BioGecko Vol 12 Issue 03 2023

ISSN NO: 2230-5807

WEB BEHAVIOURAL PATTERN MINING IN SOCIAL MEDIA NETWORKS USING MULTI-OBJECTIVE OPTIMISATION

Author 1: Biju Balakrishnan
Designation: Research Scholar
Department: Information Techno;ogy
Affiliation: St Peter's Institute of Higher Education and Research, Avadi, Chennai

Author 2: Dr.B.Shanthini
Designation: Professor & Head
Department: Information Technology
Affiliation: St Peter's Institute of Higher Education and Research, Avadi, Chennai

Abstract
The widespread availability of multimedia devices in recent years has greatly increased people ability to enjoy a wide range of forms of entertainment. The internet multimedia platforms not only give people a place to disseminate knowledge, but they also give them a way to relax and refresh their bodies and brains.In this paper, we focus on generating user profiling with clustered web user data, by formulating a multi-objective functions. The multi-objective functions are formulated based on the relative user profile attributes strength. The attribute strength of user varies based on their preference, dislikes, behavioural traits, social linkages etc.

Keywords
Behavioural Pattern Mining, Social Networks, Multi-Objective Optimisation, Web Data

1. Introduction

Due to its emphasis on human interaction, social media has amassed a massive readership [1] and has become an integral part of many people daily routines [2]. Even though the vast bulk of social media content is created and shared by normal users, many still look to their feeds as a dependable source of information on current events, public opinion, market assessments, and significant personalities [3].
Data mining techniques have been created to help achieve more with fewer resources and to mimic human intelligence [4]. Algorithms employed in data mining must be able to self-educate in order to process the vast volumes of information they amass and store. Data exploration is also used in data mining techniques to help combine disparate datasets, identify meaningful connections, and derive useful insights [5]. Information exploration is a tool that can help with this. Statistics, classification, clustering, regression, and association are all subfields of data mining technology [6].
There is no other network that comes close to its popularity as the most widely used global real-time information network. Users consistently provide current information and offer their opinions on a wide variety of issues [7]. There are a number of ways for Sina Weibo users to show their approval of other user posts, including the like, retweet, hashtag, and mention buttons [8].
Due to the heterogeneous nature of social media data, many approaches have been tried to investigate and make sense of the problem. In order to categorize user feelings, researchers have deployed various variants of machine learning techniques [9]. The emotions of the users are another part of the experience that has been studied. This work complements previous studies that have focused on the same topics—namely, identifying user characteristics and mining information models [10]-[19].
We used multi-objective optimisation into study to better understand the sentiments expressed in the posts under investigation. In this research, we build multi-objective optimisation(MOO) from the profiles of internet users based on aggregated data. To create a MOO, we first rank the significance of different user profile features. User preferences, personality traits, social connections, and other factors all contribute to a wide range of attribute strengths.Depending on the nature of the crisis at

ISSN NO: 2230-5807

hand, our research suggests that people may choose to communicate their emotions in a variety of different ways. The study results add to the growing body of knowledge about the range of emotions shared over the internet.

2. Related works

Anandhi and Ahmed [20] proposed some mechanisms to use when mining web logs.Identifying user navigational patterns in order to better understand the users' goals, promoting a more individualized experience, and developing design thinking were all aims that needed to be accomplished. They made use of density-based clustering in order to achieve their goal of detecting patterns in the sessions of the consumers. Web sessions provide data streams, which call for mining to take place more quickly.

The Map Reduce methodology serves as the foundation for the streaming processing strategy proposed by Chen et al. [21]. Scalability was maintained, and the load-balancing capabilities of the cloud were leveraged to their maximum potential. Despite this, additional work has to be done in order to make it memory-efficient and acceptable for testing in heterogeneous cloud environments.When using one of the various frameworks, such as Apache Storm, it is feasible to process an endless number of streams of data.

Iqbal and Soomro [22] are the ones to consult for information regarding streaming processing when it comes to big data analytics. There are many different contexts in which user behavior analysis can be useful; the school setting is one of them.

Kassak et al. [23] focused their attention on exit intent prediction in order to gain a better understanding of how students interact with a web-based learning platform. Attention order They created a classifier by employing a combination of stochastic gradient descent and polynomial regression in its development. The classifier was successful in revealing the dynamic run-time behavior of the users. They failed to recognize the underlying interdependence that was present in the students' actions.

Liraki et al. [24], employ the concept of weighted association rule mining (ARM) to predict the behavior of users in terms of navigational patterns. They came up with a hybrid technique for efficiently mining latent patterns by combining ARM with fuzzy C means (FCM). The method has the drawback of producing a larger number of association rules, which, in order to improve performance, need to be cut down. In most cases, recommendations are produced by utilizing various data mining techniques.

Tyagi and Bharadwaj [25]developed a user intention model for improving the user ideas. In order for collaborative filtering (CF) to be effective, it must be able to shed light on the considerations that went into previous userdecisionsfor collaborative filtering (CF) to be effective, it must be able to shed light on the considerations that went into previous users' decisions. They might be able to improve their strategy by using fuzzy logic.

For the purpose of web usage mining, Thiyagarajan et al. [26] looked into the application of weighted K-Means clustering. It has the potential to improve the functionality of the K-Means algorithm. However, in order for them to conduct an accurate analysis of user activity, their technology will need to be upgraded so that it is capable of accommodating overlapping clusters.

Moniz et al. [27] gave an approach for finding extensively shared articles in their study, which may be found here. Their approach, which is focused on maximizing efficiency, has the potential to increase output. The limited size of the data samples they have access to limits the scope of the work they have done. There may be a wide variety of preferences among Internet users. The process of modeling them can be highly beneficial for gaining relevant business insight.

Rafailidis and Nanopoulos[28]placed a significant amount of focus on a concept known as user preference dynamics, which investigates the rate at which consumer preferences shift. In recent years, the relevance of a user preferences has expanded as a result of variables such as increased exposure to novel products, learning from one errors, the bias toward more popular products, and the increase in the number of options available to consumers. Their approach has a number of flaws, one of which is that it does not take into account the passage of time. Because there can be a large number of players participating in an online game at the same time, it is essential to observe and evaluate their behavior.

ISSN NO: 2230-5807

Sapienza et al. [30] looked at non-negative tensor factorization by applying it to the scenario of a multiplayer online game. In this kind of game, the activity patterns of the players are important since they provide the required information for decision-making. They considered the passage of time as a factor while formulating a solution to the problem. However, further development was necessary in order to use the previous records of participants in games.
Lee et al. [31] conducted research on a wide variety of procedures for mining common patterns. We also looked into more modern approaches, including the damped window model, the sliding window model for streaming processing, and the slanted time window model. According to them, the existing algorithms need to be improved in some way, whether it be by sampling that is more condensed or through some other technique. This will allow for an increase in both their throughput and their efficiency in terms of resource consumption.
Shin et al. [32] made use of the CP-tree data structure in order to simplify the mining of often-recurring sets. This was done to facilitate the process. A concise explanation of a set of items and the resources that support them is provided at each tree node, which can be used to learn more about the subject matter. The solution that was based on CP trees performed far better than its predecessors.
It is the goal of this study to use a mathematical model to optimize for multiple criteria simultaneously. Multi-objective functions are solved by constructing objective functions and using optimization algorithms to find a solution that meets all of the objectives. This is done in light of the specific needs of the ideas being made. When evaluating a recommendation system, it is important to examine not only the precision of the system but also other metrics, such as the breadth of the options it provides. The MOO problem is a model that more closely reflects real-world issues in science and industry.

3. Proposed Method

In this section, we study the web-based social media recommendation system using multi-objective optimisation. The illustration of which is given in Figure 1.

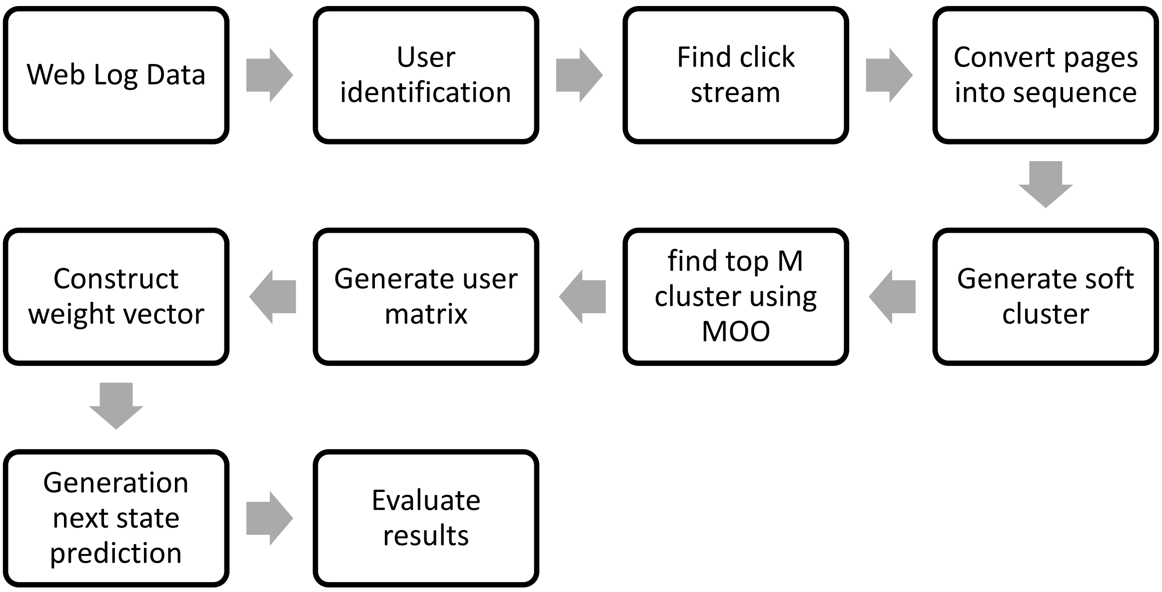

Figure 1: Proposed Framework

The issue of user feature arises when one goal begins to take precedence over another. The formula describes the meaning of the PF^* value.

$PF^*=\{F(x^*)=(f_1(x^*),f_2(x^*),\ldots f_K(x^*))^T || x^* \in PS^* ||\}$

Both MOO evolutionary algorithmsis used to operate the entire user web recommendation system. initial emphasis on developing potential answers. It focuson improving the similarity functions. increases the amount of variance supplied by the recommendation algorithm.

Traditional collaborative filtering approaches undergo variety of difficulties, which includes lack of diversity in the recommendations. In this research, we introduce the MOO algorithm, which increases idea drift flexibility and diversity of recommendations that merges the process of prediction using kernel density estimation with the MOO.

This leads in an improved concept drift adaptability and the scoring prediction algorithm initially pairs people together based on how similarly they were scored. This helps the system be more accurate in its predictions. The next phase is for the existing evaluation algorithm to produce predictions for all unrated occurrences using rating matrix. These projections will be used to guide future reviews.

The first thing that the algorithm does is develop a model based on the user preferences and habits. The kernel density is incorporated into the process of constructing the user model. This shows that the distribution score of the user across the totality of the item space provides a more accurate assessment of the degree to which individuals share interests. The system examines the user earlier ratings of products in order to develop predictions about how they would rank new ones. It then takes use of those predictions in order to establish which row vector in the user rating matrix best shows the user interest in a certain product.

The rating matrix for the user-items are shown in Table 1, where *R* denotes a rating record and a value of 0 indicates that the user did not give a rating to the item in question.

Table 1:User-Item Rating Matrix

	$Item_1$	**$Item_2$**	**…**	**$Item_n$**
$User_1$	$R_{1,1}$	$R_{1,2}$	…	$R_{1,n}$
$User_2$	$R_{2,1}$	$R_{2,2}$	…	$R_{2,n}$
…	…	…	…	…
$User_n$	$R_{n,1}$	$R_{n,2}$	…	$R_{n,n}$

In order to establish a neighborhood set using the similarity measure, the interest model is used to compute interest similarity between users in order to produce a neighbor set, and the distribution is derived using the kernel density. Before beginning any form of calculation, it is vital to limit or remove the potential influence of applying different scales to determine the degree of similarity. The Pearson correlation coefficient is applied in this study in order to evaluate the degree of congruence between two different collections of data.

The system first calculates the overall average rating before moving on to calculating the Pearson correlation coefficient. After that, it estimates the individual rating standard deviation from the mean. By leveraging the dissimilarity between users in order to identify the degree to which they are similar, the corresponding recommendation algorithm is able to achieve more trustworthy prediction results from the mean. By leveraging the dissimilarity between users in order to identify the degree to which they are similar, the corresponding recommendation algorithm is able to achieve more reliable prediction results. This is accomplished by decreasing the noise provided by users who have varied scoring requirements.

The user is given with material that has been proved to be well-liked by the user contemporaries through the deployment of a certain recommendation approach. The algorithm achieves adaptive weight reduction with the optimal integration of the static parameters of the closest neighbor with the user activity. It then performs weight reductions of varying widths based on the similarity items before finalizing the solutionsselection. This approach is done until all feasible solutions have been examined.

In order for the strategy to have a rapid rate of convergence, the objective function needs to be adjusted in accordance with the Pareto dominance principle. When combined with a number of different strategies for the preservation of diversity.

4. Results and Discussions

In this study, the algorithm performance is measured via offline tests. The data set of user activities is split in two parts: the training/test set of user actions. A user interest model is trained on the training set, and then predictions about the user future actions are made using the test set.

The experiment dataset was selected from Netflix and some of the most well-known films ever made are included in this dataset. To date, there have been around 480,189 signups, and the database currently has 17,770 reviews of various films. Training, testing and validation sets are all used in the proposed study.

Therefore, it is possible that a more accurate description of the algorithm performance can be obtained by combining the prediction accuracy, computational cost and execution rate.

In this study, we use two different collections of experimental data to investigate the performance of recommendation algorithms that make use of multi-objective optimization methodologies. All of these findings are from the same experiment. The purpose of the preliminary tests is to zero in on the optimal strategy for solving the issue. The second set of trials compares the proposed algorithm to the MOO algorithms in terms of accuracy and variety.

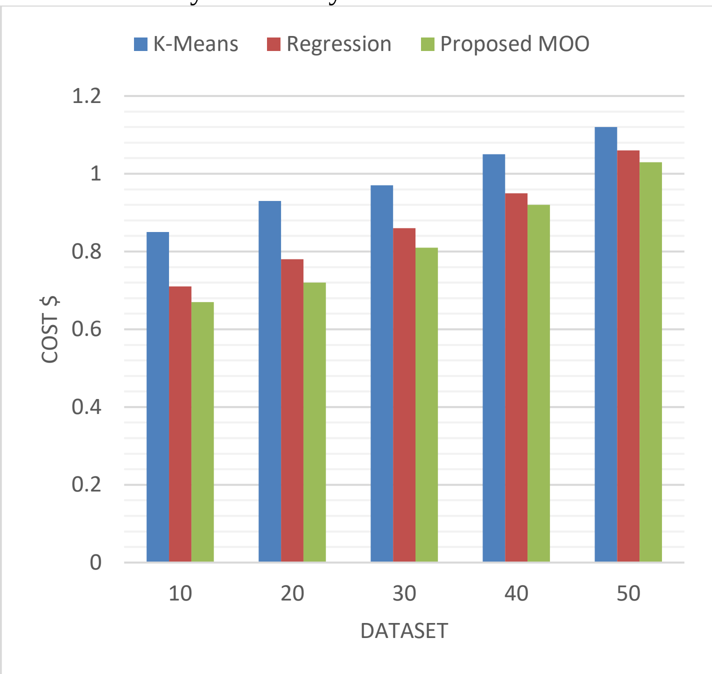

Figure 2: Communication Cost

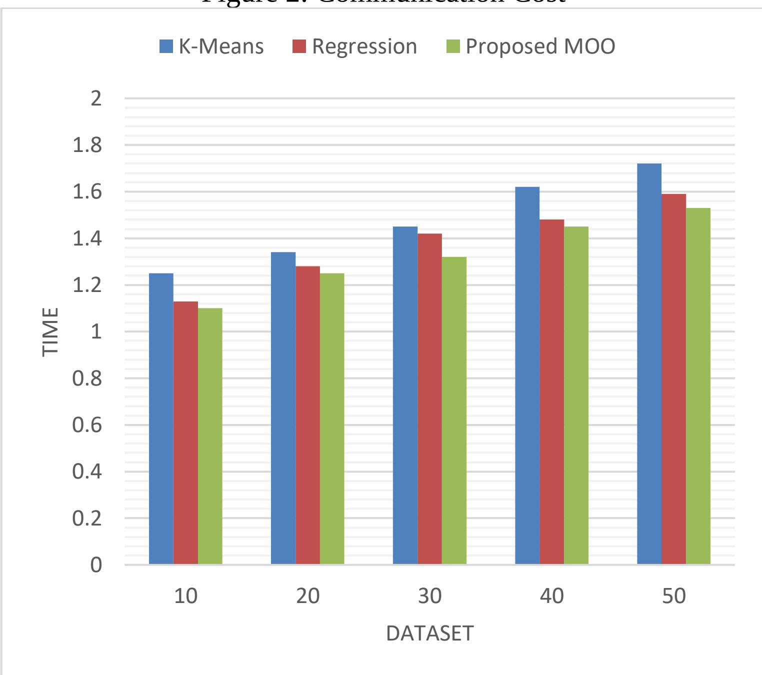

Figure 3: Execution Time (ms)

The original proposed use of the classic MOO algorithm was for filtering information such as emails and news articles. The algorithm starts by identifying a set of users whose interests are similar to those of the target user; it then makes more intuitive recommendations of items presents the final recommendation. Similarity calculations help with all of these processes. The approach has been widely adopted by other recommendation platforms.

MOO framework has a quicker time for identifying online users in contrast to the existing methods, and the Scalable approach. This is owing to the fact that analysts are able to estimate a user personality feature and then extrapolate those preferences for products using data gathered from the user social media profiles.

Using the dataset and the proposed framework, we were able to determine the product that possessed the one-of-a-kind online social network. The amount of time required to authenticate a person identification over the internet is greatly decreased.

Research conducted within the MOO framework has demonstrated a 37% reduction in the amount of time needed to identify web users in comparison to the PCA and LDA feature extraction approach, an 18% reduction in accuracy required is found in testing when compared to the training method, and an 11% reduction in time required when compared to the validation method.

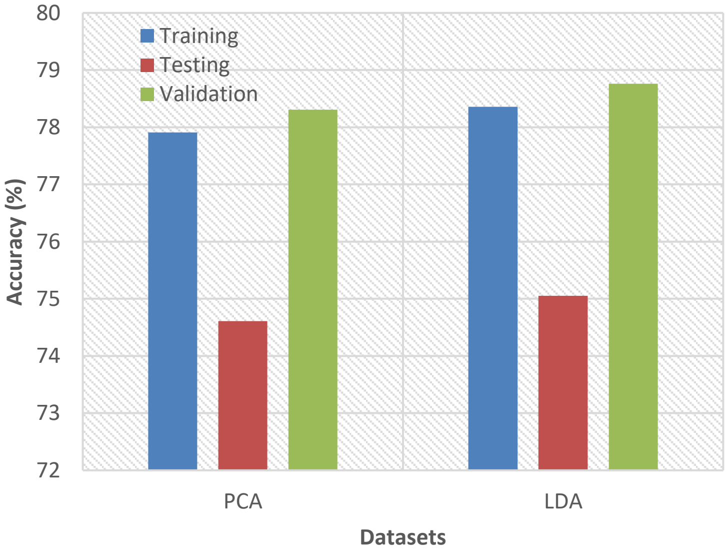

Figure 2: Accuracy

Figure 3: F-Measure

Figure 4: Precision

Figure 5: Recall

We established a methodology based on the testing of linear-temporal logic models with the purpose of applying it to the task of evaluating structured e-commerce blogs. Runs of prepared queries were conducted in attempt to detect patterns in the user behavior throughout the broad spectrum of activities that they participated in throughout the course of the session. The intended approach, on the other hand, did not study the supplemental behavior patterns and did not contribute to the automatic detection of these patterns.

In order to spot trends in regular Internet measurements collected locally by end users, MOO was developed as a tool. MOO recognized user service groups that had common Internet usage characteristics and categorized them together. Those real statistics were utilized by the free service, which gathers Internet analytics, in order to calculate MOO.

Mining ordinary behavioral patterns from the data provided by a wide array of sensors has been made easier with the development of a system that is scalable. The constructed framework was able to recognize the typical patterns of activity with the fine-grained temporal viewpoint that is typical of individuals. Both final consumers and those that supplied information-based services to final consumers, commonly known as providers, benefited from the trends.

Thus, MOO is a framework that was designed to analyse data from social media platforms in order to make conclusions about the personalities of users and, as a result, their preferences concerning particular items. On the other hand, during the e-recruitment processes that were taking place in the crowdsourcing market, there was no usage of a personality prediction engine to insure a good personality and organizational culture.

5. Conclusions

Comparisons are made between numerous ways for identifying users on the web by mining their patterns of behavior. According to the conclusions of the research, the approach of evaluating linear-temporal logic models is insufficient for conducting analyses and contributing to the automatic detection of extra behavioral patterns. The current iteration of the MOO technique does not result in an increase in clustering accuracy, as evidenced by the data of the survival research. In addition, there was no employment of a personality prediction engine in any of the e-recruitment processes that took place in the crowdsourcing market. This meant that a suitable personality-organizational culture fit was not attained. The numerous technologies for mining user behavioral patterns on the web each have their own set of advantages and limitations, which have been discovered through a variety of research based on previously established methodologies.

References

[1]. Gan, W., Lin, J. C. W., Fournier-Viger, P., Chao, H. C., & Yu, P. S. (2019). A survey of parallel sequential pattern mining. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *13*(3), 1-34.

[2]. Tarus, J. K., Niu, Z., &Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing*, *22*(8), 2449-2461.

[3]. Gan, W., Lin, J. C. W., Fournier-Viger, P., Chao, H. C., & Philip, S. Y. (2019). HUOPM: High-utility occupancy pattern mining. *IEEE transactions on cybernetics*, *50*(3), 1195-1208.

[4]. Awal, G. K., & Bharadwaj, K. K. (2018). Harnessing Collective Intelligence Through Pattern Mining in Social Computational Systems. In *Handbook of Research on Pattern Engineering System Development for Big Data Analytics* (pp. 91-110). IGI Global.

[5]. Radhakrishna, V., Aljawarneh, S. A., Kumar, P. V., & Janaki, V. (2018). A novel fuzzy similarity measure and prevalence estimation approach for similarity profiled temporal association pattern mining. *Future Generation Computer Systems*, *83*, 582-595.

[6]. Chakravarthy, D. G., &Kannimuthu, S. (2021). Mining interesting actionable patterns for web service composition. *Journal of Ambient Intelligence and Humanized Computing*, *12*(6), 6181-6187.

[7]. Zhang, C., Dai, Q., Du, Z., Gan, W., Weng, J., & Yu, P. S. (2021). Utility-driven Mining of Contiguous Sequences. *arXiv preprint arXiv:2111.00247*.

[8]. Chen, X., Xiao, R., Xin, D., Lin, X., & Lin, L. (2018, November). Constructing a novel spark-based distributed maximum frequent sequence pattern mining for IoT log. In *Proceedings of the 8th International Conference on Communication and Network Security* (pp. 112-116).

[9]. Rathish, C. R., & Rajaram, A. (2017). Robust early detection and filtering scheme tolocate vampire attack in wireless sensor networks. Journal of Computational andTheoretical Nanoscience, 14(6), 2937-2946.

[10]. Zheng, L., Xia, D., Zhao, X., Tan, L., Li, H., Chen, L., & Liu, W. (2018). Spatial–temporal travel pattern mining using massive taxi trajectory data. *Physica A: Statistical Mechanics and its Applications*, *501*, 24-41.

[11]. Balasubramaniam, T., Nayak, R., & Yuen, C. (2018, September). Nonnegative coupled matrix tensor factorization for smart city spatiotemporal pattern mining. In *International Conference on Machine Learning, Optimization, and Data Science* (pp. 520-532). Springer, Cham.

[12]. Adamov, A. Z. (2018, October). Mining term association rules from unstructured text in Azerbaijani language. In *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-4). IEEE.

[13]. Anwar, T., & Uma, V. (2020, October). Book recommendation for elearning using collaborative filtering and sequential pattern mining. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)* (pp. 1-6). IEEE.

[14]. Lin, J. C. W., Fournier-Viger, P., Tseng, V. S., & Philip, S. Y. (2021). IEEE Access Special Section Editorial: Utility-Pattern Mining: Theoretical Analytics and Applications. *IEEE Access*, *9*, 16604-16607.

[15]. Harikrishnan, G., & Rajaram, A. (2017). Improved throughput based recognitionconnection denies for aggressive node in wireless sensor network. Journal ofComputational and Theoretical Nanoscience, 14(12), 5748-5755.

[16]. Margret, M. K., Ponni, A., & Priyanka, A. (2021, May). Frequent Pattern Mining Using Db-Scan Algorithm. In *Journal of Physics: Conference Series* (Vol. 1916, No. 1, p. 012054). IOP Publishing.

[17]. Xiao, Z., Fu, X., Zhang, L., & Goh, R. S. M. (2019). Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey. *IEEE Transactions on Intelligent Transportation Systems*, *21*(5), 1796-1825.

[18]. Dias, V., Teixeira, C. H., Guedes, D., Meira, W., & Parthasarathy, S. (2019, June). Fractal: A general-purpose graph pattern mining system. In *Proceedings of the 2019 International Conference on Management of Data* (pp. 1357-1374).

[19]. Choi, H. J., & Park, C. H. (2019). Emerging topic detection in twitter stream based on high utility pattern mining. *Expert systems with applications*, *115*, 27-36.

[20]. Jevalaksshmi, S., Shankari, K. H., Mathivilasini, S., Jabeen, T. N., Maheswari, K., &Akila, D. (2021, September). Prediction and Analysis in High Utility Pattern Mining Using List Based HUPM Algorithm. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-5). IEEE.

[21]. Sun, G., Wang, W., Dai, J., & Van Gool, L. (2020, August). Mining cross-image semantics for weakly supervised semantic segmentation. In *European conference on computer vision* (pp. 347-365). Springer, Cham.

[22]. Premanand, R.P., Rajaram, A. Enhanced data accuracy based PATH discovery usingbacking route selection algorithm in MANET. Peer-to-Peer Netw. Appl. 13, 2089–2098(2020).

[23]. Anandhi, D., & Ahmed, M. I. (2014). An improved web log mining and online navigational pattern prediction. *Research Journal of Applied Sciences, Engineering and Technology*, *8*(12), 1472-1479.

[24]. Chen, C. C., Shuai, H. H., & Chen, M. S. (2017). Distributed and scalable sequential pattern mining through stream processing. *Knowledge and Information Systems*, *53*(2), 365-390.

[25]. Iqbal, M. H., & Soomro, T. R. (2015). Big data analysis: Apache storm perspective. *International journal of computer trends and technology*, *19*(1), 9-14.

[26]. Kassak, O., Kompan, M., &Bielikova, M. (2016). Student behavior in a web-based educational system: Exit intent prediction. *Engineering Applications of Artificial Intelligence*, *51*, 136-149.

[27]. Anitha, B., & Rajaram, A. (2014). Efficient Position Based Packet Forwarding ProtocolFor Wireless Sensor Networks. Journal Of Theoretical & Applied InformationTechnology, 69(2).

[28]. Liraki, Z., Harounabadi, A., &Mirabedini, J. (2015). Predicting the Users' Navigation Patterns in Web, using Weighted Association Rules and Users' Navigation Information. *International Journal of Computer Applications*, *110*(12), 975-8887.

[29]. Tyagi, S., & Bharadwaj, K. K. (2012). Enhanced new user recommendations based on quantitative association rule mining. *Procedia Computer Science*, *10*, 102-109.

ISSN NO: 2230-5807

[30]. Thiyagarajan, R., Thangavel, K., &Rathipriya, R. (2014). Recommendation of web pages using weighted K-means clustering. *International Journal of Computer Applications*, *86*(14), 44-48.

[31]. Rajaram.A., Dr.S.Palaniswami . Malicious Node Detection System for Mobile Ad hocNetworks. (IJCSIT) International Journal of Computer Science and InformationTechnologies, Vol. 1 (2) , 2010, 77-85

[32]. Moniz, N., Torgo, L., Eirinaki, M., & Branco, P. (2017). A framework for recommendation of highly popular news lacking social feedback. *New Generation Computing*, *35*(4), 417-450.