JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Frequent Pattern sub Graph generation on Online Social Network data

Aparna Annaso Patil[1], Prathmesh S. Powar[2]

[1] M.Tech. Student, Department of Computer Science & Engineering , Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India.

[2] Asst. Professor, Department of Computer Science & Engineering , Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India

***Abstract* – Across online social networks, a sizable portion of society is now being brought onto the online platform. To build interconnection between people over great distances, this will be crucial. Such social media networks have proven to be among the most useful strategies for successfully fostering user interaction. Users can effortlessly exchange and communicate with a multitude of people. A great deal of information is produced by this relationship and the countless posts and other elements on social media websites, from which critical information may be extracted. This is problematic because to the huge volumes of data involved and the assumption that it would have been difficult to determine the connections and relationships between the data in order to retrieve it. For this, this paper explores relevant research and develops a successful method for converting data from social media into a graph database which use Neo4j, and also the implementation of term weight, Shannon information gain, and pattern protocol, including the ECLAT algorithm and fuzzy classification. This approach has been evaluated extensively through the use of experimentations which has achieved highly satisfactory results.**

Keywords: *ECLAT, Shannon Information Gain, Fuzzy Classification, Neo4j.*

I. INTRODUCTION

The need for social media has grown substantially during the past several years. An overwhelming number of individuals and organizations are now using the technology to conduct out personal and professional tasks as a result of the quick development of online systems and applications. To understand their customers and market their products or services, organizations, institutions, and private citizens must assess this pertinent information and present media or internet services in instantaneously. The demand for online multimedia stream suggestions has grown as social media in the format of streaming has proliferated. Offering streaming services over social networks is essential for several uses, including investigative journalism, online commercial advertising, and recreation.

The use of social networking websites has contributed to the growth of a far better level of global communication and connection. For many of these individuals, the internet community has become a necessary part of daily life. This is demonstrated by the vast array of unique applications that have already been created to function effectively when used online. People’s livelihoods have improved as a consequence of the internet's rapid adoption. The emergence of multiple platforms that significantly simplify the lives of customers has been made possible by developments in the digital realm.

Huge volumes of data are produced over and over on the internet, particularly on social networking sites. This information is very important because it has the potential to produce data that is relevant. But managing and interpreting such vast amounts of data is extremely difficult. This is a result of the enormous number of people who disclose their information on social media platforms. Social media and social networking platforms can exchange multimedia data that is recorded in a variety of formats, including photographs, videos, and text-based content. The unavailability of a practical method for assessing this data is among the problems that this research aims to overcome.

Data administration and comprehension are considerably improved when machine learning algorithms are used in association with graph databases. The innovations that have been developed in the modern world with internet access as their base are immensely distinctive and socially beneficial overall. The online community has proven to be a very valuable resource that has contributed to the rapid growth of all facets of life. The demand for an immersive environment has led to the enormous popularity of social networks. Nearly every day, users submit a significant amount of data to social media networks, engage with their followers and online friends, and integrate media. The material on the network also includes marketing advertisements and product interaction activities like promotion.

Similarly, recent researches such as SOCMI is proposed by L. Li et al. [1] to solve the bottleneck problem of frequent pattern mining in large social networks. Its fundamental idea is to use a pathgraph to store the appearance of patterns, making it easier for SOCMI to expand patterns and compute frequency during mining. Furthermore, ASOCMI is presented to obtain improved performance on dense huge networks such as social graphs by adopting a rapid exploration technique and a pathgraph generation approach of edges to overcome the exponential time consumption of acquiring neighbor information. Experiment findings suggest that the proposed methodologies can outperform conventional approaches by up to two orders of magnitude.

Additionally, W. Fang et al. [2] created a multi-objective pattern mining issue model to simultaneously maximize support, occupancy, and utility values to mine high-quality patterns to fulfill the many criteria of users. Then, for addressing the problem model, an upgraded MOEA (dubbed MOEA-PM) was presented. There are two novel population initialization procedures shown and explored. The effectiveness of the two proposed initialization approaches was thoroughly analyzed and debated. An extra tool is presented and utilized to speed up algorithm convergence. MOEA-PM, unlike classic pattern mining algorithms, does not need the specification of parameters such as minSup, minOccu, and minUti.

This research article on personality classification is classified in 5 segments. The segment 1 is provides the introduction to this publication, whereas section 2 describes the related works done on this topic in the past. Section 3 presents the proposed idea of our technique and section 4 discusses the evaluation paradigm along with the results obtained. Finally, Section 5 provides the conclusion and the directions for future enhancement to this research.

II. Literature Survey

S. U. Rehman et al. examined the connections between FSPs and opt-SGPs. The FSPs were uncovered in the author's previous work utilizing the proposed FSM framework A-RAFF. The opt-SGPs were identified utilizing the PSO-based optimization strategies described in this research. To find the link, the PSO-based optimization approach is proposed. To structure the particle in the suggested PSO process, several graph measurements and features were utilized from the literature [3]. In addition, a scoring function that works as a fitness function is created depending on several node attributes gathered from the literature. One of the most significant benefits of this research will be the reduction of FSPs by selecting those FSPs in the final result set that were also proven to be optimized. Several tests were carried out to validate the notion of a link between FSPs and opt-SGPs.

S. Jamshidi Nejad et al. suggest an approach for explicit information retrieval from Persian reviews for both single-word and multi-word aspects, with the latter being a more difficult task. To that end, the authors propose the development of a weighted directed graph as an ADG structure in the proposed technique. They may be able to extract the problematic aspect compounds by traversing selected pathways of the ADG network using their created technique. The graph pieces, including edges and weights, are produced from the output of FP-frequent Growth's pattern detection method [4]. The proposed technique is used for reviews in Iran's domestic hoteling area. To build and retrieve data, they used the Neo4J graph database system and its Cypher query language.

J. M. -T. Wu et al. created two tighter upper-bounds for HAUIs to further limit the search area while maintaining its downward closure property and correctness. A recursive method generates the processing order and prunes unneeded elements in the proposed algorithm. To decrease evaluation time, a transaction-rival pruning technique was also created to establish a hard constraint for all candidate item sets [5]. Experiments on five real-world datasets show that the new technique beats the prior EHAUPM algorithm considerably. They also discover certain drawbacks to the suggested TUB-HAUPM method. However, they find that the suggested TUB-HAUPM procedures are acceptable for use in a parallel computing environment.

P. Das et al. illustrate an unsupervised strategy for extracting criminological relationships from newspapers. The suggested clustering approach finds major crime trends, which can be useful in both criminology and the criminal justice sector. This experimental research effort sheds light on three distinct facets of crime against women in India. They called the clusters to depend on the most common context word however, it is possible that some of the context words in the cluster do not represent the same criminal feature as the cluster label [6]. In such a situation, they can gather context terms that define the same meaning.

C. Yang et al. investigated how to determine users' true identities by examining the sequences of their accessed objects in cyberspace, and conducted an IPTV user matching case study. They originally proposed an n-level multi-item-set user behavior fingerprint matching framework for UI. Then, by merging two generic similarity distances, they presented a new similarity distance [7]. Finally, the authors presented a fusion decision technique to increase accuracy at the expense of more rejections. Using a large-scale IPTV dataset, the authors conducted comprehensive tests. In behavior fingerprinting and similarity metrics, the authors proved that the current technique outperforms the state-of-the-art. Furthermore, they confirmed the key results gained from the IPTV dataset using two other datasets.

WGIN, a session-based recommendation model that takes into account the repeated link effect, is proposed by P. Das et al. as a novel architecture. This model collects user preferences from both the graph and sequence representations. Using a repeated weighted graph neural network, the suggested technique not only analyses the complex structure of transitions between items occurring in session sequences but also combines several levels of information to extensively study user preferences [8]. A Transformer-based module also employs a long multi-interest attention mechanism and a short single-interest attention mechanism to extract long-term and short-term user preferences. Extensive trials on two real-world datasets show that the proposed strategy outperforms current cutting-edge session-based recommendation algorithms.

To discover HTFUIs, the FHTFUP method depends on the two-phase upper bound model, and FP-tree was presented by T. -P. Hong et al. With the downward-closure feature maintained the model may locate all potential itemsets (HTFUUBIs). When compared to the generate-and-test technique, using FP-tree reduces the number of explosive candidates [9]. During the experimental assessment, the two methods were evaluated with correspondingly modified parameters on the synthesized and two actual datasets. On diverse datasets, the findings reveal that the new strategy has a lower execution efficiency than the previous method.

D. Jiaman et al. present an association rule-based classifier chains technique (ARECC for short). The technique begins by combining the concept of common patterns to build a computation technique of relevant information between labels depending on association rules. The pairwise dependency and degree of reliance between the labels are then utilized to form a directed acyclic graph. The directed acyclic graph is then topologically sorted to acquire the sequence that indicates label dependency, and the resultant topological sequence is employed as the learning order of the labels in the classifier chains approach [10]. Finally, the association rules are employed to change and update the classifier chains model's prediction outcomes. ARECC can effectively increase classification performance, according to experimental results on a range of public multi-label datasets.

P. -Y. Huang et al. developed the DFP method to improve the mining efficiency of association rules using a distributed approach. In the offered publications, the distributed technique cannot be readily applied for database processing. As a result, as client communication grew, the transmission cost would quickly climb. According to the authors' research, the proposed DFP technique overcomes this challenge by moving the mining result from the performing client to the server, because the integrated information is sent through the server rather than among clients [11]. The results of their experiments also demonstrated that the proposed approach had significantly more computational efficiency than DistEclat and BigFIM, needing just 45 percent and 14.2 percent of the transmission costs of DistEclat and BigFIM, respectively. By estimating the maximum amount of TIDs for each client, the DFP approach avoids memory limits and repetitive scans, and the remaining TIDs are handled by an enlarged mechanism known as MP.

M. Drlik et al. present an alternate strategy for utilizing identified frequent itemset that does not include any temporal information in the research of behavioral changes of VLE stakeholders over time. As previously stated, while the discovery of behavioral patterns in VLE stakeholders' behavior from pre-processed data is frequently dependent on the analysis of sequences or time series, the chosen approach of frequent itemsets analysis examines a set of activities that the stakeholder has visited or in which he/she has participated in e-learning courses over time [12]. The suggested technique is distinct in that it is not concerned with estimating trends or identifying stakeholders who exhibited similar behavior over the studied time.

An improved balanced parallel FP-growth method is presented by A. Essam et al., along with a load-balancing technique that seeks to distribute items fairly between groups. Second, an improved technique enhances the conditional pattern base creation by deleting uncommon components from conditional patterns before generating conditional FP-trees. The authors' findings reveal that the EBPFP balancing techniques balance group load better than the PFP and BPFP balancing techniques [13]. The MPFP and BPFP were surpassed by the EBPFP. Furthermore, like other mining algorithms, FP-Growth in distributed processing systems needs a considerable volume of data to be communicated across the network. Therefore, one of the major issues for FP-Growth, particularly in this era of huge data, is bandwidth constraint.

Y. Zhou et al. proposed the FPBS technique, which tries to integrate data mining and optimization into the population-based search methodology The technique uses a modular and component-based approach to enable a wide range of applications, including subset selection and permutation issues in particular [14]. On the well-known QAP, the authors established the feasibility of the suggested FPBS approach. Extensive computational results on prominent QAPLIB benchmarks revealed that the resultant FPBSQAP algorithm outperforms relatively current state-of-the-art techniques.

III. PROPOSEDSYSTEM

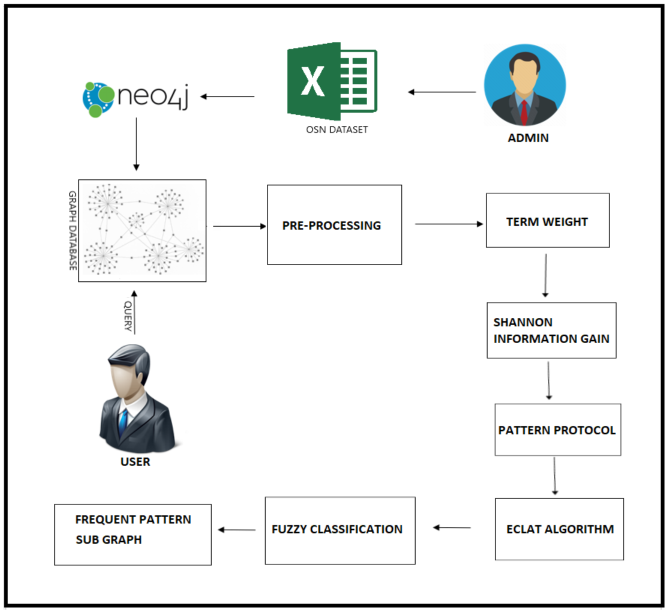

Figure 1: System Overview

The proposed approach for the purpose of achieving the frequent pattern sub graph on Online Social Network dataset consisting of conversational tweets posted on the Twitter website. The entire process is divided into consistent steps that are outlined in the section below.

Step 1: Data Collection and Preprocessing –The proposed technique requires the input of a dataset containing responses of individuals in the form of tweets on the Twitter Online Social Network. For this purpose a dataset containing tweets has been extracted from the URL - https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter?resource=download.

This dataset contains the tweet conversational data that has recorded the conversation between two individuals using tweets on the Twitter OSN. This is an important event for our implementation of the system and is the appropriate data and the content of tweets which is highly suitable for our implementation. The tweets contain attributes such as tweet id, author id, inbound, created at, tweet content, response tweet id, and in response to tweet.

The tweets need to be preprocessed before providing it to the system to reduce the incidences of any error or redundancy that may impact the performance negatively. The preprocessing approach has been elaborated in the next step of the approach.

Step 2: Preprocessing – The preprocessing approach is the initial logical step of the approach that is designed to facilitate the processing of the tweet before it can be provided to the system. The extracted dataset achieved in the previous step is provided as an input in this step of the procedure. The dataset is in the workbook format, therefore the JXL library is being used to interface this file with the java code. The dataset is then converted into a double dimension list that can be easily processed by the system.

The process of preprocessing is effective in realization of the conditioning of the input tweets that are provided. This is highly crucial as the performance of the execution depends on the effectively of this step of the procedure. The presence of any unnecessary data can be detrimental to the system as it could result in an error reducing the accuracy. The redundant data can also take longer to process which can be problematic to achieve efficiency. The steps for preprocessing are defined below.

Special Symbol Removal – The input string containing the tweet is provided to this initial step of the procedure where the special symbols are removed. These special symbols are used to provide grammatical pauses and other nuances to the text. This is redundant in our implementation and can be removed without any negative effects to the tweet. These symbols, such as ?, , etc are removed from the tweet.

Stemming – This is the second step in the preprocessing approach where the special symbol removed tweet is provided as an input. This step deals with making the input lightweight which can considerably reduce the processing time of the system. Majority of the words in the English language come from the same root word and just differ from one another just by a postfix. These words are then stemmed to their specific root words, which does not change the meaning of the word. For example, the word sleeping will be truncated to sleep, which doesn’t change the semantics of the word while making the tweet lightweight.

Stop Word Removal – The stemmed tweet is provided to this step of the preprocessing for the purpose of stop word removal. In the English language, stop words are words that are used to provide connection between two sentences or combine two different parts of the same sentence. This is not required by our approach, and the removal of the stopwords does not provide any penalty in terms of comprehension of the sentences. Therefore, the words such as and, is, the are eliminated from the tweet to achieved the preprocessed tweet which is provided to the next step for further processing.

Step 3: Term Weight – The preprocessed string that is achieved the previous step is being used to provide input to this step of the procedure. This input string is being utilized for the purpose of extracting the terms or the words from the preprocessed tweets. These words are being used for the purpose of achieving the term weight of the extracted words. The term weight extracts and evaluates the frequency of the words in the input string. This is done by identifying the space character in the string to isolate the words by splitting the string at these junctures.

The extracted words are then added to a list and this process is repeated consistently for all the tweets in the preprocessed list. Once the list with all the words is achieved, the hash set function is utilized for the purpose of achieving the unique words in the list. These unique words are then used for the purpose of counting the frequency of their presence in the preprocessed string. The Term Weight is an essential feature that is extracted for the purpose of achieving the effective and useful realization of the entropy estimation in the next step of the approach.

Step 4: Shannon Information Gain–This step of the procedure utilizes the term weight achieved previously for the purpose of estimating the entropy. This step of the procedure analyzes and measures the distribution factor of the words in the tweets provided as an input. The entropy is estimated using the Shannon Information Gain approach through the equation 1provided below.

$$IG(E) = -(P/T)\log(P/T) - (N/T)\log(N/T) \quad \text{--------(1)}$$

Where

P= Frequency of the present count
N= Non presence count
T= Cluster Elements Size.
IG(E) = Information Gain for the given Entity

Step 5: ECLAT Algorithm - The by the using the vertical intersection of the words system identifies the most obvious words for rule mining using powerset. Where all these words are extracting by the comparative recursion of the combination of the words.Then after fetching the important words from all the documents system will perform association rule using Apriori Algorithm with the step stated below.

Let T be the training data with n attributes $A_1, A_2 \ldots A_n$and C is a list of class labels. A particular value for attribute Ai will be denoted ai, and the class labels of $C$are denoted cj.An item is defined by the association of an attribute and its value (Ai, ai), or a combination of between 1 and$n$ different attributes values, e.g. < (A1, a1)>, < (A1, a1), (A2, a2)>, (A1, a1), (A2, a2), (A3, a3)> … etc.

A rule r for multi-label classification is represented in the form:$(A_{i1},a_{i1}) \wedge (A_{i2},a_{i2}) \wedge \ldots \wedge (A_{1m},a_{im}) \rightarrow c_{i1} \ldots . c_{im}$where the condition of the rule is an item and the consequent is a list of ranked class labels.The actual occurrence (*ActOccr*) of a rule r in T is the number of cases in T that match r*'s* condition. The support count (*SuppCount*) of r is the number of cases in T that matches r*'s* condition, and belong to a class ci. When the item is associated with multiple labels, there should be a different *SuppCount* for each label.

A rule r passes the minimum support threshold (*MinSupp*) if for r, the $SuppCount(r)/|T| \geq MinSupp$, where $|T|$ is the number of instances in T. A rule r passes the minimum confidence threshold (*MinConf*) if $SuppCount(r)/ActOccr(r) \geq$ MinConf. Any item in T that passes the *MinSupp is* said to be a frequent item.

In the final step proposed system will perform vertical frequent pattern mining using éclat algorithm as shown below.

Algorithm 1 ECLAT Algorithm

Input: Alphabet A with ordering ≤ multiset T ⊆ P (A) of sets of Items, Minimum support value minsup Є ℕ.
Output: Set F of frequent Itemsets and their support counts.
F :={(Ø,|T|) }.
CØ: = {(x,T({x}))| x Є A}.
C'Ø: = freq ($C_{Ø}$):= {$(x,T_x)|(x,T_x)$ Є $C_{Ø}$, $|T_x| \geq$ minsup }
F: ={Ø}.
Add frequent supersets (Ø, $C'_{Ø}$).

Function add Frequent Supersets ():
Input: frequent Itemsets p Є P (A) called prefix, incidence matrix C of frequent 1-item-extentions of p.
Output: add all frequent extensions of p to global variable F.
for (x, T_x) Є C **do**
q:= p U {X}.
C_q:={$(y,T_x \cap T_y)$ | (y,T_y) Є C, y > x}.

$C'q := freq(C_q) := \{(y,T_y) \mid (y, T_y) \in C_q, \ |T_y| \geq minsup \}$
If $C'_q \neq \emptyset$ **then**
Add frequent supersets (q,C'_q).
End if
$F := F \cup \{(q, |T_x|)\}$
End for

IV RESULT AND DISCUSSIONS

The presented approach for the purpose of achieving Frequent Pattern Sub Graph has been achieved using the Java Programming language and the coding has been performed using the NetBeans IDE. The realization of the approach has been done on a development machine with standard configuration that consists of 8 GB RAM, 500GB HDD powered by an Intel Core i5 processor.

The experimental evaluation has been performed using Precision and Recall performance metric. The assessment details have been stipulated in the section given below.

Precision and Recall Performance Metric

Precision and recall are regarded as the one of the best parameter to quantify the performance of our system. Precision is known for the positive predictive values that indicates the amount of relevant information matched through the system. Precision can be stated as the ratio of number of relevant patterns matched for the given number of user comments to the sum of number of relevant and irrelevant patterns matched for the given number of user comments. Relative effectiveness of the system can be evaluated thoroughly by using precision parameters.

Recall generally indicates the part of relevant results matched over the matched relevant results. Recall can be defined as the ratio of number relevant patterns matched to the sum of relevant patterns not matched. Absolute accuracy of the system can be properly denoted by this system.

Precision can be more effectively explained as below
A = The number of relevant patterns matched for the given number of comments
B= The number of irrelevant patterns matched for the given number of comments
C = The number of relevant patterns is not matched for the given number of comments

So, precision can be given as

Precision = (A / (A+ B)) *100
Recall = (A / (A+ C)) *100.

No of Given Comments	Relevant Patterns Extracted (A)	Irrelevant Patterns Extracted (B)	Relevant Patterns not Extracted (C)	Precision = (A / (A+ B)) *100	Recall = (A / (A+ C)) *100
25	23	2	2	92	92
50	46	1	4	97.87234043	92
75	74	0	1	100	98.66666667
100	93	3	7	96.875	93
125	120	2	5	98.36065574	96
150	141	5	9	96.57534247	94
175	172	2	3	98.85057471	98.28571429
200	188	2	12	98.94736842	94
225	214	7	11	96.83257919	95.11111111
250	236	6	14	97.52066116	94.4

Table 1: Precision and Recall Results

On conducting the different experiments for the different given number of inputs of comments system records the result for both precision and recall as mentioned in the above table 1.

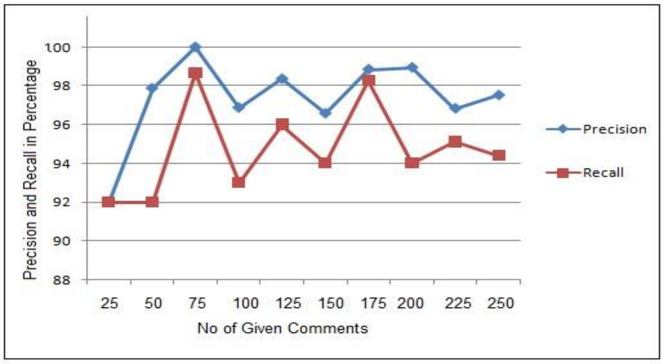

Figure 2: Performance Evaluation through Precision and Recall

The graph in above figure 1 clearly indicates that proposed model for sub graph matching achieves 97.3 % of average precision and 94.7 % of average recall for the given user comments data. This is high and indicates successful deployment of the proposed idea.

V CONCLUSION AND FUTURE SCOPE

The presented approach for the realization of the frequent pattern sub graph formation on social media content has been elaborated in this research article. The existing system stated that there has been an exponential increase in the number of users utilizing social media and other networking websites which has inadvertently led to the establishment of a complex relationship between various users. This relationship is based on similar interest and tastes which is highly lucrative for businesses and the customers alike. These similar users are then bundled together in a relational graph database with the help of various identification of similar features and matching their subgraphs for an accurate cluster. To identify the similarity between the users, there have been systems that are developed to identify the matching sub graphs. Doing this does not achieve any satisfactory results due to the restrictive and strict nature of the matching process. Therefore, the proposed system utilizes pattern identification in conjunction with frequent pattern analysis with the combination of ECLAT algorithm and Fuzzy classification, for achieving the frequent pattern sub graph and identifying the correlation between the users efficiently. The approach has been evaluated with extensive experimentation that achieves highly satisfactory results.

The future research directions can involve the implementation of cloud infrastructure to achieve improved an effective realization of frequent pattern subgraph generation.

REFERENCES

[1] L. Li, P. Ding, H. Chen and X. Wu, "Frequent Pattern Mining in Big Social Graphs," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 3, pp. 638-648, June 2022, DOI: 10.1109/TETCI.2021.3067017.

[2] W. Fang, Q. Zhang, J. Sun and X. Wu, "Mining High-Quality Patterns Using Multi-Objective Evolutionary Algorithm," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 8, pp. 3883-3898, 1 Aug. 2022, DOI: 10.1109/TKDE.2020.3033519.

[3] S. U. Rehman, S. Asghar, and S. J. Fong, "Optimized and Frequent Subgraphs: How Are They Related ?," in IEEE Access, vol. 6, pp. 37237-37249, 2018, DOI: 10.1109/ACCESS.2018.2846604.

[4] S. Jamshidi Nejad, F. Ahmadi-Abkenari, and P. Bayat, "A Combination of Frequent Pattern Mining and Graph Traversal Approaches for Aspect Elicitation in Customer Reviews," in IEEE Access, vol. 8, pp. 151908-151925, 2020, DOI: 10.1109/ACCESS.2020.3017486.

[5] J. M. -T. Wu, J. C. -W. Lin, M. Pirouz and P. Fournier-Viger, "TUB-HAUPM: Tighter Upper Bound for Mining High Average-Utility Patterns," in IEEE Access, vol. 6, pp. 18655-18669, 2018, DOI: 10.1109/ACCESS.2018.2820740.

[6] P. Das, A. K. Das, J. Nayak, D. Pelusi and W. Ding, "A Graph-Based Clustering Approach for Relation Extraction From Crime Data," in IEEE Access, vol. 7, pp. 101269-101282, 2019, DOI: 10.1109/ACCESS.2019.2929597.

[7] C. Yang, L. Wang, H. Cao, Q. Yuan, and Y. Liu, "User Behavior Fingerprinting With Multi-Item-Sets and Its Application in IPTV Viewer Identification," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2667-2682, 2021, DOI: 10.1109/TIFS.2021.3055638.

[8] Z. Yang, H. Wang, and M. Zhang, "WGIN: A Session-Based Recommendation Model Considering the Repeated Link Effect," in IEEE Access, vol. 8, pp. 216104-216115, 2020, DOI: 10.1109/ACCESS.2020.3041772.

[9] T. -P. Hong, C. -Y. Lin, W. -M. Huang, K. S. -M. Li, L. S. -L. Wang and J. C. -W. Lin, "Using Tree Structure to Mine High Temporal Fuzzy Utility Itemsets," in IEEE Access, vol. 8, pp. 153692-153706, 2020, DOI: 10.1109/ACCESS.2020.3018155.

[10] D. Jiaman, Z. Shujie, L. Runxin, F. Xiaodong and J. Lianyin, "Association Rules-Based Classifier Chains Method," in IEEE Access, vol. 10, pp. 18210-18221, 2022, DOI: 10.1109/ACCESS.2022.3149012.

[11] P. -Y. Huang, W. -S. Cheng, J. -C. Chen, W. -Y. Chung, Y. -L. Chen and K. W. Lin, "A Distributed Method for Fast Mining Frequent Patterns From Big Data," in IEEE Access, vol. 9, pp. 135144-135159, 2021, DOI: 10.1109/ACCESS.2021.3115514.

[12] M. Drlik, M. Munk and J. Skalka, "Identification of Changes in VLE Stakeholders' Behavior Over Time Using Frequent Patterns Mining," in IEEE Access, vol. 9, pp. 23795-23813, 2021, DOI: 10.1109/ACCESS.2021.3056191.

[13] A. Essam, M. A. Abdel-Fattah and L. Abdelhamid, "Towards Enhancing the Performance of Parallel FP-Growth on Spark," in IEEE Access, vol. 10, pp. 286-296, 2022, DOI: 10.1109/ACCESS.2021.3137789.

[14] Y. Zhou, J. -K. Hao and B. Duval, "Frequent Pattern-Based Search: A Case Study on the Quadratic Assignment Problem," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 3, pp. 1503-1515, March 2022, DOI: 10.1109/TSMC.2020.3027860.

*****