

Usage and Research Challenges in the Area of Frequent Pattern in Data Mining

P. Alagesh Kannan¹, Dr. E. Ramaraj²

¹ Assistant Professor, Department Of Computer Science, MKU College, Madurai, Tamil Nadu, India

² Professor, Department of Computer Science and Engg, Alagappa University, Karaikudi Tamil Nadu, India

Abstract: Frequent pattern mining is an important chore in the data mining, which reduces the complexity of the data mining task. The usages of frequent patterns in various verticals of the data mining functionalities are discussed in this paper. The gap analysis between the requirements and the existing technology is also analyzed. State of art in the area of frequent pattern mining was thrashed out here. Working mechanisms and the usage of frequent patterns in various practices were conversed in the paper. The core area to be concentrated is the minimal representation, contextual analysis and the dynamic identification of the frequent patterns.

Keywords: Frequent pattern, Association, Clustering, Classification

I. Introduction

Frequent patterns are the subset of the given dataset with the occurrence frequency that satisfies the user specified threshold and above [1]. Identification of the frequent patterns is a thrust area in the field of data mining which has the applications in association mining, correlation analysis. It is an important part of finding the interesting relationships within the data given. Other functional areas of data mining such as clustering, classification indexing also use the frequent patterns.

Frequent pattern mining was first proposed by Agrawal et al [2] for market basket analysis in the form of association rule mining. It analyses customer buying habits by finding associations between the different items that customers place in their “shopping baskets”. The concept is explained as

Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of all items. A k-itemset α , which consists of k items from I, is frequent if α occurs in a transaction database D no lower than $\theta|D|$ times, where θ is a user-specified.

In this paper the usage of frequent patterns in the area of finding the association, classification and clustering is discussed thoroughly. A deep review of the state of art techniques was reviewed. In the section 5 the research challenges posed in the field of frequent pattern mining is discussed.

II. Fundamental Approaches for Mining Associations Using Frequent Patterns

The basic concepts which used frequent patterns to mine the association between data are discussed in this section. Apriori, FP-growth and Eclat are the three basic approaches emerged by using frequent patterns.

II.A Algorithms using Apriori approach

Agrawal and Srikant [3] observed an interesting downward closure property, called Apriori, among frequent k item sets: A k-itemset is frequent only if all of its sub-item sets are frequent. This implies that frequent item sets can be mined by first scanning the database to find the frequent 1-itemsets, then using the frequent 1-itemsets to generate candidate frequent 2-itemsets, and check against the database to obtain the frequent 2-itemsets. This process iterates until no more frequent k-item sets can be generated for some k. This also motivated its alternative [4] where the algorithm uses all existing information between database passes to avoid checking the coverage of redundant sets.

It follows a level wise search or the breath first search. The extension of this technique reflected in many of the algorithms with some deviations for efficiency. [5] Used the direct hashing and pruning for efficient large item set generation with effective reduction on transaction database size, [6] used the partitioning technique to read the database at most two times to generate all significant association rules. Toivonen [7] suggested an algorithm which uses random sample to form association and later check with the whole database to avoid approximation. The approach is probabilistic and some time requires a second pass.

Dynamic item set counting and implication rules are used in [8] to reduce the number of passes. An incremental updating technique is proposed for efficient maintenance of discovered association rules when new transaction data are added to a transaction database in [9]. Parallelization process is done for finding the association rules in [10], which employs a clue and poll technique to address the uncertainty due to partial knowledge. In the paper [11] provides a hard and tight combinatorial upper bound to answer the question of the maximal number of candidate patterns that can be generated in the passes.

Even though the Apriori throws new light in the frequent pattern generation by the reduction of the size of the candidate sets, the problem with which it suffers is enormous number of the candidate set generation and the repeated scanning of database, which is costly for the large databases.

II.B Algorithms using Frequent Pattern growth approach

Han et al. [12] presented an approach to generate frequent patterns without candidate set generation, which is a bottleneck in apriori approach. It built a compact frequent pattern tree structure and from that the frequent patterns are extracted by traversing recursively the tree. The pattern growth is achieved by the concatenation of the suffix pattern with the frequent patterns generated from a conditional frequent pattern tree.

This approach tackles the problem by identification of shorter frequent patterns and from that the long patterns are constructed recursively by concatenation. [13] and [14] uses the hyper-structure mining of frequent patterns for building alternative trees. An array-based implementation of prefix-tree-structure for efficient pattern growth mining by Grahne and Zhu is suggested in [15].

This model reduces the irrelevant information by the deletion of the infrequent items and the more frequently occurring, the more likely to be shared. It recursively work for the frequent pattern and it is done in an incremental mode. It doesn't search for the pattern and match instead it count local frequent pattern and then built the tree.

This technique lacks in few conditions like the tree may not fit into the memory, it takes time to build, and when the support threshold is high time is wasted as pruning could be done on single items. Another important problem is support can only be calculated once the entire data-set is added to the tree.

II.C Algorithms using the vertical exploration approach

Zaki [16] proposed Equivalence CLAss Transformation (ECLAT) algorithm by exploring the vertical data format. It uses the depth first search strategy with set intersection. It states that, when the database is stored in the vertical layout, the support of a set can be counted much easier by simply intersecting the covers of two of its subsets that together give the set itself [17].

This algorithm generates the candidate item sets using the join step proposed in the apriori. After this all the items in the database are reordered in ascending. This reduces the number of intersections to be computed. It doesn't fully exploit the monotony property; the number of candidate itemsets that are generated is much larger as compared to a breadth-first approach such as Apriori.

CHARM [18] is an efficient algorithm for enumerating the set of all closed frequent itemsets. CHARM is unique in that it simultaneously explores both the itemset space and transaction space, unlike all previous association mining methods which only exploit the itemset search space. It avoids enumerating all possible subsets of a closed itemset when enumerating the closed frequent sets, which rules out a pure bottom-up search.

III. Use of Frequent Patterns for Classification

The application of frequent patterns in classification appeared in sporadic studies and achieved initial success in the classification of relational data, text documents and graphs [19]. Frequent patterns reflect strong associations between items and carry the underlying semantics of the data. They are potentially useful features for classification. Frequent pattern-based classification could exploit the state-of-the-art frequent pattern mining algorithms for feature generation, thus achieving much better scalability than the method of enumerating all feature combinations.

Frequent pattern based classification could be employed with feature generation, feature selection, and model learning. Frequent patterns form the efficient features since each pattern is a combination of single features and they are frequent. Frequent pattern is a form of non-linear feature combination over the set of single features. With inclusion of non-linear feature combinations, the expressive power of the new feature space increases.

Direct Discriminative Pattern Mining (DDPMine) performs a branch-and bound search for directly mining discriminative patterns without generating the complete pattern set. Instead of selecting best patterns in a batch, a "feature-centered" mining approach is proposed that generates discriminative patterns sequentially on a progressively shrinking FP-tree by incrementally eliminating training instances. The instance elimination effectively reduces the problem size iteratively and expedites the mining process [20].

Discriminative Pattern Mining Approach for the Classification of Software Behaviors for Failure Detection is suggested in [21], which first mines a set of discriminative features capturing repetitive series of events from program execution traces. It then performs feature selection to select the best features for classification. These features are then used to train a classifier to detect failures.

Frequent patterns allow the construction of high-level sets of compound features which can, in many cases, capture more discriminative information [19]. Frequent patterns are used exhaustively in image classification in [22][25], understanding the scene [26], object recognition and object-part recognition [27].

Effective method for using item set mining to discover a new set of mid-level features called Frequent Local Histograms is suggested by Basura Fernando et. Al [28].

PatMat a classifier based on frequent pattern is built with the use of trie data structure [29]. Monowar et.al [30] suggested an approach for fingerprint classification using data mining approach. Initially, it generates a numeric code sequence for each fingerprint image based on the ridge flow patterns. Then for each class, a seed is selected by using a frequent item sets generation technique. These seeds are subsequently used for clustering the fingerprint images.

Jeroen De Knijf describes a classification method for XML data based on frequent attribute trees. From these frequent patterns the emerging patterns are selected, and use these as binary features in a decision tree algorithm [31]. In this method they used k different minimum support values; one for each class label. To determine an appropriate minimum support value for each class, they started with a high support value, and lowered it gradually until a sufficient number of high quality patterns were produced.

A unified framework for mining multiple domain datasets and design an iterative algorithm called HTMiner[32] capture the associations among different kinds of data. HTMiner discovers essential heterogeneous patterns for classification and performs instance elimination. This instance elimination step reduces the problem size progressively by removing training instances which are correctly covered by the discovered essential heterogeneous pattern.

Iyad Batal [33] said the even though some frequent patterns can be important predictors, using all frequent patterns in the classifier is not a good option for the following reasons like large number of patterns, many non predictive patterns and many spurious patterns. It encompasses that the research opening is present in the frequent pattern mining for classification.

IV. Use of Frequent Patterns for Clustering

Clustering is the unsupervised approach for learning where the process is used for finding the data distribution and patterns in the datasets where their class labels are not known. Pattern based clustering does not require a globally defined similarity measure. Different clusters can follow different patterns on different subsets of dimensions. On the other hand, the clusters are not necessary exclusive. That is, an object can appear in more than one cluster. The generality and flexibility of pattern-based clustering may provide interesting and important insights in some applications where conventional clustering methods may meet difficulties.

An agglomerative hierarchical clustering algorithm to find clusters among the periodic item sets was suggested in [34]. Since the fuzzy number is invariant with respect to shifting, they define similarity measure using the variance of fuzzy intervals associated with frequent item sets.

Frequent pattern for text clustering was discussed in [35]. Here they have used the measure of mutual overlap of frequent sets with respect to the sets of supporting documents. They have presented two algorithms for text clustering, FTC which creates flat clustering and HFTC for hierarchical clustering. FTC works in a bottom-up fashion. Starting with an empty set, it continues selecting one more element (one cluster description) from the set of remaining frequent itemsets until the entire document collection is contained in the cover of the set of all chosen frequent itemsets. In each step, FTC selects one of the remaining frequent itemsets which has a cover with minimum overlap with the other cluster candidates, i.e. the cluster candidate which has the smallest entropy overlap (EO) value. The documents covered by the selected frequent itemsets are removed from the collection D, and in the next iteration, the overlap for all remaining cluster candidates is recomputed with respect to the reduced collection. A frequent item-based approach of clustering is promising because it provides a natural way of reducing the large dimensionality of the document vector space.

Pattern preserving clustering [36] was discussed by Hui Xion et. al. They suggest a better cluster interpretation than traditional clustering approaches by considering the patterns found in clusters. Pattern-based clustering algorithms determine clusters based on the similarities of the patterns among objects across the relevant dimensions, instead of the absolute distance values among objects [37]. Pattern-based algorithm is the bi cluster model proposed by Cheng et al. [38], which tries to measure the coherence of the genes and the conditions in a sub matrix of a DNA array. P Cluster [39] was introduced to cluster objects by computing maximal candidate attributes set (MCAS) and maximal candidate objects set (MXOS) iteratively.

The Maple method [40] advances further this idea to find the maximal pattern-based clusters. Redundant clusters are avoided completely by mining only the maximal pattern-based clusters. The idea is to report only those non-redundant pattern-based clusters, and skip their trivial sub-clusters. It conducts a depth-first, progressively refining search to mine maximal pattern-based clusters. By mining maximal pattern-based clusters, the number of clusters can be reduced substantially. Moreover, many unnecessary searches for sub-clusters can be pruned and thus the mining efficiency can be improved dramatically as well.

An expression pattern based bi clustering technique, CoBi, for grouping both positively and negatively regulated genes from microarray expression data was discussed in [41]. Regulation pattern and similarity in degree of fluctuation are accounted for while computing similarity between two genes. Unlike traditional bi

clustering techniques, which use greedy iterative approaches, it uses a Bi-Clust tree that needs single pass over the entire dataset to find a set of biologically relevant bi clusters.

Clustering based on Frequent Word Sequence (CFWS) is projected in [42]. CFWS uses frequent word sequence and K-mismatch for document clustering. The difference between word sequence and word item set is that word sequence considers words' order while word item sets ignores words' order. Document Clustering Based on Maximal Frequent Sequences (CMS) is proposed in the paper [43]. A frequent sequence is maximal if it is not a subsequence of any other frequent sequence. The basic idea of CMS is to use maximal frequent sequences (MFS) of words as features in vector space model (VSM) for document representation and then k-means is employed to group documents into clusters.

Frequent Itemset-based Hierarchical Clustering (FIHC) is proposed by B.Fung and K.Wang[44]. Two kinds of frequent item are defined in FIHC: global frequent item and cluster frequent item. FIHC develops four phases to produce document clusters: finding global frequent itemsets, initial clustering, tree construction, and pruning.

Maximum Capturing (MC) for document clustering is to produce natural and comprehensible document clusters was discussed in [45]. To produce natural clusters to use frequent itemsets for representation and measure similarities of documents based on cooccurrences of frequent itemsets in documents was the notion. To make document clusters comprehensible, most frequent itemsets in a document cluster were assigned as the topic of the cluster. Because documents with largest number of common frequent itemsets were assigned into a same cluster, cluster topics will be the most representative frequent itemsets in the cluster and thus distinguish clusters from each other.

Fuzzy Frequent Itemset-Based Hierarchical Clustering (F^2 IHC) approach, which uses fuzzy association rule mining algorithm to improve the clustering accuracy of Frequent Itemset-Based Hierarchical Clustering (FIHC) method, was discussed in [46]. In this approach, the key terms will be extracted from the document set, and each document is pre-processed into the designated representation for the following mining process. Then, a fuzzy association rule mining algorithm for text is employed to discover a set of highly-related fuzzy frequent itemsets, which contain key terms to be regarded as the labels of the candidate clusters. Finally, these documents will be clustered into a hierarchical cluster tree by referring to these candidate clusters.

In order to tackle the challenges in the high dimensional clustering, sub space clustering is introduced. For subspace clustering, the determination of subspaces possibly containing clusters is a critical and time-consuming process. Jihong Guan et. al [37] discussed a new pattern-based subspace clustering algorithm CPT by using Pattern tree. The key point is that CPT adopts the pattern tree to discovering the subspace by scanning the database only once.

V. Research Challenges in Finding Frequent Patterns

In the previous sections we have discussed about the need, importance and the usage of the frequent patterns in various verticals of data mining functionalities. This section let us discuss about the challenges in the field of research in frequent pattern mining.

The important challenge is the set of frequent patterns derived by most of the current pattern mining methods is too huge for effective usage. There are proposals on reduction of such a huge set, including closed patterns, maximal patterns, approximate patterns, condensed pattern bases, representative patterns, clustered patterns, and discriminative frequent patterns [1]. Much research is still needed to substantially reduce the size of derived pattern sets and enhance the quality of retained patterns.

The real bottleneck of the problem is not at the efficiency but at the usability. Typically, if the minimum support is high, mining may generate only commonsense patterns, however, with a low minimum support, it may generate an explosive number of results. This has severely restricted the usage of frequent-pattern mining [47]. Reducing the number of uninteresting patterns is an active and emerging research area [50]. Mining compressing patterns is NP-Hard and belongs to the class of inapproximable problems. Future work should be concentrated on further improvements to the mining algorithm using ideas from compression but keeping the focus on usefulness for data mining [51].

Unlike mining static databases, mining data streams poses many new challenges. In addition to the one-scan nature, the unbounded memory requirement and the high data arrival rate of data streams, the combinatorial explosion of item sets exacerbates the mining task [48]. The main challenges in the data streams are handling of the continuous flow, modeling changes of mining results over time, Data stream pre-processing, Model overfitting, Data stream mining technology [49].

The main research work on pattern analysis has been focused on pattern composition (e.g., the set of items in item-set patterns) and frequency. A contextual analysis of frequent patterns over the structural information can identify why that particular pattern is frequent. The deep understanding of frequent patterns is essential to improve the interpretability and the usability of frequent patterns [52].

VI. Conclusion

This paper acts as the literature review in the area of the frequent pattern mining. The paper addresses the various techniques used for the frequent pattern mining and the need of it in the various functionalities of the data mining arena. The research challenges and the area to be concentrated are also discussed.

To conclude the area where the frequent pattern mining is to be concentrated is on the dynamic, Contextual and the compressed frequent pattern mining algorithms. The dynamic nature must address the mining concept with respect to the data streams. The contextual part must focus on the area where the pattern's structural component for the understanding semantics. The compression part must spotlight the minimal requirement of the pattern for the representation.

References

- [1] Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan, "Frequent pattern mining: current status and future Directions", *Data Mining and Knowledge Discovery*, Vol. 15 (2007), pp. 55-86.
- [2] Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: *Proceedings of the 1993 ACM-SIGMOD international conference on management of data (SIGMOD'93)*, Washington, DC, pp 207-216.
- [3] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: *Proceedings of the 1994 international conference on very large data bases (VLDB'94)*, Santiago, Chile, pp 487-499
- [4] Mannila H, Toivonen H, Verkamo AI (1994) Efficient algorithms for discovering association rules. In: *Proceeding of the AAAI'94 workshop knowledge discovery in databases (KDD'94)*, Seattle, WA, pp 181-192
- [5] Park JS, Chen MS, Yu PS (1995) An effective hash-based algorithm for mining association rules. In: *Proceeding of the 1995 ACM-SIGMOD international conference on management of data (SIGMOD'95)*, San Jose, CA, pp 175-186
- [6] Savasere A, Omiecinski E, Navathe S (1995) An efficient algorithm for mining association rules in large databases. In: *Proceeding of the 1995 international conference on very large data bases (VLDB'95)*, Zurich, Switzerland, pp 432-443
- [7] Toivonen H (1996) Sampling large databases for association rules. In: *Proceeding of the 1996 international conference on very large data bases (VLDB'96)*, Bombay, India, pp 134-145
- [8] Brin S, Motwani R, Ullman JD, Tsur S (1997) Dynamic itemset counting and implication rules for market basket analysis. In: *Proceeding of the 1997 ACM-SIGMOD international conference on management of data (SIGMOD'97)*, Tucson, AZ, pp 255-264
- [9] Cheung DW, Han J, Ng V, Wong CY (1996) Maintenance of discovered association rules in large an incremental updating technique. In: *Proceeding of the 1996 international conference on data engineering (ICDE'96)*, New Orleans, LA, pp 106-114
- [10] Park JS, Chen MS, Yu PS (1995) Efficient parallel mining for association rules. In: *Proceeding of the 4th international conference on information and knowledge management*, Baltimore, MD, pp 31-36
- [11] Geerts F, Goethals B, Bussche J (2001) A tight upper bound on the number of candidate patterns. In: *Proceeding of the 2001 international conference on data mining (ICDM'01)*, San Jose, CA, pp 155-162
- [12] Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: *Proceeding of the 2000 ACM-SIGMOD international conference on management of data (SIGMOD'00)*, Dallas, TX, pp 1-12
- [13] Agarwal R, Aggarwal CC, Prasad VVV (2001) A tree projection algorithm for generation of frequent itemsets. *J Parallel Distribut Comput* 61:350-371
- [14] Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. In: *Proceeding of the 2001 international conference on data engineering (ICDE'01)*, Heidelberg, Germany, pp 215-224
- [15] Grahne G, Zhu J (2003) Efficiently using prefix-trees in mining frequent itemsets. In: *Proceeding of the ICDM'03 international workshop on frequent itemset mining implementations (FIMI'03)*, Melbourne, FL, pp 123-132
- [16] Zaki MJ (2000) Scalable algorithms for association mining. *IEEE Trans Knowl Data Eng* 12:372-390
- [17] Pramod S. and Vyas (2010), Survey Frequent Itemset Mining, *International Journal of Computer Applications* 1(15):pp 86-91
- [18] Zaki MJ, Hsiao CJ (2002) CHARM: an efficient algorithm for closed itemset mining. In: *Proceeding of the 2002 SIAM international conference on data mining (SDM'02)*, Arlington, VA, pp 457-473
- [19] Hong Cheng, Yan, X., Jiawei Han (2007); Chih-Wei Hsu, Discriminative Frequent Pattern Analysis for Effective Classification, *IEEE 23rd International Conference on Data Engineering* pp 716 - 725
- [20] Hong Cheng, Xifeng Yan, Jiawei Han, Philip S. Yu (2008), Direct Discriminative Pattern Mining for Effective Classification, *IEEE 24th International Conference on Data Engineering*, 169 - 178
- [21] David Lo, Hong Chen, Han, Khoo, Chengnian Sun, Classification of software behaviors for failure detection: a discriminative pattern mining approach, *15th ACM SIGKDD international conference on Knowledge discovery and data mining*, Pages 557-566
- [22] Nowozin, S., Tsuda, K., Uno, T., Kudo, T., Bakir, G.: Weighted substructure mining for image analysis. In: *CVPR*. (2007)
- [23] Yuan, J., Wu, Y., Yang, M.: Discovery of collocation patterns: from visual words to visual phrases. In: *CVPR*. (2007)
- [24] Yuan, J., Yang, M., Wu, Y.: Mining discriminative co-occurrence patterns for visual recognition. In: *CVPR*. (2011) 2777 -2784
- [25] Kim, S., Jin, X., Han, J.: Disiclass: discriminative frequent pattern-based image classification. In: *Tenth Int. Workshop on Multimedia Data Mining*. (2010)
- [26] Yao, B., Fei-Fei, L.: Grouplet: A structured image representation for recognizing human and object interactions. In: *CVPR*. (2010)
- [27] Quack, T., Ferrari, V., Leibe, B., Van Gool, L.: Efficient mining of frequent and distinctive feature configurations. In: *ICCV*. (2007)
- [28] Basura Fernando, Elisa Fromont, Tinne Tuytelaars, Effective Use of Frequent Itemset Mining for Image Classification, *Lecture Notes in Computer Science Volume 7572*, 2012, pp 214-227 in *Computer Vision - ECCV 2012* (2012)
- [29] Wim Pijls, Rob Potharst, classification and target group selection based on frequent patterns, *ERS -2000-40-LS*, October 2000.
- [30] Monowar H. Bhuyan, Sarat Saharia, Dhruva Kr Bhattacharyya, An Effective Method for Fingerprint Classification, *International Arab Journal of e-Technology*, 2010, 1(3): 89 - 97.
- [31] Jeroen De Knijf, FAT-CAT: Frequent Attributes Tree Based Classification, *Comparative Evaluation of XML Information Retrieval Systems*, *Lecture Notes in Computer Science Volume 4518*, 2007, pp 485-496
- [32] Dhaval Patel, Wynne Hsu, Mong Li Lee, "Integrating Frequent Pattern Mining from Multiple Data Domains for Classification," *icde*, pp.1001-1012, 2012 *IEEE 28th International Conference on Data Engineering*, 2012

- [33] Batal, I. and Hauskrecht, M. (2010), Constructing Classification Features using Minimal Predictive Patterns. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10, pages 869–878, New York, NY, USA. ACM.
- [34] Fokrul Alom Mazarbhuiya, Muhammad Abulaish, Clustering periodic frequent patterns using fuzzy Statistical parameters, International Journal of Innovative Computing, Information and Control ISSN 1349-4198 ,Volume 8, Number 3(B), March 2012.
- [35] Beil F, EsterM, Xu X (2002) Frequent term-based text clustering. In: Proceeding of the 2002 ACM SIGKDD international conference on knowledge discovery in databases (KDD'02), Edmonton, Canada, pp 436–442
- [36] Hui Xiong, Michael Steinbach, Arifin Ruslim, Vipin Kumar, “Characterizing pattern preserving clustering”, Knowledge and Information Systems, June 2009, Volume 19, Issue 3, pp 311-336
- [37] Jihong Guan, Yanglan Gan, Hao Wang, Discovering pattern-based subspace clusters by pattern tree, Knowledge-Based Systems, Volume 22, Issue 8, December 2009.
- [38] Y.Z. Cheng, M. George, Biclustering of expression data, Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, 2000, pp. 93–103.
- [39] H.X. Wang, W. Wang, J. Yang, S. Yu. Philip, clustering by pattern similarity in large data sets, Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, 2002.
- [40] Xiaoling Zhang; Moonjung Cho; Haixun Wang; Yu, P.S, MaPle: a fast algorithm for maximal pattern-based clustering, ICDM 2003. Third IEEE International Conference on Data Mining, 2003.
- [41] Swarup Roy, Dhruva K.Bhattacharyya, Jugal K.Kalita, CoBi:Pattern Based Co-Regulated Biclustering of Gene Expression Data, Pattern Recognition Letters, March 2013
- [42] Y.J. Li, S.M. Chung, J.D. Holt, Text document clustering based on frequent word meaning sequences, Data & Knowledge Engineering 64 (2008) 381–404.
- [43] H. Edith, A.G. Rene, J.A. Carrasco-Ochoa, J.F. Martinez-Trinidad, Document clustering based on maximal frequent sequences, in: Proceedings of the FinTAL 2006, LNAI, vol. 4139, 2006, pp. 257–267.
- [44] B. Fung, K. Wang, M. Ester, Hierarchical document clustering using frequent itemsets, in: Proceedings of the 3rd SIAM International Conference on Data Mining, 2003.
- [45] Wen Zhang , Taketoshi Yoshida, Xijin Tang , Qing Wanga, Text clustering using frequent itemsets, Knowledge-Based Systems 23 (2010) 379–388.
- [46] Chun-Ling Chen , Frank S.C. Tseng b, Tyne Liang, Mining fuzzy frequent itemsets for hierarchical document clustering, Information Processing and Management 46 (2010) 193–211
- [47] Dong Xin Jiawei Han Xifeng Yan Hong Cheng, Mining Compressed Frequent-Pattern Sets, Proceedings of the 31st international conference on Very large data bases, Pages 709 – 720, 2005.
- [48] James Cheng Yiping Ke Wilfred Ng, A survey on algorithms for mining frequent itemsets over data streams, Knowledge and Information Systems, Volume 16 Issue 1, July 2008.
- [49] Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy, Mining Data Streams: A Review, ACM SIGMOD Record Volume 34 Issue 2, June 2005 ,Pages 18 - 26
- [50] Jiawei Han. Mining Useful Patterns: My Evolutionary View. Keynote talk at the Mining Useful Patterns , workshop KDD 2010
- [51] Hoang Thanh Lam, Fabian Moerchen, Dimitry Fradkin, and Toon Calders. Mining compressing sequential patterns. In: SIAM Data Mining Conference 2012
- [52] Qiaozhu Mei, Dong Xin, Hong Cheng, Jiawei Han, ChengXiang Zhai, Generating Semantic Annotations for Frequent Patterns with Context Analysis, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining Pages 337-346 , 2006.