Journal of Universal Computer Science, vol. 15, no. 9 (2009), 1852-1870
submitted: 29/8/08, accepted: 25/4/09, appeared: 1/5/09 © J.UCS

An Application of the Dynamic Pattern Analysis Framework to the Analysis of Spatial-Temporal Crime Relationships

Kelvin Leong
(The Hong Kong Polytechnic University, Kowloon, Hong Kong
afkelvin@inet.polyu.edu.hk)

Junco Li
(The Hong Kong Polytechnic University, Kowloon, Hong Kong
junco.li@gmail.com)

Stephen Chan
(The Hong Kong Polytechnic University, Kowloon, Hong Kong
csschan@comp.polyu.edu.hk)

Vincent Ng
(The Hong Kong Polytechnic University, Kowloon, Hong Kong
cstyng@comp.polyu.edu.hk)

Abstract: Dynamic pattern analysis refers to analyzing the relationship of spatial patterns at different time points. Traditional spatial pattern analysis such as data clustering can find the spatial patterns extant at a geographical location at a particular time point but failing to identify spatial dynamics, or changes that occur over time in a particular place. In this paper, we present a dynamic pattern analysis framework, the *DPA framework*. This framework allows user to identify three types of dynamic patterns in spatial-temporal data: 1) similar spatial patterns at different time points, 2) interactive relationship between two geographical locations as a result of a specific reason and 3) frequent association rules related to particular types of events, geographical locations, and time points. To evaluate the proposed framework, we used it to analyze a set of reported crime data for a district of Hong Kong and compared the identified patterns with some expectations of field experts and prior empirical studies for this kind of data and patterns. In line with expert predictions, we found strong correlations between school holidays and crime clusters. On the contrary, in our data set, we could not find obvious seasonal dependency. These findings are corroborated by related empirical crime studies.

Keywords: Spatial-temporal Data Mining, Crime analysis, CSCW
Categories: I.2.1, L.6.2

1 Introduction

Traditional spatial pattern analysis approaches find spatial patterns at a particular time point. However, many spatial phenomena are spatially and temporally dynamic. That is to say, they can change their location over time [Skogan 1990] [Anastasia 1999]. Dynamic pattern analysis refers to analyzing the relationship of spatial patterns at different time points. Such relationships can be analyzed in three ways. Firstly, we can find similar spatial patterns at different time points. For example, *we can find that the spatial patterns of Saturday and Sunday are most similar to each other but dissimilar to other spatial patterns of weekdays*. Secondly, we can examine whether there is an interactive relationship between two geographical locations as a result of a specific reason. For example, *on a raining day, the number of passengers using the tunnel increases, whereas the number of passengers using the bridge decreases*. Thirdly, we can identify frequent rules related to particular types of events, geographical location and time points. For example, *most traffic accidents in the tunnel occur on Sunday*. The first type is of interest because one can investigate the degree of similarity between the spatial patterns of different periods and explore their causes. The second type is important because the users can investigate whether a spatial event has truly been reduced in specified locations or if there is only a displacement of spatial events to surrounding neighbourhoods. The third type is important because it allows users to respond different events in different times and in different locations.

Hotspot analysis [Gonzales et al. 2005] [Ratcliffe 2004] finds spatial patterns at a particular time point by clustering, that is, by classifying objects into groups, or, more precisely, partitioning a data set into subsets (clusters) based on some criteria of similarity [Roddick et al. 1999] [Ester et al. 2001] [Shekhar 2003]. A hotspot is geographical location where certain classes of event occur with a higher-than-average incidence. Hotspots are interesting because "*Everything is related to everything else but nearby things are more related than distant things*" [Tobler 1970]. Several clustering techniques include point locations, hierarchical, partitioning, density and clumping techniques etc. are used to find hotspots. According to [Gonzales et al. 2005], the general techniques for discovering crime hotspots are mean center, standard deviation distance, standard deviation ellipse, and data clustering. All of these techniques can identify hotspots at a particular time point, but they cannot identify the relationship of hotspots at different time points. In recent years, a number of data mining techniques have exploited spatial pattern analysis. These approaches include collocation pattern mining [Yan H 2004] [Xin et al. 2004], and topological pattern mining [Wang et al. 2005]. These two approaches are able to find static spatial patterns at a particular time point but cannot identify similar spatial patterns between different time points. Three approaches which can incorporate time points into spatial pattern analysis are sequential pattern mining [Srikant et al. 1995] [Srikant et al. 1996], flow pattern mining and generalized spatio-temporal pattern mining [Hsu et al. 2007]. However, all of these approaches are unable to identify the interactive relationship between two geographical locations as a result of a specific reason.

In [Leong et al. 2008], we introduced a model, STEM, to to find frequent rules among events, hotspots and time points. In this paper, we propose an extended Dynamic Pattern Analysis Framework, *DPA Framework* for analyzing dynamic spatial-temporal patterns. It can identify three different types of dynamic spatial-temporal patterns. First, we propose the concept of coefficient of correspondence to identify what spatial patterns are similar to each other over different time points. Second, building on the concept of Weighted Displacement Quotient (WDQ) [Bowers et al. 2003], we identify whether there is an interactive relationship between two geographical locations that can be attributed to a specific cause. Third, we use association mining to find hidden frequent rules related to particular types of events, geographical locations, and time points.

This paper is organized as follows: Section 2 reviews work related to hotspot analysis, WDQ, and association mining. Section 3 introduces the *DPA Framework*. Section 4 provides out experiments and results including a demonstration of outr framework applied to crime-related datasets. Section 5 concludes this paper.

2 Related Work

Section 2.1 reviews the hotspot analysis. we apply single linkage clustering to find hotspots in our framework. Section 2.2 explains the weighted displacement quotient, WDQ. This is a technique to identify whether there is an interactive relationship between two geographical locations as a result of a specific reason. Section 2.3 reviews the concepts of association rule mining. We use association rule mining to discover the hidden frequent rules among particular types of events, geographical locations and time points.

2.1 Hotspot analysis in spatial pattern analysis

In our framework, we apply single linkage clustering to find hotspots because it has advantages of simplicity of implementation for massive files. Single linkage clustering is also known as the nearest neighbour technique, which is a kind of hierarchical-agglomerative clustering. The distance between two groups is defined as the distance between the closest pair of records from each group. Let *i* and *j* are two objects in two clusters *r* and *s* respectively. The computation equation of the distance D between two clusters, *r* and *s*, is as below.

$$D(r, s) = \min[d(i, j)] \qquad (1)$$

Where

$i \in r$

$j \in s$

min [d(i,j)] = minimum distance between i and j for all i, j

2.2 Weighted Displacement Quotient (WDQ) technique for analyzing spatial displacement pattern

To explain the concept of the Weighted Displacement Quotient [Bowers et al. 2003], WDQ, consider three theoretical areas: target area (A), buffer area (B) and control area (C). Area A is a target area affected by specific reason. Area B is a buffer area, which may have been influenced by the specific reason in area A. Area C is a control area, which does not include areas A or B and is unlikely to be influenced by changes within them.

The rationale for detecting displacement using the WDQ is as follows:

- Over any given time point, buffer area B will account for a particular proportion of events within a control area C;
- If geographic displacement does occur, it should displace from the target area A into the buffer area B; and
- If displacement does occur. then, relative to the control area C, event in the buffer area B should increase while event in the target area A should decrease.

To look at changes in proportions over time the WDQ compares the situation after specific reason (t_1) with the situation before (t_0) using three formulas. They are – the success measure (SM), the buffer displacement measure (BDM) and the weight displacement quotient (WDQ).

- *Success measure(SM)*

The success measure, SM, is for the target area A in comparison with the control area C. If the SM is negative, it may indicate that the number of events has decreased as a result of a specific reason in target area A. The success measure can be calculated like formula 2.

$$SM = \frac{A_{t1}}{C_{t1}} - \frac{A_{t0}}{C_{t0}} \qquad (2)$$

A_{t1} and C_{t1} represent the number of the events in area A and C after or during the specific reason; while A_{t0} and C_{t0} are the area A and C before the reason.

- *Buffer displacement measure (BDM)*

The buffer displacement measure, BDM, is for the buffer area B in comparison with the control area C. The buffer displacement measure can be calculated like formula 3.

$$BDM = \frac{B_{t1}}{C_{t1}} - \frac{B_{t0}}{C_{t0}} \qquad (3)$$

B_{t1} and C_{t1} represent the number of events in area B and C after or during the specific reason, while B_{t0} and C_{t0} are the areas before the reason. For positive BDM number, it may suggest that some spatial displacement of events has occurred, from the target area A into the buffer area B.

The calculation of weight displacement quotient (WDQ) is a ratio of BDM and SM.

$$WDQ = \frac{BDM}{SM} \text{ , where } -\infty < WDQ < \infty \qquad (4)$$

The purpose of this calculation is to compare events existing in particular areas between two different time points, t_0 and t_1, where $t_0 < t_1$.

In our framework, we apply WDQ to examine whether an interactive relationship exists between two geographical locations as a result of a specific reason.

2.3 Association rule mining for frequent rules finding

Association rule mining [Agrawal et al. 1993] can find interesting associations and/or correlation relationships among large sets of data items. Association rule mining is a popular and well-researched technique for discovering interesting rules among large set of data items. For example, a police officer may be interested to know if certain groups of crimes consistently occur together. The basic principles and concepts of association rules are as follows. Let D be a database of transactions and I = {i1, i2, …, in} be a set of literals. Each transaction T consists of a set of items where T $\subseteq$ I. An association rule is an expression of the form: X $\Rightarrow$ Y where X and Y are set of some items (itemset) in I, and Y does not present in X. An association rule has two set of thresholds to express the degree of relationship about the rule. First,the support supp(X) of an itemset X is defined as the proportion of transactions in the data set which contain the itemset.

$$Supp(X) = \frac{X}{N} \qquad (5)$$

Where
X = Number of transactions containing X
N = Total number of transactions

$$Supp(X \cup Y) = \frac{X \cup Y}{N} \qquad (6)$$

Where
X = Number of transactions containing X
Y = Number of transactions containing Y
N = Total number of transactions

Second, the *confidence* of a rule is defined as *Conf (X$\Rightarrow$ Y)*. It can be interpreted as the probability that occurrence of X causes occurrence of Y.

$$Conf(X \Rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)} \qquad (7)$$

Where

X = Number of transactions containing X
Y = Number of transactions containing Y

For a rule to be interesting, the rule must satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. To achieve this, association rule generation is a two-step process. First, minimum support is applied to find all frequent itemsets in a database. In a second step, these frequent itemsets and the minimum confidence constraint are used to form rules.

In our approach, we apply association rule mining to discover hidden frequent rules among particular types of events, geographical locations, and time points. Section 3.3.3 explains the details of our approach.

3 The DPA Framework: preparing data and finding rules

The proposed *DPA framework* first prepares the data and then analyses it. Section 3.1 explains the data preparation process. The *DPA framework* provides two types of analysis: basis tendency and dynamics pattern. Section 3.2 explains the basic tendency analysis. Section 3.3 discusses the dynamic pattern analysis.

3.1 The data preparation process

The most common issue with time points is that the time data can be generalized in many ways depending on the context and dimension. In this paper, we apply the concept of domain generalization graph (DGG) [Hamilton et al. 1996] [Hilderman et al. 2001] to generalize time data into different time points, such as holiday, day of week and month. For example, based on DGG, user can map "*1st January 2008*" to "*holiday*" or "*weekday*". Figure 3.1 shows an example of DGG.

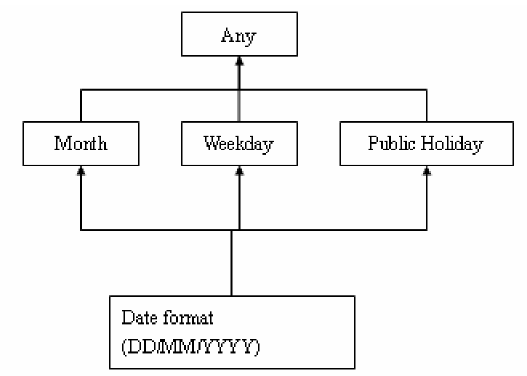

Figure 3.1: An example of DGG diagram

3.2 Basic tendency analysis

The basic tendency analysis aims to provide a general view of the distribution of events over space and time. The basic tendency analysis includes three views: these are (1) viewing distribution of event density over specific periods, (2) viewing distribution of event numbers over specific periods and (3) viewing distribution of hotspot areas over different periods.

The first function, viewing distribution of event density over specific periods, gives users a general view of the distribution of event in focus area. Event density smoothing surface is displayed on the map with a gradual change of coloring.

The second function, viewing distribution of event numbers over specific periods, is simple but essential. It is implemented to give users a general view of the distribution of event numbers in different periods. For example, the distribution of crime numbers over 12 different months, over 7 weekdays or over different long term holiday periods. This function helps user to identify the period(s) that have the highest event number within a certain period.

The third function, viewing distribution of hotspot areas over different periods, provides users a general view of the distribution of hotspot areas in the target area. Therefore, hotspots are demonstrated on the map with different colors.

3.3 Dynamic Pattern Analysis

Dynamic pattern analysis aims to allow user to identify the relationshjp of spatial patterns at different time points. Section 3.3.1 explains how to apply the concepts of coefficient of correspondence to identify what spatial patterns are similar to each other over different time points. Section 3.3.2 explains the concept Weighted Displacement Quotient (WDQ), building on this concept, we can identify whether there is an interactive relationship between two geographical locations as a result of a specific reason. Section 3.3.3 explains how to apply association mining to find hidden frequent rules related to particular types of events, geographical locations, and time points.

3.3.1 To identify similiar spatial patterns over different time points

The coefficient of areal correspondence, CAC, [Taylor 1977] is computed for any two associated areas as the area of intersection, divided by the area of union. The formula of CAC can be calculated like formula 8.

$$CAC = \frac{A_i}{A_u} \qquad (8)$$

Where

CAC = the coefficient of areal correspondence

A_i *= the area of intersection*

A_u *= the area of union*

$A_i \subseteq A_u$

The CAC can provide a simple measure of the degree of overlap between two areas, but it does not consider the density of event in the overlap area.

Building on the concept of CAC, we propose the coefficient of correspondence to identify what spatial patterns are similar to each other over different time points. The user can investigate the periods with similar spatial patterns and the reason(s) of the formation of those spatial patterns. We use coefficient C*a* to measure the similarity of two spatial patterns. Figure 3.3.1 shows an example of the coefficient of correspondence C*a*. The shaded area represents the area of correspondence between the standard deviational ellipse of the time point 1 (*t1*) and the time point 2 (*t2*) data.

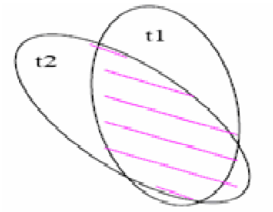

Figure 3.3.1: Area of correspondence

The formula 9 shows the calculation of the coefficient of correspondence C*a*.

$$C_a = \sqrt{\left(\frac{(A_j \times 2)}{(A_{t1} + A_{t2})} \times \frac{(E_{j1} + E_{j2})}{(E_{t1} + E_{t2})}\right)} \qquad (9)$$

Where:
- A_{t1} = *the spatial area of the time point 1 (t1)*
- A_{t2} = *the spatial area of the time point 2 (t2)*
- A_j = *the area covered jointly by both spatial areas at time points 1 (t1) and 2(t2) (shaded part)*
- E_{j1} = *the number of event in shaded area in time point 1 (t1)*
- E_{j2} = *the number of event in shaded area in time point 2 (t2)*
- E_{t1} = *the number of event in the spatial area in the time point 1 (t1)*
- E_{t2} = *the number of event in the spatial area in the time point 2 (t2)*
- $t1 < t2$

The calculation of *Ca* depends not only on the degree of overlap from one time point to another, but also considers the density of events in the joint areas. Therefore, the similarity patterns over two different time points can be calculated according to the proportion of the joint areas and the amount of event count. The range of similarity index is:

Similarity index = *Ca* $\qquad 0.0 \leq Ca \leq 1.0$

The range of possible values for the coefficient is from 0.0 (no correspondence) to 1.0 (complete congruence).

3.3.2 Analyzing the interactive relationship between geographical locations

We adopt weighted displacement quotient (WDQ) in our framework because it can examine whether there is an interactive relationship between two geographical locations as a result of a specific reason. In addition, it is very important for the user to investigate whether spatial event has truly been reduced in specified locations, or if there is only a displacement of spatial events to surrounding neighbourhoods. The details of WDQ have been discussed in section 2.2.

3.3.3 Finding of the frequent rules among events, place and time points

The third type of dynamic pattern is to find frequent rules among events, hotspots and time points by Space-Event-Time Model [Leong et al. 2008]. Let *Y* be a relational database consisting of raw data of daily transactions. Let *R* be a relational table specified for frequent rule finding purpose. *Y* generates different *R*s. Each *R* represents a specific time point, such as week, month, holiday, quarter, etc. In order to generate the frequent rule, data are ETL (extracted, transformed and loaded) into relation *R* from *Y*. Each *R* consists of three attributes: *spatial attribute*, *temporal attribute* and *event type attribute*. The *event type attribute* could be all kinds but spatial or temporal data. For example, we can use disease type for disease pattern analysis or use crime type for crime pattern discovery. Mining frequent rule is to identify significant spatio-temporal event rules.

The frequent rule mining consists of two major phases, data transformation and knowledge discovery. Figure 3.3.3a illustrates the overall concept of frequent rule mining. Firstly, spatial data are transformed by *clustering*; this process can help a user to reduce the size of the data set. As a result, the decision maker can focus on a predefined number of clusters; it makes resource planning more efficient. On the other hand, temporal data are mapped into a meaningful format that allows a user to connect the original date attribute to a more appropriate representation, such as month or holiday. The temporal mapping concept has explained in previous Section 3.1. For the event type information, the selected features will be extracted directly from the original database. The final step is knowledge discovery, in which association rule mining is applied to generate frequent rule.

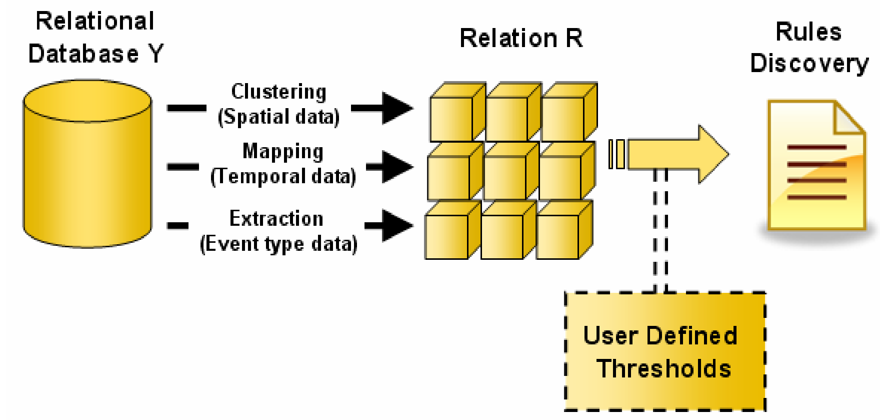

Figure 3.3.3a: Conceptual diagram of frequent rule mining

4 Experiments and results

4.1 Background:

Successful law enforcement depends upon information availability. However, the society in which we live is becoming increasingly complex and dynamic. Police forces are looking for ways to achieve their goals more efficiently and effectively within these dynamic surroundings. More often this requires intra-department co-operation. The introduction of groupware and multimedia applications into organization can improve the intra-department co-operation and has a wide range of applications ([Dustdar 2005], [Gutwin et al. 2008], [Simone et al. 1997], [Wulf et al. 1999]). The *DPA Framework* can help to support intra-department decision making. It allows police officer to allocate resources to areas with more needs and not to lower-priority areas.

4.2 Data source and experiment design

A two year reported crime dataset for a specific geographical district were collected. Each record describes a case of reported crime and is composed of the following items: date of crime, crime types and location of crimes. The data resided in the Microsoft SQL Server 2005. We use MapObjects 2.3 and MapWindows GIS for map display purpose. The programming language is Visual Basic.

4.3 Demonstration of application

Using the *DPA Framework*, decision makers of the police force can better understand the crime patterns. Thus, anti-crime resources can be deployed into right time and right place. It also facilitates decision maker to redefine the crime prevention strategy in the dynamic surrounding. In this section, we demonstrate how to use *DPA Framework* to support intra-department co-operation and decision making in crime prevention.

4.3.1 Performing basic tendency analysis

Viewing distribution of event density over specific periods

Figure 4.2 displays December's crime density distribution in the district. Based on this information, operation bureaus can send more foot patrols into the dense areas.

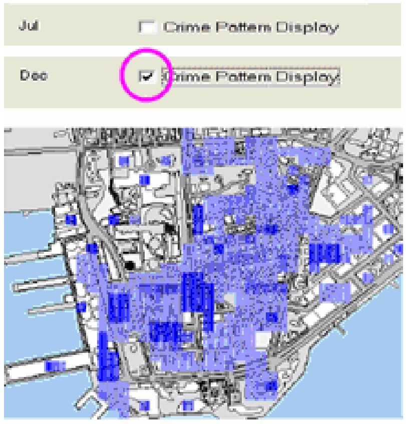

Figure 4.2: Setting the view of the distribution of crime density

Viewing distribution of event numbers over specific periods

The record of monthly crime incidents has shown an increasing trend in Figure 4.3. If current trends continue, police officer may have to consider increased in staffing.

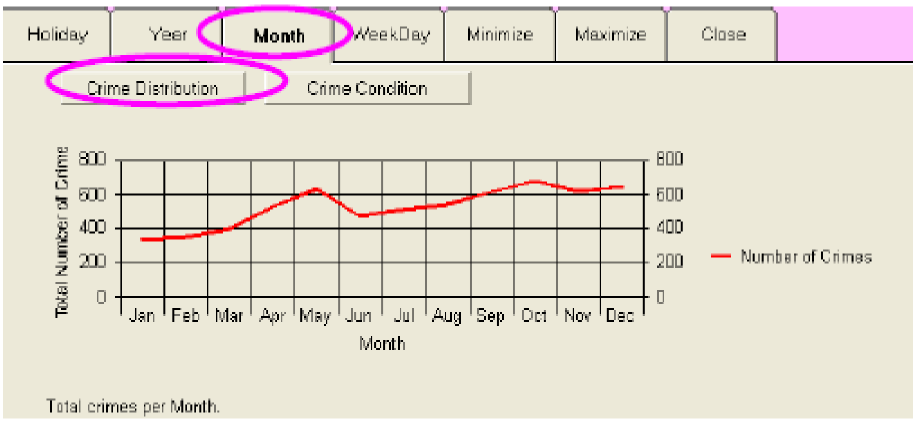

Figure 4.3: The distribution of crime numbers for all months

Viewing distribution of hotspot areas over periods

Crime hotspots are displayed on the map in Figure 4.4. The police tactical unit may consider installing closed circuit television (CCTV) in these hotspot areas.

Figure 4.4: The general view of hotspot area, hotspot pattern in July

4.3.2 Performing dynamic pattern analysis

To identify similar spatial pattern over different time points

Figure 4.5 shows the result of the pair-wise similarity indices of hotspot patterns between different months. It appears that the hotspot patterns of July and December are most similar to each other. This is, of course, only very preliminary data but it may be a hint that further investigation may be justified. This is possibly because there are two long school holidays in these two months. For example, organized crime and triad bureau may wish to investigate whether there are a large number of students involved in these crimes as victims or offenders. Operation bureaus may consider arranging more plain-clothes patrols in the hotspot areas in these two months. Crimes prevention bureaus may consider sending specific alerts to the teenagers.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1.00	-	-	-	-	-	-	-	-	-	-	-
Feb	0.69	1.00	-	-	-	-	-	-	-	-	-	-
Mar	0.52	0.54	1.00	-	-	-	-	-	-	-	-	-
Apr	0.69	0.62	0.52	1.00	-	-	-	-	-	-	-	-
May	0.57	0.55	0.50	0.59	1.00	-	-	-	-	-	-	-
Jun	0.57	0.62	0.37	0.64	0.56	1.00	-	-	-	-	-	-
Jul	0.70	0.63	0.49	0.63	0.51	0.70	1.00	-	-	-	-	-
Aug	0.66	0.58	0.57	0.66	0.62	0.55	0.72	1.00	-	-	-	-
Sep	0.73	0.67	0.63	0.70	0.58	0.61	0.72	0.74	1.00	-	-	-
Oct	0.59	0.60	0.41	0.69	0.62	0.72	0.76	0.66	0.64	1.00	-	-
Nov	0.67	0.64	0.44	0.69	0.56	0.69	0.74	0.66	0.73	0.79	1.00	-
Dec	0.74	0.67	0.49	0.63	0.51	0.62	0.81	0.66	0.69	0.66	0.70	1.00

Figure 4.5: Pair-wise similarity indices from January to December

Analyzing the interactive relationship between geographical locations

For demonstration purposes, the investigated region is divided into 12 fixed sub-regions as per Figure 4.6. Each sub-region is classified into target area A, buffer area B or control area C as per Figure 4.7. The weighted displacement quotient, WDQ, is used to measure spatial pattern displacement or diffusion.

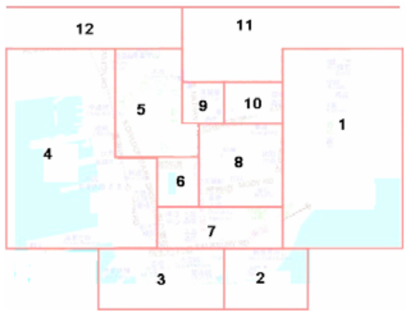

Figure 4.6: The 12 per-defined fixed sub-region

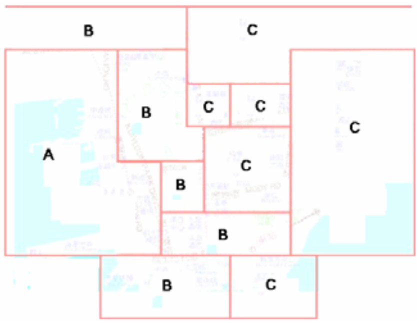

Figure 4.7: The target area, buffer areas and control areas

In Figure 4.8, the three measures are success measure (SM), buffer displacement measures and weighted displacement quotient. We assume that a crime prevention program is implemented in target area in July. The negative SM indicates that the crime prevention program in the target area A is successful when compared to the control areas. The negative BDM indicates that when the crime prevention program is implemented, crime does not rise in the buffer areas to a greater extent than in control areas, suggestion a diffusion of benefits. The positive WDQ indicates that there are no displacements during the crime prevention program. Based on the measure results, the crimes prevention bureaus may evaluate the program's impact on the neighbouring areas. In addition, the operation bureaus may arrange foot patrols schedules in different areas more effectively.

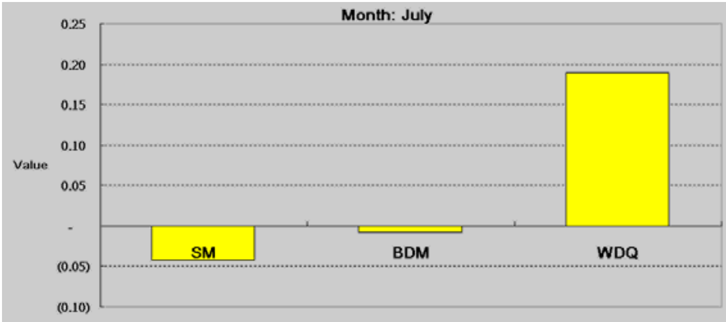

Figure 4.8: The view of the crime conditions for chosen periods

To find the frequent rules among hotspots, holiday and crime types

Some research has demonstrated the existence of a relationship between holidays and crime. [Jacob et al. 2003] suggested that the level of property crime committed by juveniles decrease by 14 percent on days when school is in session, but the level of violent crime increases by 28 percent when schools are not in session.

In this experiment, we investigate the relationship among *hotspots, holiday and crime types* for analysis. Given minimum support at 10% (equivalent to 608 instances) and minimum confidence at 60%, the top five rules with highest confidence rate are listed in Figure 4.9.

IF	THEN	Conf (%)
Hotspot = α	**Holiday =** *Non-Holiday*	97%
Offence = *Deception*	**Holiday =** *Non-Holiday*	97%
Offence = *Shoptheft*	**Holiday =** *Non-Holiday*	97%
Hotspot = δ	**Holiday =** *Non-Holiday*	96%
Offence = *Misc Theft*	**Holiday =** *Non-Holiday*	96%

Figure 4.9: The top five rules of hotspots, holiday and crime types relationship

We can find a strong relationship between hotspot α and holidays in Rule 1, it shows 97% of crimes in hotspot α occurring during non-holidays. A closer look at the environment of hotspot α may provide an explanation. The area occupied by hotspot α is a commercial zone, where only few workers stay during the holidays. As a result, opportunities (potential victims) for crime are reduced. This explanation can be applied to rule 4 as well.

Figure 4.10: Hotspot α and Hotspot δ

Figure 4.10 shows the hotspotα and hotspot δ. We find that most of "deception" (rule 2), "shop theft" (rule 3) and "miscellaneous theft" (rule 5) occurred during non-holidays. Rule 2 is easy to understand because obtaining property by deception is a kind of white collar crime, which usually happens on working days. Rule 3 may come as a bit of surprise that shop theft is not frequent on holidays. However, it is possible that many family type retail shops located in the district are closed on the holidays. Thus the crime rate drops as well. Such rules provide hints to crime analysis for further investigation. Crimes Prevention Bureaus may consider sending more foot patrols into these two hotspots during non-holiday.

To find the frequent rules among hotspots, quarter and crime types

In this experiment, we investigate the relationship among *hotspots, quarter and crime types* for analysis. Given minimum support of 3% (equivalent 182 instances) and minimum confidence of 25%, the top five rules with highest confidence rate are listed in Figure 4.11.

IF	THEN	Conf (%)
Offence = *Misc. Theft*	Quarter = *Fourth*	35%
Quarter = *Fourth*	Offence = *Misc. Theft*	32%
Hotspot = *β*	Quarter = *Fourth*	30%
Hotspot = *β*	Quarter = *Second*	29%
Quarter = *Third*	Offence = *Misc. Theft*	27%

Figure 4.11: The top five rules of hotspots, quarter and crime types relationship

Rule 1 shows "miscellaneous theft" mainly occurred in fourth quarter. Similarly, most "miscellaneous theft" occurred in the second half of the year (third quarter and fourth quarter) as per rule 2 and rule 5. On the other hand, the peak crime quarters for hotspot β were second quarter and fourth quarter. Figure 4.12 shows the hotspot β. We believe that the main reason is because two long holidays, National day and Christmas, are in second half of the year. However, a low minimum support reflect non-obvious relationships between clusters, quarter and crime types.

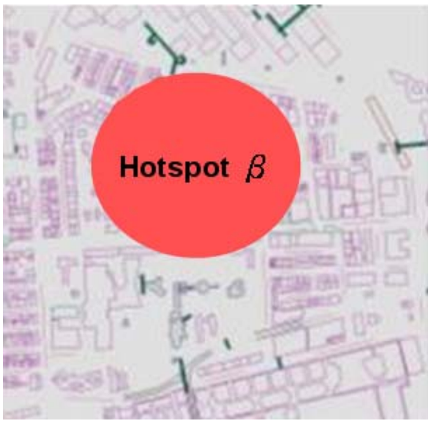

Figure 4.12: Hotspot β

On the other hand, the low confidence rate may indicate that the seasonal dependency of criminal activities is not strong. This crime pattern has been discussed in previous research. [Yan 2004] discovered this Hong Kong experience by examination of the rates of property crime in Hong Kong for the period 1991–2000. To determine the seasonal dependency, regression analysis (using dummy variables) and analysis of variance (ANOVA) were employed; the result shows no significant seasonal dependency. [Yan 2004] indicates that economic need is a more prominent factor than the effect of the season; it also can be used to support our results in Figure 4.11. The low confidence rate result implies that seasonal relationship in our data was not apparent. The higher crime rate in second half of year is mainly caused by holiday's factor (economic need). Thus, Operation Bureaus may consider arranging more plain clothes patrols during holiday.

5 Conclusion

In this paper, we have proposed a novel dynamic pattern analysis framework, *DPA Framework*. Dynamic pattern analysis refers to analyzing the relationship of spatial patterns at different time points. The three types of dynamic patterns introduced in this paper are 1) similar spatial patterns at different time points, 2) interactive relationship between two geographical locations as a result of a specific reason and 3) frequent association rules related to particular types of events, geographical locations, and time points.

In law enforcement, crime prevention is a goal that is subdivided into a wide range of specialized subtasks to several sub-divisions. The *DPA Framework* can play a key role in optimizing police resource and intra-department decision making. We demonstrated and tested the results of the framework by applying it to reported crime data in a district of Hong Kong. In section 4.3.1, we perform basic tendency analysis. The analysis provides a general view of crime distribution over space and time. Section 4.3.2 demonstrates the findings of dynamic pattern analysis. We found strong correlations between holidays and crime clusters. On the other hand, we could not find obvious seasonal dependency, at least in our test data set. These findings are corroborated by related empirical crime studies. Hopefully, the work reported in this paper could be viewed as a step towards enhancing the completeness of spatial pattern analysis.

As future work, we plan to conduct more experiments in order to examine the framework. For example, we intend to include the business nature of the building as an attribute for rule mining or refinement. Moreover, we are interested to review how to analyze the temporal data, so that more deep-lying association can be extracted. In addition, we are developing a system, building on the *DPA framework*, to support cooperative and distributed decision makers for crime prevention.

Acknowledgement

The work of the authors was supported in part by the CERG RGC Grant PolyU 5106/05E.

References

[Agrawal et al. 1993] Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules Between Sets of Items in Large Databases. In: Proceedings of 1993 ACM-SIGMOD International Conference Management of Data, Washington, D.C., 207-216, 1993.

[Anastasia 1999] Loukaitou Sideris, Anastasia: Hotspots of bus stop crime: the importance of environmental attributes. Journal of the American Planning Association. 65 (4), 395- 411, 1999.

Bowers et al. 2003] Kate J Bowers and Shane Johnson.: Measuring the Geographical Displacement and Diffusion of Benefit Effects of Crime Prevention Activity. Journal of Quantitative Criminology, 19(3), 275-301, 2003.

[Dustdar 2005] Schahram Dustdar: Reconciling Knowledge Management and Workflow Management Systems: The Activity-Based Knowledge Management Approach. Journal of Universal Computer Science, vol. 11, no. 4, 589-604, 2005.

[Ester et al. 2001] Martin Ester, Hans-Peter Kriegel, Jörg Sander.: Algorithms and Applications for Spatial Data Mining, Geographic Data Mining and Knowledge Discovery. Research Monographs in GIS, Taylor and Francis, 160 - 187, 2001.

[Felson et al. 1980] Felson, Marcus, Lawrence E.Colen: Human Ecology and Crime: A Routine Activity Approach. Human Ecology 8.4, 389-406, 1980.

[Gonzales et al. 2005] Alberto R. Gonzales, Regina B. Schofield and Sarah V. Hart: Mapping Crime: Understanding Hot Spots. U.S. Department of Justice, 2005.

[Gutwin et al. 2008] C. Gutwin, S. Greenberg, R. Blum, J. Dyck, K. Tee, G. McEwan: Supporting Informal Collaboration in Shared-Workspace Groupware. Journal of Universal Computer Science, vol. 14, no.9, 1411-1434, 2008.

[Hamilton et al. 1996] Hamilton, H.J., Hilderman, R.J., Cercone, N: Attribute-oriented Induction Using Domain Generalization Graphs. In: 8th IEEE International Conference on Tools with Artificial Intelligence, Toulouse, France, 246—253, 1996.

[Hilderman et al. 2001] Hilderman, R.J., Hamilton, H.J.: Knowledge Discovery and Measures of Interest. Kluwer Academic Publishers, 2001.

[Hsu et al. 2007] Wynne Hsu, Mong Li Lee, and Junmei Wang.: Mining Flow Patterns in spatio-Temporal Data and Mining Generalized Flow Patterns, Temporal and Spatio-Temporal Data Mining, IGI Publishing, page 157 to 208, 2008.

[Jacob et al. 2003] Jacob, B.A., Lefgren, L.: Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime. The American Economic Review. 93, 1560—1577, 2003.

[Leong et al. 2008] Leong Kelvin, Chan Stephen, Ng Vincent, Shiu Simon.: Introduction of STEM: Space-Time-Event Model for crime pattern analysis. Asian Journal of Information Technology, 7 (12): 516-523, 2008

[Ratcliffe 2004] Jerry H. Ratcliffe.: The Hotspot Matrix: A framework for the spatio-temporal targeting of crime reduction. Police Practice and Research, Vol.5 Issue 1: 5-23, 2004.

[Roddick et al. 1999] Roddick, J.-F., and Spiliopoulou.: A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research. ACM SIGKDD, Explorations Newsletter Vol.1, Issue 1, 34 – 38., 1999.

[Shekhar 2003] S. Shekhar, P. Zhang, Y. Huang, R. Vatsavai.: Trend in Spatail Data Mining. Data Mining: Next Generation Challenges and Future Directions, Chapter 3, AAAI/MIT Press, 2003.

[Simone et al. 1997] C. Simone and M. Divitini: Ariadne: Supporting Coordination through a Flexible Use of the Knowledge on Work Processes. Journal of Universal Computer Science, vol. 3, no.8, 865 - 898, 1997.

[Skogan 1990] Skogan, Wesley G.: Disorder and Decline: Crime and the Spiral of Decay in American neighborhoods. University of California Press, Los Angeles, 1990.

[Sorensen 2004] Sorensen, D.: Temporal Patterns of Danish Residential Burglary, By month, By week, and Hour of day. Denmark's Ministry of Justice, 2004.

[Srikant et al. 1995] Ramakrishnan Srikant, Rakesh Agrawal.: Mining generalized association rules, In Proceedings of the 21st VLDB Conference, Zürich, Switzerland, September 1995.

[Srikant et al. 1996] Ramakrishnan Srikant and Rakesh Agrawal.: Mining Sequential Patterns: Generalizations and Performance Improvements. In EDBT96, Avignon, France, pages 3--17, 1996.

[Taylor 1977] Taylor, P.J: Quantitative methods in geography: an introduction to spatial analysis, Chapter 5: Areal association, Houghton Mifflin, Boston, page 396. 1977

[Tobler 1970] Waldo Tobler.: A computer movie simulating urban growth in the Detroit region. Economic Geography, 236, 1970.

[UK Home Office 2005] UK Home Office: "Grid thematic maps. Crime reduction toolkits: Focus areas and hotspots", 2005.

[Wulf et al. 1999] V. Wulf, M. Krings, O. Stiemerling, G. Iacucci, M. Maidhof, R. Peters, P. Fuchs-Fronhofen, B. Nett, J. Hinrichs: Improving Inter-Organizational Processes with Integrated Organization and Technology Development. Journal of Universal Computer Science, vol. 5, no.6, 339- 366, 1999.

[Wang et al. 2005] Junmei Wang, Wynne Hsu, Mong Li Lee.: A framework for mining topological patterns in spatio-temporal databases, Conference on Information and Knowledge Management, Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany ,Pages: 429 – 436, 2005.

[Xin et al. 2004] Xin Zhang, Nikos Mamoulis, David W. Cheung, Yutao Shou.: Fast Mining of Spatial Collocations, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004), Seattle, Washington, USA, August 22–25, 2004.

[Xue et al. 2003] Xue Y , & D. E. Brown.: A Decision Model for Spatial Site Selection by Criminals: A Foundation for Law Enforcement Decision Support. IEEE Transactions on System, Man, and Cybernetics – Part C, V.33, N. 1, Feb 2003, pp.78–85, 2003.

[Yan H 2004], Yan Huang.: Discovering Colocation Patterns from Spatial Data Sets: A General Approach, IEEE Transactions on Knowledge and data engineering, Vol 16, No 12 December 2004.

[Yan 2004] Yan, Y.Y.: Seasonality of Property Crime in Hong Kong. The British Journal of Criminology. 44, 276—283, 2004.

[Yen et al. 2006] John Yen, Xiaocong Fan, Shuang Sun, Timothy Hanratty, John Dumer.: Agents with shared mental models for enhancing team decision makings. Decision Support Systems, Vol.41, Issue 3: Special issue: Intelligence and security informatics, 634 – 653, 2006.