2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)
IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V51. 115

Research On Active Object Data Association Mining Technology Based on Surveillance Video

Yujie Zheng[1]

[1]School of Computer, GuangXi, Economic Mana gement Cadre College, GuangXi, Nanning 530007

Abstract. With the rapid development of industrialization and urbanization, the public order situation becomes more and more complicated. Video surveillance is a large number of operations used to obtain characteristics of the suspect. In gang crime, the relationship between the suspects is the key clue to solving cases, which is of great significance to the detection of cases. Similar path to get active object is the relationship of the most effective method by measuring the similarity of the object activity measurement path relationship between objects, the scope for activity area is small, the path the whole situation. However, in the surveillance video, criminal gangs and sometimes appear together, sometimes separately has led to the track do not yet have similar characteristics. To solve this problem, this paper statistics on the active object in the intersection when the size of the airspace to measure the relationship between them. Firstly, we study tagging technology, which the video is converted to objects, time and space structured data. Secondly, Impact of research activities in space-time domain model objects, given time and space of the domain based on the support and confidence calculation. Finally, the data mining algorithm is given. This research study to explore spatial and temporal activities of law has significance for video annotation, and association mining technology to provide new technical means, in the criminal investigation has important application value.

Keywords: surveillance video; active object; data association mining;

1. Introduction

In view of the challenges facing the security situation in recent years, Ministry of Public Security launched police training and technology demonstration urban construction and the Eleventh Five-Year key project urban video surveillance and alarm demonstration project special science and technology research. It Hope that through technology research and integration, integration of video surveillance and alarm community resources to enhance the capacity of integrated prevention and control of public order. With the construction of video surveillance the continuous increase, video detective work has been unprecedented developments. Video technology has become the criminal investigation techniques, network surveillance technology, means of a new criminal investigation after the detection means. In practice, many large cases is able to quickly detect, much will depend on the application of video detection techniques. Video detection technology has become a new growth point of detection. Video processing and analysis services have become a public security one of the fastest growing business.

For the detection of gang crime cases, the suspect target is an important clue to the relationship, particularly gang crime is a complex, which is objective analysis of the relationship between the suspects is an indispensable means. In gang crime, the suspects are often presented together with the activities of target characteristics. This feature easily captured by surveillance video. Through video processing and analysis techniques to analyze the surveillance video of the event object relationship, will be of great significance detection of cases.

+ Corresponding author.
Email: yujiefeather@sina.com

Association data mining is a large data set hidden in the relationship of interest method. Contact is usually found in association rules or frequent item sets in the form said. Association rule is of the form X-> Y implication expression. Where X and Y are disjoint entries in the database set. Support and confidence of association rules is a measure of the strength of indicators. Rules for support are in the data set the frequency, the confidence level that contains the X, Y in the frequency of transaction. Users can set the support and confidence threshold to get interested in the association.

Traditional data association mining technique does not consider time and space factors, does not apply to attribute data with spatial and temporal analysis of the association. Similar path in recent years associated with mining activities in the object space and time in considering the development of track after a new technology. It is by calculating the time interval the maximum distance of two active objects, to calculate the trajectory of the similarity of the two active objects, and to track the activities of high similarity to the object in the group that the same group of active objects with strong association. In this paper, firstly, Surveillance video for the characteristics of the object of marking technology-based activities, through the technology of video into objects, time and space structured data. Secondly, in the data based on research showing the object in space and time characteristics of the environment, we establish of space-time of active object model. At the same time, based on the model for the spatial and temporal properties of the object with an association rule, define the support and confidence of new calculation methods, and other metrics. Finally, in the new metric, the proposed new activity object association mining algorithm, the surveillance video of the active object-oriented association mining.

2. Related Work

In 1993, Agrawal firstly proposed the concept of association rules [1]. Through the support and confidence to measure the association rules between items in the database to resolve trade association relations. In 1994, he presented classic association rule mining algorithm - Apriori algorithm based on frequent sets [2]. Mannila et al [3] proposed sampling algorithm attempts to use fewer numbers of scans or scan the transaction log to get less performance improvement. In 1995, Park et al [4] proposed the DHP algorithm, the hash table structure for association rule mining, effectively reducing the number of candidate sets. Agrawal et al [5] consider the sequence data is proposed pattern mining, aims to dig out a long sequence of frequent sub-sequences. Sequential pattern mining has become one of the important issues of data mining. In 1996, Srikant and Agrawal proposed based on Apriori algorithm GSP [6]. The algorithm is similar to the Apriori algorithm, using redundant pattern pruning strategies and the candidate hash tree data structure to achieve fast access to the candidate pattern. Because the algorithm is based on Apriori algorithm, so the algorithm also exists the need for the database scan loop, if the sequence of relatively large size of the database, you may have a large number of candidate sequential patterns and other issues. Cheung et al [7] for Aporiori algorithm for a static database, when the database changes, the need to re-scan the database problem, first proposed incremental mining frequent patterns, how changes in the database has been the frequent pattern Updated and made the first incremental mining algorithm FUP. The algorithm solves a new transaction when added to the database, frequent pattern of incremental update problem. In 1997, Lee et al [8] FUP2 updated too frequently for the problem, proposed DELI algorithm, sampling and statistical method to estimate the updated set of frequent patterns before and after the differences between which to decide whether to run an incremental algorithm. In 1998, Feng [9] remain unchanged in the database, minimum support and minimum confidence threshold is changed such as the updating algorithm IUA.

In 2000, Han Jiawei et al [10] need to be repeated for the Apriori algorithm scans the database model in dealing with long and dense data sets has shown low performance issues, made no direct generation of candidate set generation algorithm for frequent pattern FP-growth. The basic idea is to compress the entire expression database for the FP-tree, frequent pattern mining process will be converted to generate the conditions for the recursive sub-library and the corresponding conditions of the process of FP-tree[11]. FP-growth algorithm is very intensive in the data set or support threshold is low, FP-tree density is very high, counting off the effectiveness of its maintenance costs, the performance quite well. In the FP-growth algorithm is proposed, a large number of researchers studied the use of the algorithm to solve incremental sequential pattern mining and mining issues. In 2004, Pei proposed PrefixSpan for sequential pattern mining algorithm [12].

International association for the traditional rules and sequential patterns mining is very large [13]. There are many domestic scholars in-depth research. However, the activities that have spatial and temporal properties of objects associated with mining, domestic and foreign are in the beginning stages, Yida Wang and Hsiao-Ping Tsai, who do exploratory research [14]. Their similarity is by determining the path to measure the association between the degrees of active objects, object grouping of activities. This method requires the path of the object in the activities of a relatively complete for the case of relatively small areas. This article focuses on the activities of video monitoring environment object, the path is usually incomplete, and the relatively large area, direct use of the path to measure the similarity between objects associated with the degree of association may result in some important relationships are missed [15].

3. Active Object Semi-supervised Learning Integrated Video Tagging Algorithm

With traditional key frame-based multi-concept learning methods, active objects for semi-supervised learning to extract the active object area, and translate them into multi-concept learning problems. For each concept of modeling and many iterations of the semi-supervised learning, access to the initial study results. The concept of active objects and through the correlation is between improve the initial tagging results.

3.1. The Active Object Semi-supervised Learning Algorithm

(1). Feature Extraction

Color feature extraction using color histogram, assume color values for the image number of pixels is for the i, N is the total number of pixels, K is the range of possible colors, Color characteristics $h(i)$ is:

$$h(i)=\frac{n_i}{N}, i=0,1,...,k \tag{1}$$

Texture feature extraction using the Tamura texture, Tamura texture features six components, namely roughness, contrast, orientation degree, lines like the degree of regularity, and rough degree, the first three components of which the most important. Roughness of the calculation can be divided into several steps. First, calculate the size of the image:

$$A_k(\mathrm{x,y})=\sum_{i=x-2^{k-1}}^{x+2^{k-1}}\sum_{j=y-2^{k-1}}^{y+2^{k-1}} g(i,j)/2^{2\mathrm{k}} \tag{2}$$

Where k = 0, 1, ..., 5. $g(i,j)$ is Pixel intensity values. For each pixel, calculate its horizontal and vertical non-overlapping windows on the average intensity difference between them.

$$\left.\begin{aligned} E_{k,h}(x,y)&=|A_k(x+2^{k-1},y)-A_k(x-2^{k-1},y)| \\ E_{k,v}(x,y)&=|A_k(x+2^{k-1},y)-A_k(x,y-2^{k-1})| \end{aligned}\right\} \tag{3}$$

One for each pixel, enabling the maximum E value (regardless of direction) of the k value used to set the optimal size $S_{best}(x,y)=2^k$. Finally, the roughness can be calculated in the average of the whole image S_{best} to be represented as:

$$F_{crs}=\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n} S_{best}(i,j) \tag{4}$$

(2). Multiple Concept Semi-supervised Learning

Considering k more than semantic concept learning problem, the set of semantic concepts is $C=\{C_1,C_2,...,C_K\}$, Set a total of N samples $x=\{x_1,x_2,...,x_N\}, x\in R^d$, All samples are marked on the samples that the matrix Y. $Y=[y_1,y_2,...,y_N]^T$. Semantic concept in many semi-supervised learning, in order to take full advantage of unlabeled data and the links between concepts in order to obtain a better model, the concept of semi-supervised learning and more energy is defined as the following minimization problem:

$$F^*=\arg\min_F E(F)=\arg\min_F\{L(F)+E_S(F)+E_C(F)\} \tag{5}$$

$L(F)$ is the loss function, which used to punish the decision function has been marked in the output of the sample offset relative to the known labeling. $E_S(F)$ is Regularization term, which used to constrain decision-making function, so that it has the smoothness of the sample graph. $E_C(F)$ is another regularization term to ensure that decision-making function in the semantics of the concept map smoothness. The loss function $L(F)$ is defined as follows:

$$L(F)=\infty tr\left((F-Y)^T \wedge (F-Y)\right) \tag{6}$$

The smoothness of the sample mapping has defined as:

$$E_S(F)=tr\left(F^T(D-W)F\right)=tr\left(F^T \triangle_s F\right) \tag{7}$$

General similarity matrix W is calculated according to sample characteristics:

$$W(i,j)=\exp\left(-\sum_{i=1}^{d}\frac{(x_{il}-x_{jl})^2}{\sigma_l^2}\right) \tag{8}$$

$$E_c(F)=\frac{1}{2}\sum_{i,j=1}^{K}C_{ij}\left\|h_i-h_j\right\|^2=tr\left(F\triangle_c F^T\right) \tag{9}$$

3.2. The Multi-comcept Active Learning Algorithm

For many the result of the concept of semi-supervised learning, in every round of a concept and a number of selected samples were hand-marked, the concept of selection criteria is the largest performance gain expected, and then the samples were hand-learning initiative.

(1). Sample Selection

The concept of selection criteria to determine, first of all to determine the performance of multi-dimension concept of measure, where the most direct way, that measure the average performance of all the concepts.

$$pref_{avg}=\frac{1}{c}\sum_{i=1}^{c}pref_i \tag{10}$$

Where $pref_i$ denotes the i concept performance. According to this performance metric, it can be a greedy criterion, that is, always select the highest expectations of the concept of performance gain, while the expectations of each of the concept of performance gain using two rounds of the recent changes in the learning process to approximate the performance, this performance changes can be labeled according to the latest group of samples to calculate. It is noteworthy that there actually can be more widely used performance measure, given the different concepts for different weights.

$$pref_{avg}=\frac{\sum_{i=1}^{c}\lambda_i pref_i}{\sum_{i=1}^{c}\lambda_i} \tag{11}$$

This will be given greater weight to ensure that the concept of tagging accuracy. In this case, only the concept of expected performance to the corresponding weight gain can be given.

4. The Formal Definition And Data Mining Algorithm On Active Object Spatial-Temporal Association Rule

4.1. The formal definition of spatial and temporal association rules

Transaction table in the data entry time and space focus, time and space between item sets denoted as τ, Set of temporal data entries are recorded as I, Front piece $X\subseteq I\cup\tau$ and Rear parts $Y\subseteq I$. X contains at least one time and a spatial relationship between data items. Set in the time and spatial relations items item sets are referred to as temporal item sets, which is recorded as φ. $\varphi=X\cap\tau$. Events in space-time generalization of the table, the record of φ is $EGT<\varphi>$. In the space-time transaction data table, φ, X, Y and $X\cup Y$ denoted as $STT<\varphi>$, $STT<X>$, $STT<Y>$,and $STT<X\cup Y>$. Event probability is the generalization of space-time events, including the number of records in the table all the records and the ratio is expressed as:

$$probability=\frac{\left\|EGT\langle\varphi\rangle\right\|}{\left\|EGT\right\|} \tag{12}$$

Where p% probability that the event probability of occurrence. In order to avoid a small probability events have a greater degree of support and confidence in a given p% of the minimum threshold min Pr, the space-time transaction data from the table exclude the small probability of less than min Pr record of events, which can improve the spatial and temporal association. Association rule mining efficiency and credibility of the rules.

Support refers to the space-time transaction data table X and Y the number of simultaneous transactions and all matters contained in the ratio of the number, which is expressed as:

$$Support\left(X \Rightarrow Y\right) = \frac{\left\| STT\left\langle X \cup Y \right\rangle \right\|}{\left\| STT\left\langle \varphi \right\rangle \right\|} \tag{13}$$

Confidence refers to the space-time transaction data table, and the number of simultaneous transactions and the ratio of the number of transactions, which is expressed as:

$$confidence\left(X \Rightarrow Y\right) = \frac{\left\| STT\left\langle X \cup Y \right\rangle \right\|}{\left\| STT\left\langle X \right\rangle \right\|} \tag{14}$$

4.2. The Temporal association rule mining algorithm

Temporal association rule mining processes generally enter parameters are: space-time event table STET, Space data sheet STDT, r, w, minPr, minSup, and minConf. The output is temporal association rules STAR. To reduce the computational complexity of the algorithm after STET Table generalization, followed by EGT pruning to reduce computation in later steps involved in computing the number of records in order to improve the efficiency of algorithms. The steps of the algorithm described below:

(1). Main algorithm

Home rule set R is empty. Called frequent item sets generating function Get_frequent (D, minsup), generated frequent sequences back to the F. Called incremental update function Update (d, D, F, minsup), the new frequent sequences generated back to the F. Consideration of each frequent item sets frequent elements F 11. Li elements are to consider frequent elements of each sub-subsequence (li). If there is association rules subsequence (li) -> li-subsequenee (li) is greater than the credibility of the credibility of setting minconf, this association rule is incorporated into the rule set R.

(2). Frequent Item sets algorithm

Call GetF1 (D) function gets the first-order frequent item sets, the function key is to identify the database D contains the same set of data items of the item, if the same number of item sets is greater than the sum of the minimum support count, then return to these first-order Item sets. Call GetC2 (F1, h) function to obtain the second set of candidates, which function mainly to first-order frequent items in chronological order by order any combination of any two candidate sets and return the second set of candidates, in combination to note the time difference before and after the item sets the maximum time interval can not be greater than h, so that a subsequent order to ensure the maximum time interval h has the limit. If the candidate set C2 in support of item sets the minimum support is greater than the number of second order will be classified as a frequent item set F2. Called the candidate set generation function Get_candidate (Fk-1) be candidate item sets Ck, where k greater than 3. K-frequent item sets will be incorporated into the frequent item sets F, and return F

(3). Update algorithm

The original database D will be the last (n-1) * (h + l) a data item added incremental database d, because the original database D and incremental database d is continuous in time, for the sequence may be the worst in terms of Before the (nl) * (h +1) a data item is the original database D, the last data item is the incremental database d. Call count (D, Fk) k-frequent item sets calculated in the original database D, support number, set Fk.countl. Call count (d, Fk) k-frequent item sets computation in support of the incremental database d in the number of set Fk.count2. Those Fk k-frequent item sets in the database count l with count2 less than the total (D + d) the number of k-minimum support item set classified Ok. Call count (d, Ck-Fk) k-calculation of non-frequent item sets in d in the support number, set Ck-Fk. Call count (d, Wck) calculate the incremental database support in the number of set. Finally, Fk 'into the frequent item sets F, and return F. the original frequent itemsets Fk k-minus, and the Ok then get a new k-frequent item sets Fk '.

5. Conclusion

Traditional association rules for temporal properties not considered, as well as the lack of existing space activities of association rules association rules between objects defined in time and space, does not meet the objectives of surveillance video analysis activities in the relationship needs. In this paper, According to the same place at the same time there are two active object relationship between the idea of building activities of the domain space objects, the first time through the active object in the intersection when the size of the airspace to measure the temporal properties of the active object that contains the association size of the degree. The method can also be used for other activities in space-time data measured the relationship between objects. The spatial and temporal domain and temporal association rules affect the measure of the wealth of space-time theory of association rules mining technology to provide a measure of the object relationship of new technology means.

6. Acknowledgment

This work supported by “the Fundamental Research Funds for the Central Universities” in China. The project number: 3101014.

7. References

[1] Agrawal R, Imielinski T, Swami A. 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), pp207–216.1993

[2] Agrawal R, Srikant R. Fast algorithms for mining association rules. In Proceedings of the 1994 international conference on very large databases(VLDB’94), pp487–499.1994

[3] Mannila H, Toivonen H, Verkamo A I. Efficient algorithms for discovering association rules[C]. In: Prec of the AAAI Workshop onKnowledge Discovery in Databases. pp 181-192. 1994

[4] Park JS, Chen MS, Yu PS. An effective hash-based algorithm for mining association rules. In Proceeding of the 1995 ACM SIGMOD international conference on management of data (SIGMOD’95), pp175–186. 1995

[5] Agrawal R, Srikant R. Mining sequential patterns. In: Proceedings of the 1995 international conference on data engineering(ICDE’95), pp 487–499.1995

[6] Srikant R, Agrawal R. Mining sequential patterns: generalizations and performance improvements. In: Proceeding of the 5th international conference on extending database technology. pp3-17. 1996

[7] Cheung, D. W., Han, J., Ng, V., Wong, C.Y. Maintenance of discovered association rules in large databases: an incremental updating technique[C].In Proc.1996 Int. Conf. Data Engineering, pp106-114. 1996.

[8] Cheung, D.W., Lee, S. D., Kao, B. A general incremental technique for maintaining diseovered association rules[e]. In: Proc. of the15th Int. conf. On databases systems for advanced applications, 1997.

[9] Lee, S.D., Cheung, D.W. Maintenance of diseovered association rules: When to update[C]? In: Proc SIGMOD’97 Wbrkshop on Researeh Issue on Data Mining and knowledge Diseovery. 1997.

[10] Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. In Proceeding of the 2000 ACM-SIGMOD international conference on management of data (SIGMOD’00), pp 1–12. 2000

[11] Pei J, Han J et.al. Mining sequential patterns by pattern-growth: the prefixspan approach. IEEE Trans Knowl Data Eng 16:1424-1440 2004

[12] M.Vlachos, G. Kollios, D. Gunopoulos, Discovering similar multidimensional trajectories. In: Proceedings of the 18th International Conference on Data Engineering(ICDE’02),2002

[13] Y. Wang, E. -P. Lim, and S. -Y. Hwang, “Efficient Mining of Group Patterns from User Movement Data,” Data Knowledge Eng., vol.57, no.3, pp.240-282, 2006.

[14] Yida Wang, Ee-Peng Lim, San-Yih Hwang. Efficient algorithms for mining maximal valid groups[J]. The VLDB Journal. 17:515–535

[15] Hsiao-Ping Tsai, De-Nian Yang,and Ming-Syan Chen. Mining Group Movement Patterns for Tracking Moving Objects Efficiently. IEEE Trans Knowl Data Eng ,VOL.23, NO.2, 2011