

On Combined Approach for mining FSG in Transactionized Graph Datasets

K.Rajendran¹, Dr. T. Venugopal², A.V.Sriharsha³

¹Research Scholar, Department of Mathematics, SCSVMV University, Kanchipuram

²Professor, Department of Mathematics, SCSVMV University, Kanchipuram

³Research Scholar, Department of Information Technology, SCSVMV University, Kanchipuram

Abstract: Graph Data mining has ushered into new era with advanced data mining techniques. Mining Frequent Sub Graphs is the crucial area which appeals the ease of extracting the patterns in the graph. Typical graph data like Social Networks, Biological Networks (for metabolic pathways) and Computer Networks needs analysis of virtual networks of a category. Such graphs need be modeled as layered to distinguish the categories of relationships. Traditional Market Basket Analysis of Data mining has proven its elegance of mining Frequent Itemsets. Combining the techniques of Apriori with Collaborative Mining discriminates a new concept of mining FSG.

Keywords: Graph Data Mining, Link Analysis, FSG.

I. Introduction

Online social networks facilitate connections between people based on shared interests, values, membership in particular groups (i.e., friends, professional colleagues), etc. They make it easier for people to find and communicate with individuals who are in their networks using the Web as the interface.

There are several different online social networks, but for our purposes, we'll focus on the two that tend to be used the most by learning professionals—Facebook, LinkedIn. Each of these networks has its own unique style, functionality and patterns of usage and different people are active in these different networks. LinkedIn is primarily a professional network, designed to facilitate linkages between people who are wanting to connect for work-related purposes. Facebook was originally developed for college students to connect, so it has a more informal, social air than you find on LinkedIn. Now open to anyone, you will still find that Facebook is the preferred network for Millennials who see the encroachment of Boomers and, to a lesser extent. [1]

Common Features of Social Networks

The ability to create a Profile page—this is your main “home” on the network. Different networks offer varying abilities to personalize your page in terms of look and feel. They may also differ in terms of the types of information you would include, such as name, location, education, etc. Facebook, for example, asks for your relationship status (because it's more “social”), while on LinkedIn, which is primarily for professional use, does not.

A way to find and link to “friends” or connections—The purpose of a network is connections, so facilitating a members' ability to find and connect to other people is important. Each network offers different types of search capabilities and once you've located a potential friend, you must send an “invitation” to invite them into your personal network.

Privacy Controls—In most networks, your ability to access more detailed information about a person is based on their status as one of your connections; “friends” can see much more information than those who are not your “friends.” You can control who is actually in your personal network by effectively managing who you invite into your network and whose invitations you accept.

The ability to send public and private messages—In Facebook, you can communicate with your connections either by sending a private message or “writing on their wall.” On LinkedIn, you communicate via person-to-person messages.

Ability to share various digital objects and information—Facebook allow members to share various online items, including photos, videos and RSS feeds. LinkedIn offers some ability to share links, although it's multimedia capacities are nothing like what you find on Facebook.

The social graph in the Internet context is a graph that depicts personal relations of internet users. In short, it is a social network, where the word graph has been taken from graph theory to emphasize that rigorous mathematical analysis will be applied as opposed to the relational representation in a social network.[citation needed] The social graph has been referred to as “the global mapping of everybody and how they're related”. Several issues have come forward regarding the existing implementation of the social graph owned by Facebook. For example, currently, a social networking service is unaware of the relationships forged between

individuals on a different service. This creates an online experience that is not seamless, and instead provides for a fragmented experience due to the lack of an openly available graph between services. In addition, existing services define relationships differently.

This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks.

The Contingent Internet

Due to the increase in HyperText Transfer Protocol (HTTP) traffic on the World-Wide Web (WWW), the amount of transaction log records generated and collected on the servers grows tremendously. In order to benefit from these server log records, data mining has emerged as a tool to extract any useful patterns and analyze user access behavior on the Web. This specific type of data mining technique is known as Web usage mining [15]. In particular, the technique of mining user access patterns (also known as browsing patterns and path traversal patterns) has been applied in a wide range of applications including Web caching [11, 13], Web page recommendation [6, 8, 9], and Web personalization [10, 12].

In general, mining user access patterns can be considered as a special type of mining sequential patterns in the field of knowledge discovery and data mining. Association rule mining has recently attracted considerable attention and proven to be a highly successful technique for extracting useful information from very large databases [2, 7, 14]. For the problem of mining user access patterns, data sequences are typically user access sequences of Web pages. These access sequences are extracted from server log records via some Web data preparation techniques [5]. Applying a method for mining user access patterns on these access sequences reveals the user browsing behavior on the Web.

Data Set

Facebook data has been anonymized by replacing the Facebook-internal ids for each user with a new value. Also, while feature vectors from this dataset have been provided, the interpretation of those features has been obscured. For instance, where the original dataset may have contained a feature "political=Democratic Party", the new data would simply contain "political=anonymized feature 1". Thus, using the anonymized data it is possible to determine whether two users have the same political affiliations, but not what their individual political affiliations represent.

II. Related Work

In this research we undertake the elementary concepts of graph theory are induced into the data mining, assuming that there are large graphs such as network data in the areas of social networking, biological networks etc. The progress of the research is a fundamental approach that provokes idea for a mathematician and data miner to absolve the data mining techniques into graph theory.

The motivation is driven with principles of graph theory and machine learning algorithms i.e., Apriori and variants. The glimpse of motivation for the research is cited below:

1. The consequential results of the experiment should be simple.
2. Desire to use less theoretical jargon and find a feasible method for mining frequent subgraph.
3. Desire to discover a challenging combination of classical methods with machine leaning methods of data mining.

Our problem is entitled "Fast Frequent Subgraph Mining on Transactionised Graph Data" is the algorithmic state of the art technique that is implemented anew into the graph data mining. Frequent Itemset Mining for transactionised data is already in vogue proposed as Apriori, FP Growth, eClat, deClat and so many. For Graph data mining and clustering various algorithms exist for clustering, classification and for performing several graph based operations. DFS, BFS based algorithms, divide and conquer algorithms have been already used to accomplish trend setting problems. In the proposed problem, the framework for converting graphs into transactionised data and putting efforts to mine frequent itemsets, and then converting the fragments of the frequent itemsets into sub graphs has been accomplished.

The Combined Collaborative-Apriori (CCApriori) algorithm generates accurate FI itemsets from the graph data which is an intersection of the outputs from classical and specific algorithms on graph data. Transactionization is defined in the research work, where the graph data is converted into the business transaction and applied with classical Apriori. The algorithm is experimented on the complex data sets, bench mark real time data sets and synthetic data sets.

Apriori Algorithm on Graph Data

Apriori is a algorithm proposed by R. Agrawal and R Srikant in 1994 [2] for mining frequent item sets for Boolean association rule. The name of algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties. Apriori[2] is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

Apriori uses a “bottom up” approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length $k-1$. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k -length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

By applying Apriori – the Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent itemsets in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

Apriori Algorithm: (by Agrawal et al at IBM Almaden Research Centre) can be used to generate all frequent itemset.

Pass 1

1. Generate the candidate itemsets in C_1
2. Save the frequent itemsets in L_1

Pass k

1. Generate the candidate itemsets in C_k from the frequent itemsets in L_{k-1}
2. Join L_{k-1} p with $L_{k-1}q$, as follows:
insert into C_k
select $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$
from $L_{k-1} p, L_{k-1}q$
where $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2}, p.item_{k-1} < q.item_{k-1}$
3. Generate all $(k-1)$ -subsets from the candidate itemsets in C_k
4. Prune all candidate itemsets from C_k where some $(k-1)$ -subset of the candidate itemset is not in the frequent itemset L_{k-1}
5. Scan the transaction database to determine the support for each candidate itemset in C_k
6. Save the frequent itemsets in L_k

III. Collaborative Association Rule Mining Algorithm on Graph Data

Various algorithms and techniques for mining user access patterns have been proposed in the literature. In [11, 12], variations of the Markov model such as first-order Markov model and all K^{th} -order Markov model were applied to construct a predictive model to predict the user requests on Web pages. Their work mainly focused on the analysis of consecutive sequential access of Web pages, and hence given a currently visiting Web page, the ability to predict the next request is limited to the following adjacent Web pages on the user access sequence. For example, given a user access sequence containing n Web pages in an ordered list, (p_1, p_2, \dots, p_n) , where p_i represents a Web page, an approximation of the first-order Markov model would contain the transitional probabilities of two adjacent Web pages in the user access sequence, $Pr(p_i | p_{i-1})$, where $1 < i \leq n$.

Most of the previous research work focused on the forward and backward accesses, where forward accesses are those accesses that browse the Web pages by following the hyperlinks embedded within the Web pages, and backward accesses are those that access the Web pages by backtracking to the previous Web pages. For example, the user access sequence was divided into smaller sequences called the maximal forward references, and the effect of backward references was not considered. However, considering only the smaller sub-sequences of the user access patterns does not fully capture the user’s intention of accessing a particular set

of Web pages, since some of the Web pages may be put into a different access sequence. Another type of accesses is the jump accesses, which the user retrieves a Web page by entering the Uniform Resource Locator (URL) directly on the Web browser. In this paper, all three access types are considered when the model is constructed. We pruned out the duplicate Web pages in the access sequences, since our goal is to predict the Web pages which the user has not yet visited. A user access sequence is used to represent a data record during the mining process.

Using this user traversal structure, a shortest path problem in the Graph Theory is applied to find the “access” distances between Web pages. We refer to these distances as Minimum Reaching Distance (MRD) information. The Association Rule Mining technique is then applied to find a set of predictive rules that pass the user-specified minimum support. The MRD information is used to prune the results from ARM in order to increase the prediction accuracy and reduce the space complexity. The proposed method for mining user access sequences was applied as a collaborative filtering technique. The results from the process of mining user access patterns are a predictive rule set that is used to recommend Web pages according to the users who accessed the Web site in the past [8, 9]. Under the Markov model notion, our method can be viewed as the All Kth-Order Markov model with the lookahead ability, which allows the prediction to include multiple non-consecutive Web pages, i.e., any Web pages within the Web site which are not necessarily connected by hyperlinks. For example, the approximation of the first-order Markov model with the look-ahead ability would contain the following transitional probabilities of two Web pages including non-consecutive ones, $Pr(p_j | p_i)$, where $1 \leq i < n$ and $i < j \leq n$. The remainder of this paper is organized as follows. In Section 2, the method of mining user access patterns based on association rule mining is explained in details. The experiments on a real Web data set are given in Section 3. In the same section, the experimental results are presented and analyzed. Conclusion is given in Section 4.

We propose to apply the association rule mining (ARM) technique [2] in mining user access patterns on the Web pages. The objective is to construct a model that predicts users’ Web page requests to assist in browsing the Web pages and to reduce the access time. In ARM, a k-itemset contains k items that pass the user specified minimum support value. Hence, in essence, ARM enables one to discover interesting patterns or associations among the k-itemset in a given collection of records. Our framework applies the ARM technique to find the frequent itemsets of Web pages from the user access sequences and to construct a set of rules based on those itemsets. Generally, the number of rules constructed from ARM is large. In the original algorithm, the minimum confidence value is used to prune the rules; while in our proposed framework, the number of rules is pruned by incorporating the MRD information, which reduces the state complexity of the model. Our MRD calculation adopts the concept from the shortest-path problem in the Graph Theory. The original algorithm assumes that the traversal path follows the link structure of the graph, where a link structure is a representation of Web pages along with the embedded hyperlinks. However, the pages that a user accesses do not always follow the link structure of the Web pages, and hence an alternative Web representation based on the actual user browsing activity on the Web site needs to be constructed. Here, a user access sequence (also referred to as a browsing sequence or a traversal path) is an ordered list of Web pages accessed by a user during one session. User access sequences are extracted from the Web server log records as part of the data preprocessing step. The issues and analysis of the user access sequence are considered in many research works.

Proposed Work

Relational characteristics in data complicate the representation, and can be eased through graph structures. As Kevin Murphy et.al. has quoted “Graphical models are a marriage between probability theory and graph theory.” reinstates the importance of relationship with graph modeling and probabilistic studies. Graph data extends to be a resourceful structure in a large variety of applications and disciplines like social networks, community networks, biological networks, mutagenesis data, World Wide Web. Modern business environments use graphs for finding the products of the customers’ demand by drawing the evolution graphs.

The growth of graph theory during its first two hundred years could in no way foreshadow the spectacular progress which this area was to make. The graph representation as, a collection of nodes and links between nodes, does support all special aspects of the relational data mining process. Since the years of development of mathematical graph theory, it is said to have metrics, schematic notations to represent more qualities and characteristics of the data. A graph in convention may be large, can be reduced into an elementary set of graphs. Applications such as cheminformatics, biology and ecology demand the stratagem of representing the sub-graphs or graphs in small without prejudice to their loss of characteristic content.

Modeling objects using graphs allows us to represent arbitrary relations among entities. For example, we can convert a basket of items into a graph, or more specifically a clique, whose vertices correspond to the basket’s items, and all the items are connected to each other via an edge. Graph Model is a methodology for modeling, analyzing and understanding relationship strategy among the elements of dataset. The Graph Model methodology was first published in Kilgour et al. [1987], and the first comprehensive treatment appeared as

Fang et al. [1993]. Alternative way of modeling the various objects is to use undirected labeled graphs to model each one of object entities, items in traditional frequent itemset discovery, and the relation between them. In particular, each vertex of a graph will correspond to an entity and each edge will correspond to a relation between two entities. In this model both vertices and edges may have labels associated with them which are not required to be unique. Using such a graph representation, a problem of finding frequent patterns then becomes that of discovering subgraphs which occur frequently enough over the entire set of graphs.

Graph mining is a technique used to extract characteristic patterns from a variety of graph structured data [4]. The graph structure is a nice way of representing and explaining complex data forms and phenomena but because of its strong expressiveness its computational complexity has been a problem to extract specific patterns. However, recent development has made it possible to perform a complete search in extracting the entire subgraph in a reasonable computation time. Apriori-based Graph Mining algorithm is one of the most advanced algorithms for graph mining and is able to deal with directed/undirected and colored/uncolored graphs. While graph mining research is still in the developmental stage there is a fair amount of research being carried out already concerning its practical applicability.

We have considered various models which represent different streams of graph modeling research. As the prevailing graph models of application domain exists such as (PRG) power-law model, (PA) preferential attachment model, (SW) small-world model and some (RG) random graph models, they are probabilistic distributive and random models. Probabilistic models seem to be unrealistic, in the sense that they would assign probability 0 to most real-world graphs. Although, slight variations of these models are more appealing in that they retain the essential features of the model while assigning positive probability to every graph. In many applications the PA and SW models have been modified and used. [16]

The proposed work of this research is a medley between the classical Apriori and another Apriori variant on graph data. The Collaborative Association rule Mining (CARM) by Mei-Ling Shyu et. al. in IEEE 2005, TKDE, is one of the best comparison of the proposed algorithms to alleviate the problem of ignoring the propense item sets. The algorithms are performed in parallel and the results are intersected to defined best more group of item sets. The graph data is transactionised by identifying their categories and within the categorical groups of the data finding the association terms are very crucial. As many algorithms sort out the problem of mining the association terms, there are controversial and domain specific constraints. The CARM is especially made for the sequence data sets or vis a vis said to be transactionised data sets, and the Apriori is also made to work on the same data sets. Transactionization of graph data becomes crucial in our experiment. The following algorithm generates the transactionized group of datasets from a graph by converting the graph into an adjacency list.

GenerateAdjacencyList

```

Input: G(V,E)
Initialize List;
X = first(V)
For Each Vertex in G
    If Authority(Vertex) then
        List = Add(Vertex, Authority(Vertex));
    End If
Next
    
```

Return List;

The Apriori and CARM are combined into a GUI framework and they are executed in parallel to obtain the result. Following are the algorithms that build the framework.

| | |
|---|--|
| <p>Invoke Apriori Input G(V,E) Initialize <i>Apriori_TidSets</i> Initialize <i>Frequent_Items</i> For Each Row in the <i>Adjacency_List</i> Append_Record(<i>Apriori_TidSets</i>, Suffix(Row)) Next While(End_Of_Transactions(<i>Apriori_TidSets</i>)) <i>Frequent_Items</i> = <i>Apriori</i> {Join(<i>Apriori_TidSets</i>), Prune(<i>Apriori_TidSets</i>)}; Done Return Apriori TidSets;</p> | <p>Invoke MRD Table Input G(V,E) Tree = Generate_Tree(G) Initialize ItemCount Initialize Done() Array to False ItemCount = 2 While ItemCount <= Height_of_Tree For Each Node in the Level If HasLeaf(Node) Then Table(ItemRowIndex, ItemColIndex) = ConcatenateNodes(ItemCount, Node); Else Set Table.Done(ItemIndex) = True End If Next Done Return Table;</p> |
|---|--|

The complexity of execution of the algorithms in parallel reduces the search of association terms or frequent itemsets from the transactionised graph data sets.

CCApriori

```
Input: G(V,E)
Adjacency_List = GenerateAdjacencyList(G);
ItemSets_Of_MRD = Invoke_MRD_Table(Adjacency_List);
ItemSets_Of_FI = Invoke_Apriori(Adjacency_List);
Collaborative_FSG = { ItemSets_Of_MRD } U { ItemSets_Of_FI
};
Visualize(Collaborative_FSG);
End
```

Experimental Setup and Results

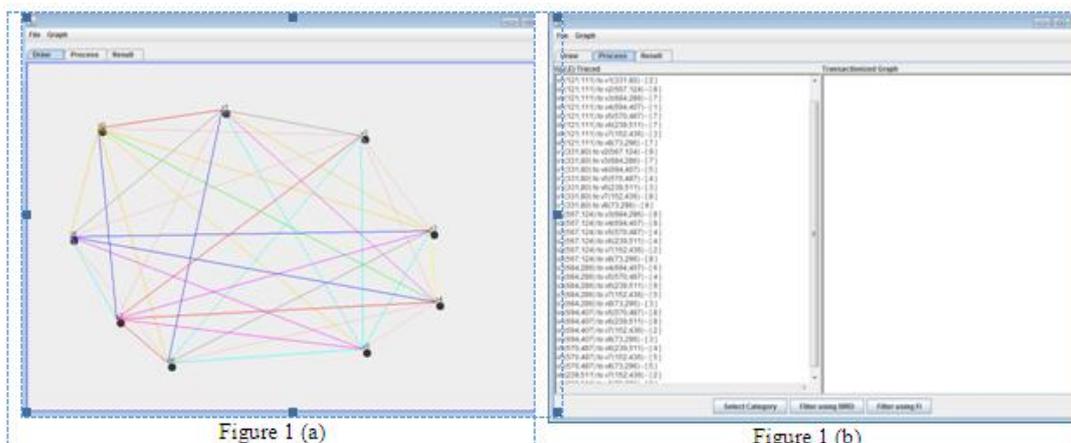
The experiment is conducted on synthetic data sets and Stanford Network Analysis Project data as bench mark and real time data sets. Synthetic data sets for the experiments are generated using interface in the project, where user can plot vertices on the plane and a large connected graph can be generated. The following algorithm generates the synthetic data sets for the experiment.

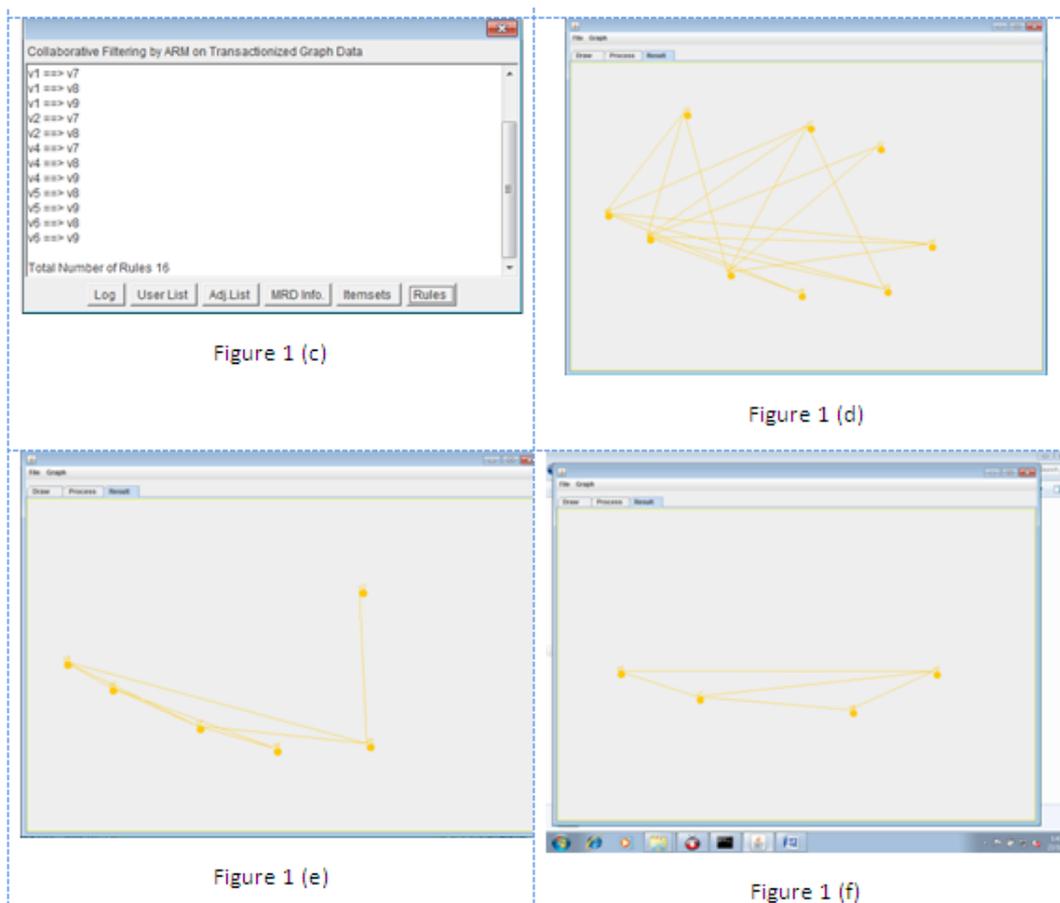
Generate Syn Graph Data

```
Initialize G
Initialize V, Source_Verx, Dest_Verx ( Vertex with Coords)
Initialize E ( Source_Verx, Dest_Verx)
Generate G
V = Read_All_Vertices(G);
For Each X in V
For Each Y in V
E = Draw_Edge(Coords(X), Coords(Y));
E.Category = GenerateCategoryOfRelationship();
Next
Next
Return G(V,E);
```

Operating Environment:

An operating environment is usually not a full operating system but is a form of middleware that rests between the OS and the application. The project is setup on Java 7 platform. The input is in the form of files for real-time data. The total software consists of java classes. The package is designed using AWT and Graphics. Following are the screenshots of the application run on standard PC, with 1 Gb RAM and 120 Gb hard disk space. Though there application is very small size, the standard requirements of RAM and processing capacity of the PC should be considered high for handling huge data sets.





The graph data selected from Stanford Network Analysis Project is ego-relationship data among the people of the Facebook social network. The experiment has a combined algorithm of Apriori and Collaborative Association Rule Mining using Minimum Reaching Distance. The classical Apriori can yield good results (FI) using candidate itemsets, the CARM-MRD generates the frequent itemsets with nearest distance of web graph data. The combination accelerates the production of FI and promises the accuracy as well. In general all generic FI mining algorithms should yield to good results which are similar in category. The CARM-MRD and Apriori are the challenging algorithms to generate the FI on two fields of data sets, like graph and market-basket data. The Combination of both algorithms as CCApriori results in FI which are unique and qualitative. The CCApriori has been experimented around one thousand times and found that the 70% of the results are proven. The results of CARM-MRD and the Apriori are intersected to find the final result.

A preprocessing and live graph drawing program was developed to collect the hyperlinks embedded within the Web pages for the link structure. From our observations, the maximum number of Web pages with unique URLs is developed which is equal to 3,948. The user log records are used to construct the user access sequences. Once all the user access sequences are identified, two-third of the data set with 34,362 user access sequences is used as the training data set; while one-third of it have 17,182 user access sequences and are used as the test data set.

Many graph color tables were drawn for analyzing the result comparatively.

| | | | | | | | | | |
|---|---|------|------------|--------|------------|--------|--------|--------|------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | | Blue | Light Blue | Yellow | Orange | Pink | Purple | Red | Light Blue |
| 1 | | | Light Blue | Pink | Pink | Yellow | Orange | Pink | |
| 2 | | | | Yellow | Yellow | Pink | Orange | Grey | |
| 3 | | | | | Light Blue | Pink | Red | Purple | Grey |
| 4 | | | | | | Yellow | Green | Yellow | Pink |
| 5 | | | | | | | Green | Purple | Yellow |
| 6 | | | | | | | | Purple | Blue |
| 7 | | | | | | | | | Light Blue |
| 8 | | | | | | | | | |

Following the sample of an observation from the results acquired during experimentation.

| CARM | FI |
|-----------|-----------|
| v0 ==> v7 | v5 ==> v7 |
| v2 ==> v7 | v5 ==> v8 |
| v3 ==> v6 | v6 ==> v7 |
| v3 ==> v7 | v6 ==> v8 |
| v4 ==> v7 | v7 ==> v8 |
| v5 ==> v7 | |

For a graph with 9 vertices and all drawn edges of random categories, v5 ==> v7 is the common rule from these experiments.

THUS v5 ==> v7 FORMS THE FREQUENT SUB GRAPH FROM BOTH EXPERIMENTS.

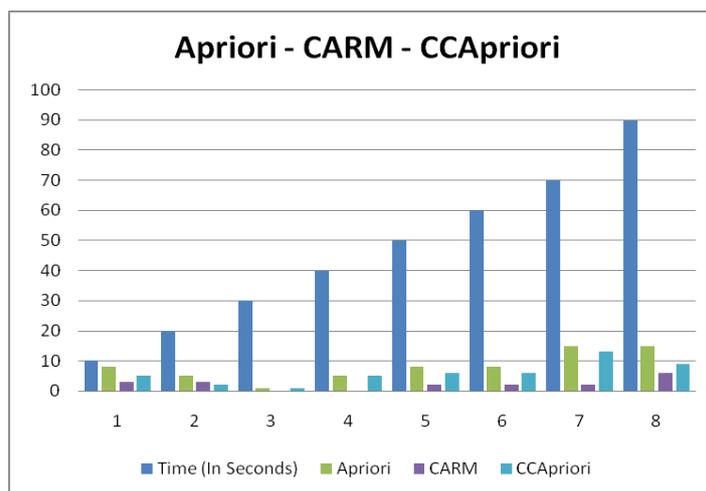


Chart 1

The chart shows the comparison between the three algorithms individually, for the maximum datasets of 110 of 3 categories, in various durations. The Y-axis shows the time in seconds and the X-axis shows the selected experiments in observations of the experiment. When the time taken for the running of CCApriori experiment the obvious results are shown with an improvement. But the results of the fixed data sets that belong to single domain will not vary after a threshold point of time. In the chart there is a gradual improvement of CCApriori upto seven observations but after that the time increases the results are normalized.

In this experiment, the precision and recall values are compared for the following four different approaches:

1. Association rule mining using link structure information (association rule mining (link));
2. Markov model using link structure information (Markov model (link));
3. Association rule mining using the MRD information (association rule mining (MRD)); and
4. Markov model using the MRD information (Markov model (MRD)).

IV. Conclusion

A comprehensive analysis is made on the models of the graph data sets; understand the features and structures of the graphs. Irrespective of the complexity of the algorithms in FI mining, the concept of mining subgraphs without using graph traversal, but only converting the datasets into transactions and mining as they were in a transaction database of a business is one of the important achievement, where the processing complexity of the algorithms on the transactionised databases are plane-sweep algorithms, compared to that of algorithms on the graphs. For an algorithm on graphs, it takes lot of recursive complexity to discover the frequent itemsets. In our research, we proposed a model call mereo topological model which consists of various layers of the datasets, as well as the layer based knowledge can be depicted as the transactionised data sets to discover the frequent sub graphs.

Reference

- [1]. www.wikipedia.com
- [2]. Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules. In Proc. 20th Intl. Conf. Very Large Data Bases, 487-499.
- [3]. Aristides Gionis, Mining the graph structures of the web, © 2008, Yahoo! Research, Barcelona, Spain, and University of Helsinki, Finland.
- [4]. Bettina Berendt, The semantics of frequent subgraphs: Mining and navigation pattern analysis, © 2005, ACM/WebKDD'05, Chicago, Illinois, USA.
- [5]. Martin J Wainwright, Michael I. Jordan, "Graphical Models, Exponential Families, and Variational Inference", Foundations and Trends in Machine Learning, Vol. 1 No.s 1 -2, ©2008.
- [6]. Michihiro Kuramochi, George Karypis, Frequent Subgraph Discovery, © 2001, IEEE.
- [7]. Nicholas Asher, Laure Vieu, Elsevier-Lingua, "Subordinating and coordinating discourse relations", 115 © 2005 pp. 591-610.
- [8]. Peixiang Zhao, Jeffrey Xu Yu, Fast Frequent Free Tree Mining in Graph Databases, ©2006, Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06).
- [9]. Sofus A. Macskassy, Claude C. Nanjo, Graph Mining using Graph Pattern Profiles, ©2008, ICAI 2008 - Knowledge Bases for Learning.
- [10]. Xifeng Yan, Jiawei Han, gSpan: Graph-Based Substructure Pattern Mining, © 2003, ACM/IEEE.
- [11]. Akihiro Inokuchi, Takashi Washio and Hiroshi Motoda, An Apriori-based Algorithm for Mining Frequent Substructures from Graph Data. © 2000, ACM/IEEE/PKDD2000.
- [12]. Ivona Bezáková, Adam Kalai, Rahul Santhanam, "Graph Model Selection using Maximum Likelihood", Appeared in 23rd International Conference on Machine Learning, Pittsburg, PA, © 2006.
- [13]. Jun Huan, Wei Wang, Jan Prins, Efficient Mining of Frequent Subgraphs in the Presence of Isomorphism, © 2003, ACM/IEEE.
- [14]. Karsten M. Borgwardt, Hans-Peter Kriegel, Peter Wackersreuther, Pattern Mining in Frequent Dynamic Subgraphs, © ICDM 2006, Hongkong.
- [15]. J. McAuley and J. Leskovec. Learning to Discover Social Circles in Ego Networks. NIPS, 2012.
- [16]. Graph Model Selection using Maximum Likelihood, Appeared in 23rd International Conference on Machine Learning, Pittsburg, PA, 2006.
- [17]. Graphical Models, Exponential Families, and Variational Inference, "Martin J Wainwright and Michael I. Jordan", Foundations and Trends in Machine Learning, Vol. 1 No.s 1 -2, ©2008.
- [18]. Introduction to Graphical Modeling, David Edwards © 1995, 2000 Springer-Verlag New York. pp. 4-11.
- [19]. Application of Graph based Data Mining to Biological Networks by Chang Hun You. University of Texas at Arlington, December 2005.
- [20]. K.Rajendran, Dr. T. Venugopal, A.V.Sriharsha, A Survey of Frequent Sub-Graph Mining. In the Proceedings of National Conference on Intelligent Systems and Communications, Organized by SPMVV, Tirupati – 2nd & 3rd December, 2009.
- [21]. K.Rajendran, Dr. T. Venugopal, A.V.Sriharsha, Graph-based modeling for Graph Data Mining, © In the Proceedings of CAMIST, 11-13, January 2010. pp 104. Published by MacMillan.
- [22]. K.Rajendran, Dr. T. Venugopal, Comprehensive Analysis of Complex Data Structure for Graph Data Mining, © JATIT, December 2010, Vol 23. No 1.

About Authors



Mr.K.Rajendran has completed M.Sc. and M.Phil in Madras University, Madras. He has profound experience of 16 years in working with defence and teaching. He has worked with graphs and their social applications (social networks). His areas of research are pattern detection and matching in large graphs, graph mining.



Prof. T.Venugopal has headed various programmes in Sri Chandrasekharendra Saraswathi Maha Vidyalaya (University), Kancheepuram. He has worked on various types of problems in graph, functional analysis, and wavelet theory with graph mining.



A.V. Sriharsha is B.Tech in Computer Science & Engineering from Andhra University and M.Tech in Information Technology from Sathyabhama University, Chennai. His main research interest includes Data Mining, Information Retrieval, Database Management Systems and Knowledge-based Software Architectures.