International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319 – 6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 9; September 2017; Page No. 5811-5814
DOI: http://dx.doi.org/10.24327/ijcar.2017.5814.0809

Research Article

GENERATING TEMPORAL ASSOCIATION RULES FOR INFREQUENT ITEMS IN CONCEPT LEVEL VIDEO DATASETS

K.Kumar and P.Sudhakar

PRIST Univeristy, Tamil Nadu, INDIA
Department of CSE, Annamalai University, Tamil Nadu, INDIA

A R T I C L E I N F O

Article History:

Received 19th June, 2017
Received in revised form 3rd July, 2017 Accepted 18th August, 2017
Published online 28th September, 2017

Key words:

Association, mining, temporal, itemset

A B S T R A C T

Due to the increasing rate of video data over the World Wide Web and smart phones, it is becoming very essential to extract useful information from visual data. The video data consists lot of objects which in turn contain important visual information. Video data also contains an amount of direct as well as indirect or hidden information about its objects. The users can access direct information from video by viewing it. To access hidden information from the video mining techniques such as classifications, clustering, regression, outlier detection and association rules etc, can be applied. We focus on concept level video associations and propose an algorithm to generate temporal association rule for the same.

INTRODUCTION

Association rule mining among videos has been extensively studied in data mining research. The main goal of association rule mining[1] is to discover relationships among set of items in a transactional database. An association rule can be interpreted as “if itemset A occurs in a transaction, then itemset B will also likely occur in the same transaction”.

However, in recent years, there has been an increasing demand for mining the videos and infrequent items within it. Since exploring interesting relationship among infrequent items has not been discussed much in the literature, in this paper, we propose practical and effective scheme to mine infrequent items[5] in concept level video dataset. Our algorithm can also be applied to frequent items with bounded length. Our schemes compare favorably to Apriori[1] and FPgrowth[2] under the situation being evaluated.

Review of Literature

Agrawal and Imielinski discussed mining sequential patterns in [3], as well as mining quantitative association rules in large relational tables in [4], while Bayardo considered efficiently mining long patterns from a database in [5] and Dong and Li studied efficient mining of emerging patterns in [6]. On the other hand, Kamber et al. [7] proposed using data cubes to mine multi-dimensional association rules and Lent et al. used the clustering method in [8]. While most researchers focus on association analysis of rules [9–14], Brin et al. analyzed the correlations of association rules in [15]. With the development of data mining techniques, quite a few researchers have worked on alternative patterns; for example, Padmanabhan et al. discussed unexpected patterns in [16], Liu et al. and Hwang et al. studied exception patterns in [17–19], and Savasere et al., Wu et al. and Yuan et al. discussed negative association in [20–22] respectively. J. Ding discussed association rule mining among rare items in [26]. He designed a new disk-based data structure, called Transactional Co-Occurrence Matrix (TCOM) to store the data information. This structure combines the advantages of transactional oriented (horizontal) layout and item oriented (vertical) layout of the database. So any itemsets could be randomly accessed and counted without full scan of the original database or the TCOM. He also constructed a compressed matrix structures, called Reduced Transactional Co-Occurrence Matrix (RTCOM) which reside in the memory. This matrix only contains the items, which are of interest in application. Then the infrequent patterns and the valid association rules among infrequent items can be mined out. Although this is a significant theoretical advancement in the subject, it is quite costly to implement this algorithm.

Video mining Stages

Feature extraction & Transformation

Includes Image features at frame level as well as motion descriptors, camera metadata like camera motion, date, place, time etc among which Motion calculation is computationally expensive.

*Corresponding author: **K.Kumar**
Research Scholar, PRIST Univeristy, Tamil Nadu, INDIA

Video Classification

Human body motion recognition, goal detection, gestures etc. Not enough training data for to learn event of interests and domain knowledge dependence.

Video Clustering

Clustering of shots[28], segments[27] for video, for indexing etc. Scalability for large video clusters is an issue.

Video Association

Finding semantic event boundaries and temporal distance thresholds. Such videos and their association is challenging tasks before researchers.

Video Mining trends

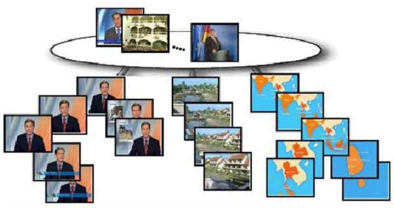


Fig. 1 Intra video mining- Clustering news, weather, stocks
Inter video mining

Inter video mining

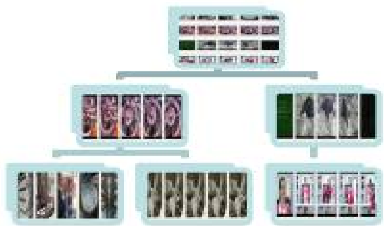


Fig. 2 Inter video mining- Clustering Youtube video clips

Fig. 3 Concept level video – Match point with Federer

Problem Identification

The traditional algorithms discover valid rules by exploiting support and confidence requirements and use a minimum support threshold to prune its combinatorial search space. Two major problems may arise when applying such strategies.

(1) If the minimum support is set too low, this may increase the workload significantly, such as the generation of candidate sets, construction of tree nodes, comparisons and tests. It will also increase the number of rules considerably, which causes the traditional problem of algorithms suffering from extremely poor performance. In addition, many patterns involving items with substantially different support level are produced, which usually have a weak correlation and are not really interesting to users.

(2) If the minimum support threshold is set too high, many interesting patterns involving items with low support are missed. Such patterns are useful for identifying associations among rare but expensive items such as diamond necklaces, rings and earrings, as well as the identification of identical or similar web documents, etc.

Recently, there are some growing interests in developing techniques for mining association patterns without a support constraint [23–25]. The algorithms proposed in [23] are limited to dealing with identifying pairs of similar columns. The approaches presented in [24] and [25] employ a confidence-based pruning strategy instead of the support-based pruning adopted in traditional association rule mining. The mining of support-free association discovers rules in the patterns with high support, cross-support where items have widely differing support levels, and low support. In fact, patterns with a high minimum support level often are obvious and well known; patterns with cross-support level have extremely poor correlation, and patterns with low support often provide valuable new insights. Due to lack of practical and efficient methods for mining rules among infrequent items, we propose Hash Based Video Association Algorithm to explore interesting associations among infrequent items with memory-resident data structure.

Hash-Based Video Association Algorithm

Step 1: Find all combinations of the index values of infrequent items, where the index values are the positions of infrequent items in matrix Inf1.

Step 2: Compute hash value of each combination by exploiting the hash function, h(x, y), provide by library, where the hash value is the index of the hash table. Then hash (i.e., map) them into different buckets of the hash table and increase the corresponding count. Hash collision occurs when different itemsets are assigned same hash value.

Step 3: Use function interest(X, Y) to prune uninteresting itemsets.

Step 4: Employ constraints correlation(X, Y), CPIR(X, Y) to capture rules of strong interest.

InputD: database; minSupp: minimum support;
minConf: minimum confidence;
minInte: minimum interest
Output AR*:* association rules

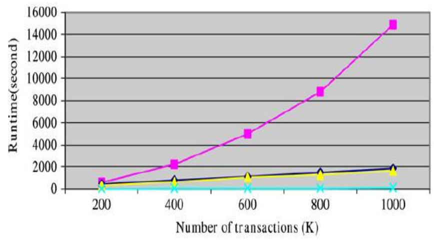
scan the database D and find all infrequent 1- Itemsets (Inf1)

(1) create a hash table
(2) read database 2nd time
(3) for each transaction Ti in D do{
(4) identify all infrequent items;
(5) find all combinations of them;
(6) hash each combination to hash table and obtain a hash index and increase the value of hash index by 1
(7) for each k-itemset I in hash table do for ¥itemsets X,Y, (X union Y) = I and (X intersection Y) = null do
If interest (X,Y) < minInte then Infk:= Infk-{I};
end end

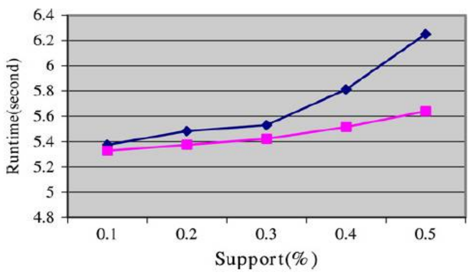
(8) for each infrequent k-itemset of interest {X union Y} in Infk do If Correlation(X,Y)>1 && CPIR(Y|X) >=minConf then AR :={X => Y}
If Correlation(X,Y)>1 && CPIR(X|Y)>=minConf then
AR := {Y => Y}end
(9) return AR

RESULTS AND DISCUSSION

We select the well-known IBM synthetic database generated by the generator in [27]. This choice gives us the flexibility of controlling the size and dimension of the database. The parameters for generating a synthetic database are the number of transactions D, the average transaction size T and the average length I of so-called maximal potentially large itemsets. The number of items in the database is 1000. We conduct experiments on a varied average transaction size and differing number of transactions to test our algorithm's efficiency and scalability. The program is developed in the language JAVA. In this paper, the runtime includes both CPU time and I/O time. In all of the following experiments, we limit the length of association rules to 5.Since our algorithms can be applied to mine association rules among frequent items with bounded length, we compare our schemes with the most influential algorithms for association rule mining, which are the Apriori algorithm and FPGrowth.



Comparing HBVAS against Apriori, FP Growth and HBVAS for limited ransactions



Comparing HBVAS against Matrix Based Scheme (proposed by Ling Zhou et al(2007)) on runtime

Conclusion and future work

In this paper, we have proposed HBVAS for efficient discovery of association rules among infrequent items, which also can be applied to mine association rules efficiently among frequentitems with limited length. In our method, we need to traverse database twice, and we exploit thepruning function interest (X, Y) to reduce the search space considerably, as well as using interestingness measures, correlation (X, Y) and CPIR (X, Y), to extract rules of strong interest. However, HBVAS is a limited option for mining association rules with any length among frequent items due to the expensive cost of hash collision.

References

1. R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the Association for Computing Machinery-Special Interest Group on Management of Data, May 1993, pp. 207-216.
2. R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Databases, VLDB, September 1994, pp. 487-499.
3. R. Agrawal, R. Srikant, Mining sequential patterns, in: International Conference on Data Engineering, ICDE, 1995, pp. 85–93.
4. R. Srikant, R. Agrawal, Mining quantitative association rules in large relational tables, in: Proceedings of the Association for Computing Machinery-Special Interest Group on Management of Data, ACM, 1996, pp. 1–12.
5. R.J. Bayardo, Efficiently mining long patterns from database, in: Proceedings of the 1998 ACM SIGMOD International Conference on Management of data, 1998, pp. 85–93.
6. G. Dong, J. Li, Efficient mining of emerging patterns: Discovering trends and differences, in: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD, August 1999, pp. 43-52.
7. M. Kamber, J. Han, J.Y. Chiang, Metarule-guided mining of multi-dimensional association rules using data cubes, in: Proceeding of 3rd International Conference on Knowledge Discovery and Data Mining, KDD, 1997, pp. 207-210.
8. B. Lent, A. Swami, J. Widom, Clustering association rules, in: Proceeding of International Conference Data Engineering, 1997, pp. 220-231.
9. R. Agrawal, T. Imielinski, A. Swami, Database mining: A performance perspective, IEEE Transactions on Knowledge and Data Engineering 5 (6) (1993) 914-925.
10. J. Han, J. Pei, Y. Yin, Mining frequent pattern without candidate generation, in: Proceeding of ACM SIGMOD International Conference Management of Data, 2000, pp. 1–12.
11. M. Chen, J. Han, P. Yu, Data mining: An overview from a database perspective, IEEE Transactions on Knowledge and Data Engineering (6) (1996) 866–881.
12. H. Mannila, H. Toivonen, A. Verkamo, Efficient algorithm for discovering association rules, in: Knowledge Discovery and Data Mining, KDD, 1994, pp. 181-192.
13. A. Savasere, E. Omiecinski, S. Navathe, An efficient algorithm for mining association rules in large databases, in: Proceeding of the 21stInternational Conference on Very Large Databases, VLDB, September 1995, pp. 432-444.
14. R. Srikant, R. Agrawal, Mining generalized association rules, in: Proceedings of the 21th International Conference on Very Large Data Bases, VLDB, September 1995, pp. 407–419.
15. S. Brin, R. Motwani, C. Silverstein, Beyond market basket: Generalizing association rules to correlations, in: Special Interest Group on Management of Data, 1997, pp. 265-276.
16. B. Padmanabhan, A. Tuzhilin, Small is beautiful: Discovering the minimal set of unexpected patterns, in:

Proceeding of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 54-63.

17. H. Liu, H. Lu, L. Feng, F. Hussain, Efficient search of reliable exceptions, in: Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining, PAKDD, 1999, pp. 194–204.
18. [18] F. Hussain, H. Liu, E. Suzuki, H. Lu, Exception rule mining with a relative interestingness measure, in: Pacific–Asia Conference on Knowledge Discovery and Data Mining, PAKDD, 2000, pp. 86–97.
19. S. Hwang, S. Ho, J. Tang, Mining exception instances to facilitate workflow exception handling, in: Proceedings of the Sixth International Conference on Database Systems for Advanced Applications, 1999, pp. 45-52.
20. A. Savasere, E. Omiecinski, S. Navathe, Mining for strong negative associations in a large database of customer transactions, in: Proceedingsof the Fourteenth International Conference on Data Engineering, ICDE, 1998, pp. 494-502.
21. X. Wu, C. Zhang, S. Zhang, Efficient mining of both positive and negative association rules, ACM Transactions on Information Systems (2004) 381-405.
22. X. Yuan, B. Buckles, Z. Yuan, J. Zhang, Mining negative association rules, in: Proceedings of the Seventh International Symposium on Computers and Communications, ISCC, 2002, pp. 623-629.
23. E. Cohen, M. Datar, S. Fujiwara, A. Gionis, P. Indyk, R. Motwani, J. Ullman, C. Yang, Finding interesting associations without supportpruning, in: Proceedings of the 16th International Conference on Data Engineering, ICDE, 2000, pp. 489-500.
24. K. Wang, Y. He, D. Cheung, Y. Chin, Mining confident rules without support requirement, in: Proceedings of ACM International Conference on Information and Knowledge Management, CIKM, 2001, pp. 89-96.
25. H. Xiong, P. Tan, V. Kumar, Mining strong affinity association patterns in data sets with skewed support distribution, in: Proceedings of theThird IEEE International Conference on Data Mining, ICDM, 2003, pp. 387–394.
26. J. Ding, Efficient association rule mining among infrequent items, Ph.D. Thesis, University of Illinois at Chicago, 2005.
27. Pravin M. Kamde, Dr. S. P. Algur, “A Survey on Web Multimedia Mining” communicates for May 2011 issue in the *International Journal of Multimedia and its Application* (IJMA), ISSN 0975-5578.
28. Daniel Getica Perez, Ming-Tin-Sun and Chang Gu, “Semantic Video Object Extraction Based on Backward Tracking of Multivalued Watershed”

How to cite this article:

K.Kumar and P.Sudhakar (2017) 'Generating Temporal Association Rules for Infrequent Items in Concept Level Video Datasets', *International Journal of Current Advanced Research*, 06(09), pp. 5811-5814.
DOI: http://dx.doi.org/10.24327/ijcar.2017.5814.0809

****