INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT AND MANAGEMENT

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Business Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email addresses, infoijrcm@gmail.com or info@ijrcm.org.in.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION**:

DATED: _____________

THE EDITOR

IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF ____.**

(e.g. Computer/IT/Finance/Marketing/HRM/General Management/other, please specify).

DEAR SIR/MADAM

Please find my submission of manuscript titled '____' for possible publication in your journal.

I hereby affirm that the contents of this manuscript are original. Furthermore it has neither been published elsewhere in any language fully or partly, nor is it under review for publication anywhere.

I affirm that all author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if our/my manuscript is accepted, I/We agree to comply with the formalities as given on the website of journal & you are free to publish our contribution to any of your journals.

NAME OF CORRESPONDING AUTHOR:

Designation:

Affiliation with full address & Pin Code:

Residential address with Pin Code:

Mobile Number (s):

Landline Number (s):

E-mail Address:

Alternate E-mail Address:

2. **INTRODUCTION**: Manuscript must be in British English prepared on a standard A4 size paper setting. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of the every page.
3. **MANUSCRIPT TITLE**: The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.
4. **AUTHOR NAME(S) & AFFILIATIONS**: The author (s) full name, designation, affiliation (s), address, mobile/landline numbers, and email/alternate email address should be in italic & 11-point Calibri Font. It must be centered underneath the title.
5. **ABSTRACT**: Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para.
6. **KEYWORDS**: Abstract must be followed by list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.
7. **HEADINGS**: All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
8. **SUB-HEADINGS**: All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.
9. **MAIN TEXT**: The main text should be in a 8 point Calibri Font, single spaced and justified.
10. **FIGURES &TABLES**: These should be simple, centered, separately numbered & self explained, and titles must be above the tables/figures. Sources of data should be mentioned below the table/figure. It should be ensured that the tables/figures are referred to from the main text.
11. **EQUATIONS**: These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.
12. **REFERENCES**: The list of all references should be alphabetically arranged. It must be single spaced, and at the end of the manuscript. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s**) for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio," Ohio State University.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITE

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Economic and Political Weekly, Viewed on July 05, 2011 http://epw.in/user/viewabstract.jsp

FREQUENT PATTERN MINING USING DYNAMIC PROGRAMMING

V. G. RAJALEKSHMI
ASST. PROFESSOR
DEPARTMENT OF MATHEMATICS
S.D. COLLEGE
ALAPPUZHA
KERALA

DR. M. S. SAMUEL
DIRECTOR
DEPARTMENT OF COMPUTER APPLICATIONS
MAR ATHANASIOS COLLEGE FOR ADVANCED STUDIES
TIRUVALLA
KERALA

ABSTRACT
Data mining methodologies have been developed for exploration and analysis of large quantities of data to discover meaningful patterns and rules. Frequent pattern mining is an important model in data mining. In this pattern mining given a data base of customer transactions, the task is to unearth all patterns in the form of sets of items appearing in a sizable number of transactions. Here the attempt is to discover frequent pattern using the Dynamic programming methods in operation research. Here we can calculate the frequent pattern item set by stages of operations.

KEYWORDS
Data mining, Dynamic programming, Frequent pattern mining.

INTRODUCTION
Frequent item set mining [1] is popular frame work for pattern discovery. Frequent patterns are patterns that appear in a data set frequently. For example set of items such as milk and bread, that appear frequently together in a transaction data set is a frequent item set. The concept of frequent item set was first introduce for mining transaction data base and was first proposed by Agarwal. et. al (1993).
Frequent mining searches for declining relationship in a given data sets. Frequent item set mining leads to the discovery of associations and correlations among items in large transactional or relational data sets. With massive amounts of data, continuously being collected and stored, many industries are becoming interested in mining such patterns from their data bases. The discovery of interesting correlation relationship among huge amounts of business transaction records can help in many business decision making process, such as catalog design cross marketing and customer shopping behavior analysis.
A typical example of frequent item set mining is market basket analysis. This process analyzes customer buying habits by finding associations between the different items that customers place in their 'shopping baskets'. The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. For instance, if customers are buying milk, how likely are they to also buy bread on the same trip to the supermarket? Such information can lead to increased sales by helping retailers do selective marketing and plan their shelf space.
This paper discusses frequent pattern by Dynamic programming methods. Here the problem is divided into different stages and at each stage frequent patterns are found. Section 2 deal with an overview of some related work for mining frequent patterns. In section 3 some definitions are given. Section 4 is about Dynamic programming methods. Section 5 discuss frequent mining model with DP. In section 6 an example is discussed. Final section is the conclusion.

RELATED WORKS
As an important data mining problem, frequent pattern mining plays an essential role in many data mining tasks, such as mining associations [2,8], sequential patterns [18,13] maximum patterns and frequent closed patterns [5,12,16] classification [6,15] and clustering [4]. There have been many algorithms developed for fast mining of frequent patterns which can be classified into two categories. The first category, candidate generation and test approach, such as Apriori [2] as well as many subsequent studies, are directly based as an anti-monotone. If a pattern with K items is not frequent, any of its super-patterns with (K+1) or more items can never be frequent. A candidate generation and test approach iteratively generates the set of candidate patterns of length (K+1) from the set of frequent patterns of length K (K≥1), and check their corresponding occurrence frequencies in the data base.
The Apriori algorithm achieves good reduction on the size of candidate sets. However, when there exist a large number of frequent patterns or long patterns, candidate generation and test methods may still suffer from generating huge numbers of candidates and taking many scans of large data bases for frequency checking.
Recently, other category methods, pattern-growth methods, such as FP growth [11] and Trade projection [3] have been proposed. A pattern – growth methods uses the Apriori property. However, instead of generating candidate sets, it recursively partitions the data base in to sub-data bases according to the frequent patterns found and searches for local frequent patterns to assemble longer global ones. In [10] shows that H-Mine has high performance in various kinds of data, out performs the previously developed algorithms in different settings, and is highly scalable in mining large data bases. In [19] developed an algorithm termed – A
Dynamic approach for frequent patterns mining using Transposition of Data base for mining frequent patterns which are based on Apriori algorithm and used Dynamic function for longest common subsequence. In [7] output from hidden Markov models into the association rule mining frame work, demonstrating the potential for frequent pattern mining in the field of scientific modeling and experimentation. New probabilistic formulations of frequent item set based on possible world semantics are introduced in [21]. In [17] a statistical approach to the success probability for the simple shopping model where every item has the same probability and all the items and all the transactions are independent is given. A class of models called Item Set Generating models (IGM) that can be used to formally connect the process of frequent item sets to discover with the learning of generative models are presented in [20].

FREQUENT PATTERN MINING
Definition: Let I = $\{x_1,x_2 ...... x_n\}$ be a set of items. An item set X is a subset of items, ie, $X \subseteq I$. For the sake of brevity, an item set X = $\{x_1, x_2 .... x_m\}$ is also denoted as X = $x_1\, x_2 ... x_m$. A transaction T = (tid, X) is a 2-tuple, where tid is a transaction-id and X is an item set. A transaction T = (tid, X) is said to contain item set Y if and only if $Y \subseteq X$. A transaction database TDB is a set of transactions. The number of transactions in TDB containing item set X is called the support of X, denoted as

sup(X). Given a transaction database TDB and a support threshold min-sup, an item set X is a frequent pattern, or a pattern in short, if and only if sup (X) $\geq$ min-sup.

The problem of frequent pattern mining is to find the complete set of frequent patterns in a given transaction database with respect to a given support threshold.

DYNAMIC PROGRAMMING

The method of dynamic programming (DP) was developed in 1950's through the work of Richard Bellman who is still the doyen of research workers in this field. The essential feature of the method is that a multivariable optimization problem is decomposed into a series of stages, optimization being done at each stage with respect to one variable only. Richard Bellman gave it rather the undescriptive name of dynamic programming. A more significant name would be recursive optimization.

To case a verbal problem into a multistage structure is not always simple: often it is very difficult and even looks mysterious. But once done, recursive optimization is easy to apply. Recursion equations are a standard type and the related computer programming is a standard routine.

FREQUENT ITEM MINING BY DYNAMIC PROGRAMMING

Let I = {x_1,x_2 ...... x_n} be a set of items. Let T = {T_1, T_2 .....T_k} be a set of data base of customer transactions. Each transaction T_i is a collection of items purchased by a customer in one visit to a store. A non-empty set of items is called an item set. Our aim is to find the items whose frequency exceeds a user defined threshold referred to as a frequent item set.

Let the number of an item in the whole transaction is greater than or equal to 'p' then it is considered as frequent item. First we search for a single item which occurs as frequent item. From that we will get sets of frequent items containing one element. After that we check for frequent item sets with two element and then check for frequent items set with three elements. Continue the process so as to get frequent item sets with maximum number of elements. We use D.P method by dividing the problems into stages and find the optimal solution for each stage. Each stage consists of some state variables and return functions.

	x_1	x_2	....	....	x_n
T_1	$p_1(x_1)$	$p_1(x_2)$	....	....	$P_1(x_n)$
T_2	$p_2(x_1)$	$P_2(x_2)$	....	....	$P_2(x_n)$
....	....	....	....	....	....
.....	....	....	....	....	....
T_k	$p_k(x_1)$	$p_k(x_2)$	....	....	$p_k(x_n)$

The above table gives the relation between the transaction and items where $p_{k}(x_n)$ is the probability of x_n in the k^{th} transaction. $p_j(x_i)$ = 1 if x_i € T_j and equal to zero other wise .

The first stage is to find all the frequent item sets with single item. Here we have to check which among. x_1 ..x_2 ...... x_n are frequent items.

The return function for the variable x_i is

$$F_1(x_i) = \sum_{J=1}^{k} p_j(x_i) \text{ is the return function.}$$

If $\sum_{J=1}^{k} p_j(x_i) \geq p$ then xi is frequent item if it is less than P then it is not a frequent item. So we can remove the item from further stages.

Let I = (x_{11},x_{12} ...... x_{1q}) are the frequent item from stage 1. The Table for stage 1 is given follows:

Item	count in the transaction
x_1	$\sum_{j=1}^{k} p_j(x_1)$
x_2	....
...	....
....	....
....	....
x_n	$\sum_{j=1}^{k} p_j(x_n)$

Now we enter the second stage with variables from the first stage as frequent. Here our attempt is to find frequent item sets with two elements.

Here the return function for the variables (x_{1i}, x_{1k}) is

$$F_2(x_{1i}, x_{1k}) = \sum_{j=1}^{k} p_j(x_{1i})\, p_j(x_{1k} / x_{1i})$$

where x_{1i} and x_{1k} element of {x_{11}, x_{12}.............. x_{1q}} and $p_j(x_{1k} / x_{1i})$ = 1 if (x_{1i}, x_{1k}) € T_j

= 0,other wise

if $\sum_{j=1}^{k} p_j(x_{1i})\, p_j(x_{1k} / x_{1i}) \geq p$ then (x_{1i}, x_{1k}) is a frequent pattern or associate items

Table for second stage

	x_{12}	....	....	x_{1q}
x_{11}	$\sum_{j=1}^{k} p_j\ (x_{11})p_j\ (x_{12}/\ x_{11})$	....	....	....
x_{12}	______	....	....	....
....	....	....	....	....
....		....	....	....
	.....	....		
....	....	....	....	....
	....	....	....	.....
x_{1q}	$\sum_{j=1}^{k} p_j\ (x_{1q})p_j\ (x_{12}/\ x_{11})$	....	......	....

From the second stage give frequent item set of two items. In the third stage examines for frequent item set of three elements. The return function

$$\sum_{j=1}^{k} p_j\ (x_{2q})p_j\ (x_{2r,}x_{2s}/x_{2q})\ \_\_\_\_\ A$$

Where $p_j\ (x_{2r,}x_{2s}/x_{2q}) = 1$ if $x_{2r,}x_{2s,}x_{2q}$ € T_j and zero other wise . $x_{2r,}x_{2s,}x_{2q}$ is a frequent item set if $A \geq p$.

Continued for getting frequent item set containing maximum number of items which satisfies the minimum threshold.

EXAMPLE

Following is a transactional data given the items in each transaction.

T_iD	List of items
T_1	11, 12, 15
T_2	12, 14
T_3	12 , 13
T_4	11, 12, 14
T_5	11, 13
T_6	12, 13
T_7	11, 13
T_8	11, 12, 13, 15
T_9	11, 12, 13

Item I = {11, 12, 13, 14, 15}

Let the number of item in the whole transaction is ≥ 2 then it is a frequent item.
The table showing the relation between items and transactions are follows.

	11	12	13	14	15
T_1	1	1			1
T_2		1		1	
T_3		1	1		
T_4	1	1		1	
T_5	1		1		
T_6		1	1		
T_7	1		1		
T_8	1	1	1		1
T_9	1	1	1		

For stage 1 we search for single set frequent items.

Items	$\sum_{j=1}^{k} pj(x_i)$
11	6
12	7
13	6
14	2
15	2

Here all items satisfy the minimum threshold. Therefore the frequent item from stage 1 is 11, 12, 13, 14 and 15. The second stage is to find the frequent item sets with two elements.

	11	12	13	14	15
11	-	4	4	1	2
12	4	-	4	2	2
13	4	4	-	0	1
14	1	2	0	-	0
15	2	2	1	0	-

Here the item set satisfy the condition $\sum_{j=1}^{k} p_j (x_{1i}) p_j (x_{1k} / x_{1i}) \geq 2$ are

{{11, 12} {11, 13} {11,15}{12,13} {12, 14) {12, 15}.}

In the third stage is for finding frequent item set with three elements.

	11	12	13	14	15
{11, 12}	-	-	2	1	2
{11, 13}	-	2	-	0	1
{11, 15}	-	2	1	0	-
{12, 13)	2	-	-	0	1
{12, 14}	1	-	0	-	0
{12, 15)	2	-	1	0	-

Here the item set satisfies the condition $\sum_{J=1}^{k} p_j (x_{2q}) p_j (x_{2r}, x_{2s}/x_{2q}) \geq 2$ are {11, 12, 13} and

{11, 12, 15}.

Now for stage four is to find frequent item sets containing four elements.

	11	12	13	14	15
{11, 12,13}	-	-	-	0	1
{11, 12, 15}	-	-	1	0	-

Here no item sets satisfies the minimum threshold condition. Therefore there does not exist frequent set containing four elements. Therefore maximum number of elements in the frequent item set is three and they are {11, 12, 13}, {11, 12, 15).

CONCLUSION

Determining frequent item sets is one of the most important field of mining. It is well known that the way of candidates are defined has great effect on running time and memory need and this is the reason for the large number of algorithms. Here presented a new research trend on frequent mining using dynamic programming. Dynamic programming is a simplest method in operation research. The concept of DP has been extended to infinitely multi stage cases also.

REFERENCES

Agrwal.R, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIG MOD Conference on Management of Data*, pages 207-216, May 1993.

Agarwal.R and R. Srikant. Fast algorithms for mining association rules. In VLDB 94, pages 487-499.

Agarwal.R, C.Aggaral and V.V.V. Prasad. A tree projection algorithm for generation of frequent item sets. In *J of Parallel and Distributed Computing,* volume: 61, issue: 3, March 2001, pages: 350-371.

Agrawal.R, J. Gehrike, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *SIG MOD 98* pages 94-105.

Bayardo.R.J, Efficiently mining long patterns from databases. In *SIGMOD 98, pages 85-93.*

B. Liu, W. Hsu and Y. Ma. Integrating classification and association rule mining. In *KDD 98*, pages 80-86.

Christopher Besemann and Anne Denton. Integration of profile hidden Marko model output into association rule mining.

H. Mannila, H. Toivonen, and A.I. Verkamo. Efficient algorithms for discovering association rules. In *KDD 94*, pages 181-192.

Jiawei Han and Michelin Kamber, Second edition. Data mining concept and techniques.

Jian Pei, Jiawei Han, Hongjum Lu, Shojiro Nishio, Shiwei Tang, Dongqing Yang. H-Mine: Hyper-structure mining of frequent patterns in large data base.

J. Han, J. Pei and Y. Yiri. Mining frequent patterns without candidate generation. In *SIGMOD 00, paged 1-12.*

J. Pei, J. Han and R. Mao. CLOSET: An efficient algorithm for mining frequent closed item sets. In *Proc. 2000 ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery (DMKD 00)*, pages 11-20.

J. Pesi, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal , and M.C. Hsu. Prefix-projected pattern growth. In JCDE 01, pages 215-224.

K.V. Miital, Optimization methods in operation research and system analysis.

K. Wing, S. Zhou and S.C. Liew. Building hierarchical classifiers using class proximity. In *VLDB 99*, pages 363-374.

M. Zaki. Generating non-redundant association rules. In *KDD 00*, paged 34-43.

Nele Dexters, University of Antwerp, Middelheimlaan, 1, 2020 Antwerp, Belgium. Approximating the probability of an item set being frequent

R.Srikant and R.Agrwal. Mining sequential patterns: Generalizations and performance improvements. In *EDBT 96*, paged 3-17.

Sunil Joshi, Dr. R.S. Jadon, Dr. R.C. Jain, An implementation of frequent pattern mining algorithm using dynamic function.

Srivastan Laxman, Prasad, Naldurg, Raja Sripada Microsoft Research Labs

Ramarathnam Venkatesan. Connections between mining frequent item sets and learning generative models

Thomas Bernecker, Hans-Kriegel, Matthias Renz, Florian Verhein, Andreas Zuefle.

Probabilistic frequent item set mining uncertain databases

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Commerce, IT and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mails i.e. **infoijrcm@gmail.com** or **info@ijrcm.org.in** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator