technical University, Kapurthala, Punjab. Her area of interests includes Databases, Data Mining and Machine Learning.

Dr. Kanwalvir Singh Dhindsa is currently working as Professor (CSE) and Incharge, ERP at Baba Banda Singh Bahadur Engineering. College, Fatehgarh Sahib (Punjab), India. With an overall experience of 17 years in the academia & research related with Computer Science and IT, He is an avid researcher having guided dissertations of many M.Tech PG students & is also currently guiding 7 PhD scholars. He is also on the reviewer panel of many esteemed referred and peer-reviewed journals. His current research interests include Cloud Computing, Web Engineering, Big Data, Internet of Things, Database & Security, and Mobile Computing. He is also Life Member-CSI, Fellow IETE, Member IEI, and Life Member-ISCA.