© 2018 IJSRST | Volume 4 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

A Review on Pattern Mining Research Issues

Krishna Chaitanya Sanagavarapu
Associate Network Engineer, Telecom Gateway, Frisco, TX

ABSTRACT

A common guide on model extraction investigation. Almost these research issues primarily handle three pattern extraction facets such as the types of models to be detected, mining strategies, and areas. Few studies, yet, incorporate many features; for instance, distinct fields can be required for mining distinct models, which obviously lead to the growth of novel detection methodologies. This paper provides a study of basic algorithms and research issues and also about real life applications.

Keywords : Research Issues, Data Mining, Real Applications

I. INTRODUCTION

BASIC ALGORITHMS AND EXTENSIONS

The First algorithm is developed by Agrawal and Srikanth (1993) for market analysis is the way of relationship rule extraction by using algorithm Apriori, which requires five types of issues:

(i) Proficient and scalable methods to extract regular patterns
(ii) To detect focused regular patterns
(iii) Effect to information analysis and applications of mining
(iv) Regular pattern mining applications
(v) Research Information

(i) Proficient and Scalable Techniques to Extract Regular or Frequent Models: All the approaches under this category are Apriori (Agrawal and Srikant), Frequent Pattern Growth (J.Han) and ECLAT algorithms (Mohammed. Z).

II. METHODS AND MATERIAL

Illustration of Apriori Algorithm

Let a transactional database consists of four transactions with associated Item-sets. Let min_sup = 2 then there exist only one frequent Item-set with the size of three i.e, {b, c, e}. The step-by-step procedure for frequent pattern mining is as follows:

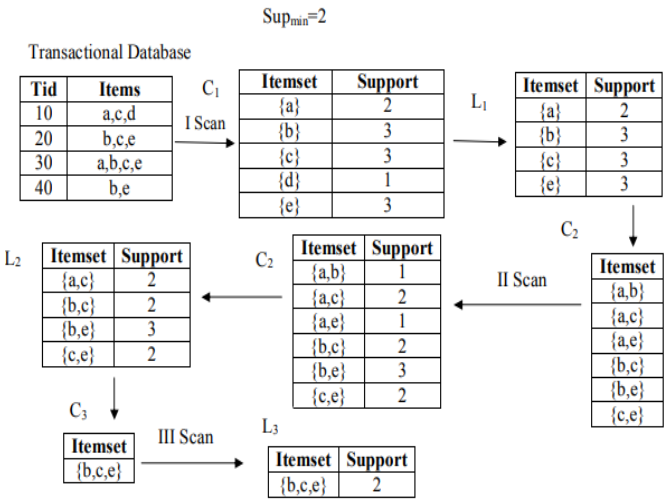

Figure 1

Comments on Apriori Algorithm

One of the major properties of this technique is that, a j-item set is regular only, iff all its subsequent item

sets are regular and a non-frequent item set and its superset cannot be regular.

Advanced Techniques over Apriori Algorithm

The extensions of this method are sampling (1996), hashing strategy (1995), Separating (1995), decreasing transactions (1995), incremental mining (1996), vigorous Item-set totaling (1997), parallel and distributed mining (1997) techniques.

Mining Frequent

Item-sets without Candidate Generation The algorithms considered under this category are Frequent Pattern Growth (2000), H-Mine (2001) and Prefix Tree Structure (2003). The input transactional dataset for Frequent Pattern Growth algorithm is given below

Transaction_id	List of Item_Ids
T100	I_1,I_2,I_5
T200	I_2,I_4
T300	I_2,I_3
T400	I_1,I_2,I_4
T500	I_1,I_3
T600	I_2,I_3
T700	I_1,I_3
T800	I_1,I_2,I_3,I_5
T900	I_1,I_2,I_3

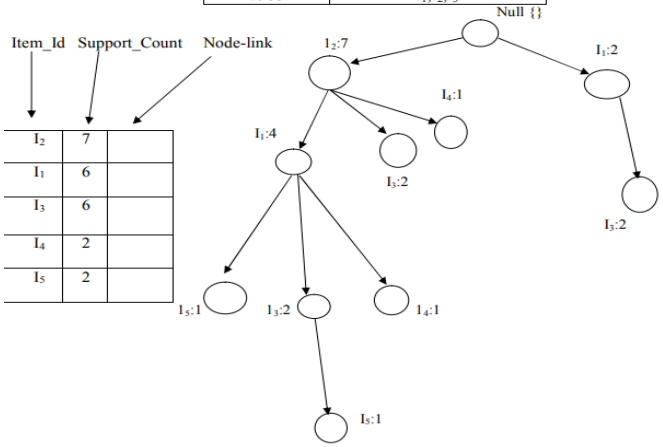

Figure 2

Item	Conditional Pattern Base	Conditional FP-Tree	Frequent Patterns Generated
I_5	$\{\{I_2,I_1:1\},\{I_2,I_1,I_3,:1\}\}$	$\langle I_2:2,I_1:2\rangle$	$\{I_2,I_5:2\},\{I_1,I_5:2\},\{I_2,I_1,I_5:2\}$
I_4	$\{\{I_2,I_1:1\},\{I_2:1\}\}$	$\langle I_2:2\rangle$	$\{I_2,I_4:2\}$
I_3	$\{\{I_2,I_1:2\},\{I_2:2\},\{I_1:2\}\}$	$\langle I_2:4, I_1:2\rangle, \langle I_1:2\rangle$	$\{I_2,I_3:4\},\{I_1,I_3:4\},\{I_2,I_1,I_3:2\}$
I_1	$\{\{I_2:4\}\}$	$\langle I_2:4\rangle$	$\{I_2,I_1:4\}$

Figure 3

There are two types of data on which association rule analysis is performed such as horizontal and vertical data format. Apriori algorithm can be applied on horizontal form of data. There are also many techniques to extract regular item sets based on vertical information form such as Equivalence Class Transformation (ECLAT) algorithm, proposed by Mohammed. Z (2000) to find both short and long frequent item sets quickly The advantage is that no need to discover the support of (j+1) item sets repeatedly.

ECLAT Algorithm Example:

Itemset	TID_set
$\{I_1,I_2\}$	$\{T_{100},T_{400},T_{800},T_{900}\}$
$\{I_1,I_3\}$	$\{T_{500},T_{700},T_{800},T_{900}\}$
$\{I_1,I_4\}$	$\{T_{400}\}$
$\{I_1,I_5\}$	$\{T_{100},T_{800}\}$
$\{I_2,I_3\}$	$\{T_{300},T_{600},T_{800},T_{900}\}$
$\{I_2,I_4\}$	$\{T_{200},T_{400}\}$
$\{I_2,I_5\}$	$\{T_{100},T_{800}\}$
$\{I_3,I_5\}$	$\{T_{800}\}$

Itemset	TID_set
$\{I_1,I_2,I_3\}$	$\{T_{800},T_{900}\}$
$\{I_1,I_2,I_5\}$	$\{T_{100},T_{800}\}$

Figure 4

Mining Multilevel, Multi-Dimensional and Quantitative Relationship Rules

There are many concepts to mine relationship rules by various levels of generalization (1996) and the multi-dimensional relationship rules engage greater than single attribute or predicate. Association Rule Clustering System uses quantitative association rules (1997).

Motivation for Closed and Maximal Patterns

A huge model will encompass an exponential quantity of slighter, regular sub patterns. If min_sup is low then huge frequent item sets will be generated so that there is a need to search for closed regular models to reduce the search space.

A model 'Y' is a closed regular pattern in a database 'E', if 'Y' is regular in 'E' and there is no proper super pattern 'Z' such that 'Z' has the similar support as 'Y' in 'E'. i.e., C = {Y | Y ∈ F and no Z ⊃ Y such that sup (Y) = sup (Z) }.

A pattern 'Y' is closed if all supersets of 'Y' have exactly fewer support, i.e., support (Y) > support (Z), for all Z ⊃ Y. A model 'Y' is a maximal model, if 'Y' is regular, and there is no super-pattern 'Z' such that Y ⊂ Z and 'Z' is regular in 'E'. i.e., W= {Y | Y ∈ F and no Z ⊃ Y, such that Z ∈ F}.

The accompanying relationship comprises between the collection of all congested and effective regular item sets i.e., W⊆ C ⊆ F. Consider the table 1.1, as the source transactions with items, then the nineteen regular item sets demonstrated in the table constitute the collection C. The set of all regular j-item sets are:

C (1) = {X, Y, Z, G, H},
C (2) = {XG, ZH, GH, XY, XH, YZ, YG, YH},
C (3) = { XYG, XGH, YZH, YGH, XYH } and
C (4) = {XYGH}.

There are exclusively 2 maximal regular item sets such as XYGH and YZH with min_sup =3.

The various algorithms developed for closed pattern mining are A-Close (1999), Closet (2000), CHARM (2002), Closet+ (2003), FP-Close (2003), AFOPT (2003). The foremost strategy to extract maximal model detection is Max-miner projected by Bayardo (1998) and it is Apriori-based, level-wise, BFS method. The key property is to discover max item set superset occurrence & subset infrequency reducing technique. The second algorithm developed is MAFIA, introduced by Burdick (2001) that utilizes vertical bitmaps to reduce the number of transaction lists. It improves the totaling proficiency.

Yang (2004) proposed a theoretical study of the worst-case complexity to extract maximal models. Numerating effective item sets complexity is depicted to be NP-hard. The length dispersion of frequent and maximal regular item set grouping is described by Ramesh Agrawal et al. (2003). The other algorithms also developed to mine multidimensional, enormous models are CARPENTER (2003), COBBLER (2004), and TD-Close (2006).

Sequential model extraction is also a major activity in knowledge mining. The significant algorithms designed for this category are generalized sequential patterns (1996), SPADE (2001), Concept – BFS and Apriori Pruning, CloSpan (2003), BIDE (2004), Acyclic Graphs of events (1997), Regular Expressions specifications (1998), multilevel and multidimensional series model detection (2001), CLUSEQ: a series grouping strategy (2003), IncSpan: an Incremental series model extraction strategy (2004), SeqIndex: sequence indexing (2004), Parallel detection of congested series models (2005), MSPX: Maximal series models by utilizing many samples (2005).

Periodicity analysis also an important activity in traffic risk analysis application. The various algorithms developed for this issue for example to mine periodic relationship rules (1998), Inter-transaction association rules (1998), Mining partial periodicity (1999), and partially periodic mining the event patterns with hidden periods (2001), Detecting asynchronous periodic models in time sequence data (2003), Discovering regular closed limited orders from string series (2005). The other types of patterns are mining structural patterns and Mining interesting frequent patterns (constraint, compressed). Apriori (AGM, FCS), Pattern growth based (gSpan, MoFA, SPIN, TreeMinerV, FREQT) are examples of first category and ExAnte (2003), TFP: Top-k most

frequent closed patterns (2005) are the sample algorithms for second category.

There are many measures used to mine interesting and correlation analysis such as lift, $\chi 2$ (1997), cosine and all confidence (2002), Unexpectedness (2003), Relatedness (2004) and User's feedback (2006).

III. RESULTS AND DISCUSSION

REAL-LIFE APPLICATIONS

i. Historic Examples: Fluoride and healthy gums near Colorado river
ii. Modern Examples
   a. Cancer clusters to investigate environment health hazards.
   b. Crime hotspots for planning police patrol routes
   c. Find the quick paths between two specified locations.
      i. The set of main web pages viewed by users, PCs and discounts also view the online shopping and checkout web pages.
      ii. Inference: The particular discount proposal leads to additional purchase of PCs.
      iii. The Consumers who purchase Milk and Sugar besides purchase
      iv. Bread Inference: A shopping mall co-locates bread and Milk in the same floor.

IV. CONCLUSION

This work is employed to find regular and maximal cyclic models proficiently by using Enhanced Tree Mining Strategy and Extensive Regular Model Extraction Algorithm from spatiotemporal datasets for changed instances. This work can be extended to real time applications such as weather forecasting, forest fire prevention and biological data analysis. This paper provided a study of basic algorithms and research issues and also about real life applications

V. REFERENCES

[1]. Sriramoju Ajay Babu, Dr. S. Shoban Babu, "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications", Volume 1, Issue 1, Jan-Mar 2014 [ISSN : 2349-0020]

[2]. Dr. Shoban Babu Sriramoju, Ramesh Gadde, "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol 1, Issue 1,Jan-Mar 2014 [ISSN : 2349-0020].

[3]. Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju, "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2014 [ISSN : 2249-4510]

[4]. Shoban Babu Sriramoju, " Review on Big Data and Mining Algorithm" in "International Journal for Research in Applied Science and Engineering Technology", Volume-5, Issue-XI, November 2017, 1238-1243 [ISSN : 2321-9653], www.ijraset.com

[5]. Guguloth Vijaya, A. Devaki, Dr. Shoban Babu Sriramoju, "A Framework for Solving Identity Disclosure Problem in Collaborative Data Publishing" in "International Journal of Research and Applications", Volume 2, Issue 6, 292-295, Apr-Jun 2016 [ISSN : 2349-0020]

[6]. Shoban Babu Sriramoju, "A Framework for Keyword Based Query and Response System for Web Based Expert Search" in "International Journal of Science and Research" Index

Copernicus Value(2015):78.96 [ISSN : 2319-7064].

[7]. Sriramoju Ajay Babu, Dr. S. Shoban Babu, "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications", Volume 1, Issue 1, Jan-Mar 2014 [ISSN : 2349-0020]

[8]. Dr. Shoban Babu Sriramoju, Ramesh Gadde, "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol 1, Issue 1,Jan-Mar 2014 [ISSN : 2349-0020].

[9]. Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju, "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2014 [ISSN : 2249-4510]

[10]. Shoban Babu Sriramoju, "An Application for Annotating Web Search Results" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[11]. Shoban Babu Sriramoju, "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2, Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[12]. Mounica Doosetty, Keerthi Kodakandla, Ashok R, Shoban Babu Sriramoju, "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2012 [ISSN : 2249-4510]

[13]. Shoban Babu Sriramoju, Madan Kumar Chandran, "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering", Vol 4, Issue 2, February 2014 [ISSN : 2277 128X]

[14]. Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao, "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol V, Issue II,Febrauary 2014 [ISSN : 2321-3469]

[15]. Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Analysis on Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and Management", Vol VII, Issue X, November 2014 [ISSN : 2249-4510]

[16]. Shoban Babu Sriramoju, "Mining Big Sources Using Efficient Data Mining Algorithms" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2, Issue 1,January 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[17]. Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays for Image Processing" in "International Journal of Information Technology and Management" Vol V, Issue I, August 2013 [ISSN : 2249-4510]

[18]. Dr. Shoban Babu Sriramoju, "A Review on Processing Big Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-1, January 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[19]. Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis around the study of Distributed Data Mining Method in the Grid Environment :

Technique, Algorithms and Services" in "Journal of Advances in Science and Technology" Vol-IV, Issue No-VII, November 2012 [ISSN : 2230-9659]

[20]. Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis on Effective, Precise and Privacy Preserving Data Mining Association Rules with Partitioning on Distributed Databases" in "International Journal of Information Technology and management" Vol-III, Issue-I, August 2012 [ISSN : 2249-4510]

[21]. Shoban Babu Sriramoju, Dr. Atul Kumar, "A Competent Strategy Regarding Relationship of Rule Mining on Distributed Database Algorithm" in "Journal of Advances in Science and Technology" Vol-II, Issue No-II, November 2011 [ISSN : 2230-9659]

[22]. Shoban Babu Sriramoju, Dr. Atul Kumar, "Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts" in "International Journal of Information Technology and management" Vol-I, Issue-I, August 2011 [ISSN : 2249-4510]

[23]. Monelli Ayyavaraiah, "Review of Machine Learning based Sentiment Analysis on Social Web Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 4, Issue 6,March 2016 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[24]. Anusha Medavaka, P. Shireesha, "Analysis and Usage of Spam Detection Method in Mail Filtering System" in "International Journal of Information Technology and Management", Vol. 12, Issue No. 1, February-2017 [ISSN : 2249-4510]

[25]. Anusha Medavaka, P. Shireesha, "Review on Secure Routing Protocols in MANETs" in "International Journal of Information Technology and Management", Vol. VIII, Issue No. XII, May-2015 [ISSN : 2249-4510]

[26]. Anusha Medavaka, P. Shireesha, "Classification Techniques for Improving Efficiency and Effectiveness of Hierarchical Clustering for the Given Data Set" in "International Journal of Information Technology and Management", Vol. X, Issue No. XV, May-2016 [ISSN : 2249-4510]

[27]. Anusha Medavaka , P. Shireesha, "Optimal framework to Wireless Rechargeable Sensor Network based Joint Spatial of the Mobile Node" in "Journal of Advances in Science and Technology", Vol. XI, Issue No. XXII, May-2016 [ISSN : 2230-9659]

[28]. Anusha Medavaka,"Enhanced Classification Framework on Social Networks" in "Journal of Advances in Science and Technology", Vol. IX, Issue No. XIX, May-2015 [ISSN : 2230-9659]

[29]. Anusha Medavaka, Dr. P. Niranjan, P. Shireesha, "USER SPECIFIC SEARCH HISTORIES AND ORGANIZING PROBLEMS" in "International Journal of Advanced Computer Technology (IJACT)", Vol. 3, Issue No. 6, 2015 [ISSN : 2319-7900]

[30]. Yeshwanth Rao Bhandayker , "Artificial Intelligence and Big Data for Computer Cyber Security Systems" in "Journal of Advances in Science and Technology", Vol. 12, Issue No. 24, November-2016 [ISSN : 2230-9659]

[31]. Sugandhi Maheshwaram, "A Comprehensive Review on the Implementation of Big Data Solutions" in "International Journal of Information Technology and Management", Vol. XI, Issue No. XVII, November-2016 [ISSN : 2249-4510]

[32]. Yeshwanth Rao Bhandayker, "Security Mechanisms for Providing Security to the Network" in "International Journal of Information Technology and Management",

Vol. 12, Issue No. 1, February-2017, [ISSN : 2249-4510]

Cite this article as :

Krishna Chaitanya Sanagavarapu, "A Review on Pattern Mining Research Issues", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 4 Issue 8, pp. 717-723, May-June 2018.
Journal URL : http://ijsrst.com/IJSRST196355