Journal of Engineering Sciences
Assiut University
Faculty of Engineering
Vol. 43
No. 1
January 2015
PP. 1 – 15

DEVELOPING SPATIO-TEMPORAL DYNAMIC CLUSTERING ALGORITHMS FOR IDENTIFYING CRIME HOT SPOTS IN KUWAIT

Taysir H. A. Soliman [1,*], Khulood Al Ommar [2], Youssef B. Mahdy [3]

[1], [2] Information Systems Dept., Faculty of Computers and Information, Assiut University, Egypt
[3] Computer Science Dept., Faculty of Computers and Information, Assiut University, Egypt

(Received 8 December 2014; Accepted 8 January 2015)

ABSTRACT

As crime rates are increasing worldwide, crime mining requires more efficient algorithms that can handle current situations. Identifying crime hot spot areas via clustering spatio-temporal data is an emerging research area. In this paper, dynamic clustering algorithms for spatio-temporal crime data are proposed to detect hot crime spots in Kuwait. Kuwait governorates are taken as case study: the capital, Hawalli, Al-Ahmady, Al-Jahra, Al-Farawaniya, and Mubarak Al-kebeer. In addition, different crime types are considered: act of discharge and humiliation, adultery, aggravated assault, bribery, counter fitting, drugs, embezzlement, fight or resist employee on job, forging of official documents, weapon, robbery and attempted robbery, suicide and attempted suicide, and bank theft. Applying Random subspace classification to those clustered data, 98% accuracy and 99.4% ROC are obtained, having precision (98.7%), recall (98.4%), and F1 (98.28%).

Keywords*:* Spatio-temporal data mining, hot spot detection, intelligent crime mining, random subspace classification, and clustering.

1. Introduction

National security in any country is the primary concern of the nation. As the world becomes instrumented and interconnected, spatio-temporal data are more ubiquitous and richer than ever before. Moving object (e.g., taxi, bird) trajectories recorded by GPS devices, social events (e.g., microblogs, crime) with location tag and time stamps, and environment monitoring (e.g., remote sensing images for weather forecasting) are typical spatio-temporal data that we meet daily [1]. Typical spatio-temporal data mining includes crime pattern analysis (i.e., hot spot detection) and prediction, traffic congestion prediction, and epidemic alerts. In the last decade, there has been a research focus on intelligent crime analysis using data mining. However, intelligent crime analysis is still an emerging research area because of increasing crime rates, complexity of crime data correlations, and high volume of crime datasets [2]. It helps to discover patterns of

* Corresponding author.
Email address: taysser.soliman@fci.au.edu.eg

criminal behavior that may help in the discovery of where, when and why particular crimes are likely to occur. This analysis using data mining techniques is of great importance to stakeholders, such as regional planners, politicians, police and residents.

In [2], authors propose a data mining framework for crime mining, including extracting important entities from police narrative reports, which are written in plain text, using Self-Organizing Maps (SOM) clustering method in the scope of crime analysis and then applying Neural Networks as a crime matching engine. In addition, researchers at [3] explain a general overview on applying data mining techniques for intelligent crime mining, including neural networks, Bayesian networks, and genetic algorithms in predicting and matching crime incidents. Moreover, crime pattern mining techniques have been proposed [4-6], illustrating a general overview of crime data mining in order to solve the disadvantages of using geospatial statistics and traditional crime data analysis systems. Other researchers handle time and space in their crime data analysis by distributive similarity between pairs belonging to different geospatial themes across locations [7,8].

In this paper, spatio-temporal clustering algorithms are proposed in order to detect the distribution of crime hot spots in Kuwait, including crime types. In addition, those clustered data are taken into random subspace classification for identifying the hot spot areas. The paper is organized as follows: section two explains related work; section three illustrates the proposed algorithms and applying it to six Kuwait governorates; section four discusses the results; finally, section five illustrates the conclusions and future work.

2. Related work

Hot Spot detection is one of the most popular research techniques for detecting crime areas. In the hot spot model [9], current criminal incident data are collected and clustered over space. The hot spot model only use criminal incident data, such as types of crimes, locations and time of criminal incidents. One of the main disadvantages of hot spot detection is showing the current patterns of crimes without the insight into considering the relationship between crimes and environment over time. This work addresses the need for dynamic clustering algorithms for hot spot detection over time.

There are several research directions in addressing hot spot detections, such as statistical models and data mining techniques. For example, Liu and Brown [10] applied a point pattern density model to criminal incidents, using Bayesian model to address their problem. Xue and Brown [11] and Smith and Brown [12] developed a spatial choice model. They assumed criminals made choices to pick places that could be modeled by random utility maximization. This utility maximization is over all alternatives, where the utility is defined by the gain from crimes and the risk of being caught. Brown, Dalton, and Hoyle [13] then proposed Generalized Linear Models (GLM) to compute the risk over a territory. They first partitioned the space into grids, associating each grid with a response referring whether incidents happened and features about the grid. They applied the spatial GLM to predict terrorist events, showing the spatial GLM had better prediction performance than the density models. Rodrigues and Diggle [14] combined point process models and Generalized Additive Models (GAM) to build a semiparametric point source model. In their model, features affected the risk nonlinearly. They applied the model to study the effect of installed security cameras on crimes.

One of the data mining techniques used in hot spot detection is clustering techniques. The clustering problem has been widely studied because of its applications to a wide variety of problems in customer-segmentation and target-marketing [15-17]. A broad

JES, Assiut University, Faculty of Engineering, Vol. 43, No. 1, January 2015, pp. 1 – 15

overview of different clustering methods may be found in [18], [19]. Currently, there are different kinds of crimes, such as cybercrimes and social media crimes, where clustering can play a vital role in finding criminals. Clustering techniques group data items into classes with similar characteristics to maximize or minimize intra-class similarity. For example, to detect the crime patterns via identifying offenders who conduct crimes in similar ways or distinguishing among groups belonging to different gangs.

Clustering methods can be broadly classified into two groups: partitioning clustering and hierarchical clustering. Partitioning clustering methods, such as K-means and self-organizing map (SOM) [20], divide a set of data items into a number of non-overlapping clusters. A data item is assigned to the ‘‘closest” cluster based on a proximity or dissimilarity measure. There are a number of analysis modules for identifying and assessing potential hot spots in crime analysis: the mode, the fuzzy mode, hierarchical nearest neighbor clustering, risk adjusted nearest neighbor hierarchical clustering, the spatial and temporal analysis of crime routine, K-means clustering, and the local Moran statistic [21].

Dynamic clustering has been previously proposed in [22] to solve the problem of node load in Wireless Sensor Networks (WSNs). In addition, dynamic K-means is suggested to solve fixed number of clusters in [23], which is similar to the current proposed method. However, in this paper, dynamic clustering algorithms are suggested. In addition, K-means algorithm is applied as a static clustering algorithm. Comparisons are made between the performances of the three algorithms, using a real case study of Kuwait governorates.

3. Materials and methods

In this paper, a data mining methodology is proposed, consisting of three main phases to detect crime hot spots, as shown in Fig.1. The first phase is the data gathering phase, where data is collected and cleaned from noise. The second phase is the clustering phase, where three clustering algorithms are used: two proposed dynamic clustering algorithms based on number of clusters and on number of cities per cluster and K-means clustering algorithm. Finally, the third phase identifies crime hot spots in Kuwait using random subspace classifier.

3.1. Data gathering phase

Our data is collected from police departments of Kuwait, obtaining 1000 crime cases. The 1000 cases cover Kuwait governorates. In this case, twelve types of crimes are gathered, as mentioned previously. For each record, the time of the year having high crime rates and low crime rates are determined. In addition, additional information is obtained about both the offender and the crime. For the offender, information, his/ her nationality, blood type, date of birth, his personal look, fingerprint, and address, are recorded in the database. For the crime, information as crime type, time, date, and location, are recorded. In addition to the necessary information discussed previously in section is obtained about both the offender and the crime. A sample of the data is shown in Table 1.

3.2. Clustering phase

In this phase, three clustering algorithms have been applied: K-means and two other proposed clustering algorithms dynamic A and dynamic B. K-means is one of the well-known clustering algorithms, which have been intensively applied in a variety of domains. Dynamic A and Dynamic B are based on dynamic clustering. However, dynamic is an overlapping dynamic clustering algorithm based on dynamic buffers per cluster. Each clustering algorithm has its own characteristics, such as number of clusters, number of cities per cluster, number of crimes per cluster, covered area, and hot spots. A comparison of the characteristics of each algorithm is shown in Table 2. Moreover, the data required for each algorithm are determined according to Table 3. The following sections illustrate the steps of each phase, applied to real data.

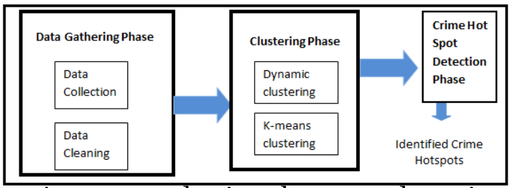

Fig. 1. Spatio-temporal crime hot spot detection phases

Table 1.
Sample of crime records

Age	Race	City	Region	Governorate	Time	Season	Country	Crime
28	White	Al-Qurain	B	Mubarak Al-Kabeer	16:00:00	Fall	Kuwait	Drugs
44	White	Al-Qurain	B	Mubarak Al-Kabeer	15:00:00	Summer	Kuwait	Murder
17	White	Al-Qurain	B	Mubarak Al-Kabeer	22:00:00	Winter	Kuwait	Kidnapping
33	White	Al-Qurain	B	Mubarak Al-Kabeer	11:00:00	Spring	Kuwait	Drugs
40	White	Al-Qurain	B	Mubarak Al-Kabeer	23:00:00	Winter	Kuwait	Rape
36	Black	Mubarak Al-Kabeer	A	Mubarak Al-Kabeer	7:00:00	Winter	Kuwait	Weapon
47	Black	Mubarak Al-Kabeer	A	Mubarak Al-Kabeer	12:00:00	Fall	Kuwait	Murder
49	Black	Mubarak Al-Kabeer	A	Mubarak Al-Kabeer	13:00:00	Summer	Kuwait	Weapon
21	Black	Mubarak Al-Kabeer	A	Mubarak Al-Kabeer	19:00:00	Summer	Kuwait	Weapon
33	Black	Mubarak Al-Kabeer	A	Mubarak Al-Kabeer	16:00:00	Fall	Kuwait	Rape

Table 2.
Characteristics of used/proposed Clustering algorithms

Criteria	K-means	Dynamic A K-means Clustering	Dynamic B K-means Clustering
Number of Clusters per governorate	Static	Variable	Variable
Number of cities per cluster	Variable	Static	Variable
Covered area	Equal	Different	Different
Number of crimes per clusters	Different	Different	Equal
Hotspots	The area of max number of crime cases	The region of max density which is equal to count of crime cases/covered area.	The smallest cluster.

3.2.1. K-means algorithm

K-means clustering algorithm partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The K-means clustering algorithm is a non-hierarchical clustering approach, where data are divided into K groups. The user is flexible to decide the value of K. The aim of this technique is to create K number of clusters so that within group sum of squares are minimized. The algorithm finds a local optimum since iterating all the possible observations is enormous. To reach the optima, algorithm is repeated several times and the best positioning K centers are found. Then, the remaining observations to the nearest cluster to minimize the squared distance are assigned.

Table 3.
Data required for all three algorithms

	K-means	Dynamic A	Dynamic B
Governorate	Total area of each governorate and its cities.	Number of cities per each governorate	Total number of specific crime cases per each governorate and its cities. Our Case on Murder Crime (98 cases from 1000)
The Capital	175	34	39
Hawalli	85	14	14
Al-Ahmady	5,120	16	10
Al-Jahra	12,750	16	5
Al-Farawaniya	204	17	25
Mubarak Al-Kabeer	104	11	5
Total	18,438	110	98

The K-means process is summarized as follows:

1. Initially, let the number of clusters be k (static)
2. Choose a set of K instances as centers of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. This process is iterated.

Taking Mubarak Al-Kebeer governorate case study, K-means algorithm clusters cities based on area to divide each governorate into approximately equal areas. So, one cluster may contain only one city and another may contain many cities as follows. For example, assume we want to cluster the cities into 4 clusters based on cities areas using K-means clustering, clusters are illustrated in Fig. 4 and their shape length and shape areas are shown in Table 4. Having a variation of the number of clusters from two-seven, Fig. 3-8 illustrate partitioning of Kuwait governorates using K-means algorithm. The disadvantage of using K-means is the static number of clusters that the user must input. In addition, crime data may vary each year in any country, so these clusters must be dynamic accordingly. Furthermore, a study of the time performance of the K-means algorithm is shown in later sections, compared with the other two algorithms.

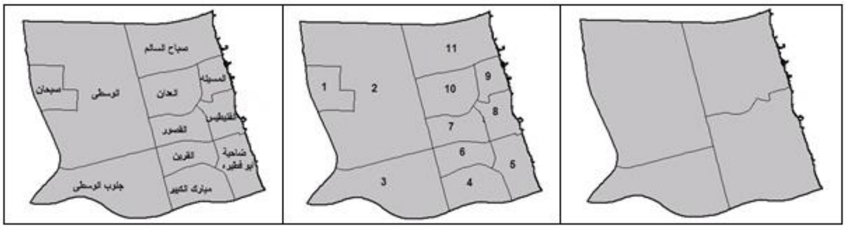

Fig. 2. Mubarak Al-Kebeer Clusters using K-means

Table 4.
Clustered Cities for Mubarak Al-kebeer using K-means Algorithm

City	SHAPE_Leng	Shape_Le_1	Shape_Area
صباح السالم	52387.07207	22002.35093	21470677.21
مبارك الكبير	52387.07207	11630.13607	8040891.188
العدان	52387.07207	10826.04931	6043534.851
الفنيطيس	52387.07207	14018.70985	5169595.344
صبحان	52387.07207	10718.41904	5889952.275
القرين	52387.07207	11913.40189	8892157.672
المسيلة	52387.07207	12738.21197	4001915.848
الوسطى	52387.07207	34049.01978	45866515.09
جنوب الوسطى	52387.07207	18540.99239	15125837.66
القصور	52387.07207	10318.33816	5439104.356
ضاحية أبو فطيرة	52387.07207	12405.42417	7479375.58

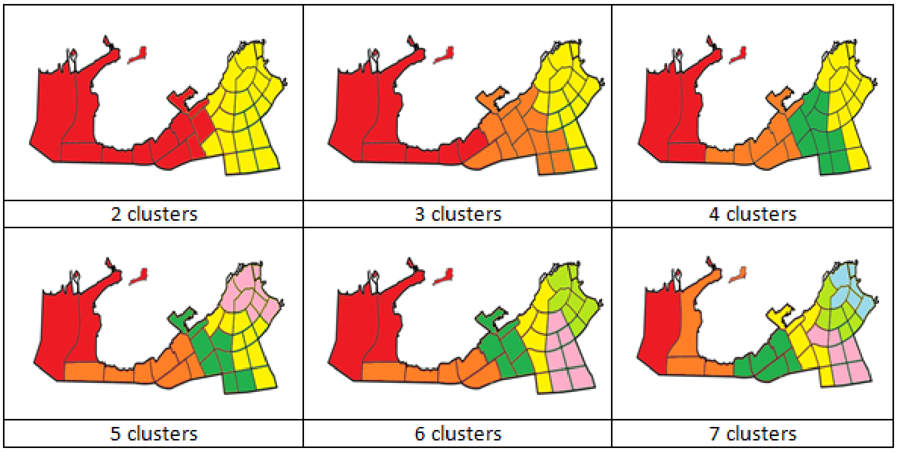

Fig. 3. Application of K-means clustering to the Capital Governorates

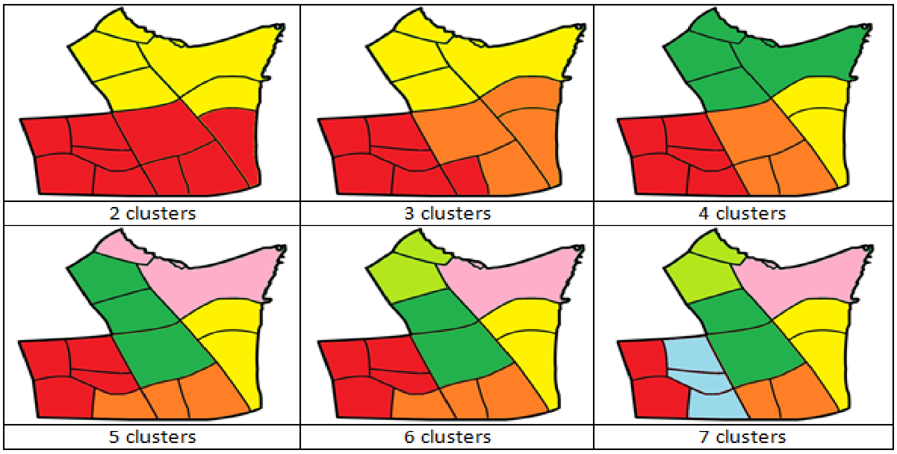

Fig. 4. Application of K-means Clustering Algorithm to Hawalli

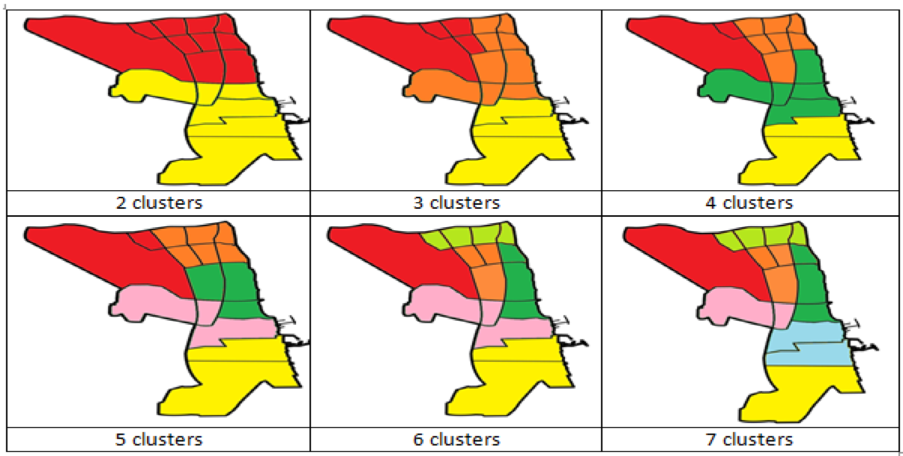

Fig.5. Application of K-means clustering Algorithm to AlAhmady

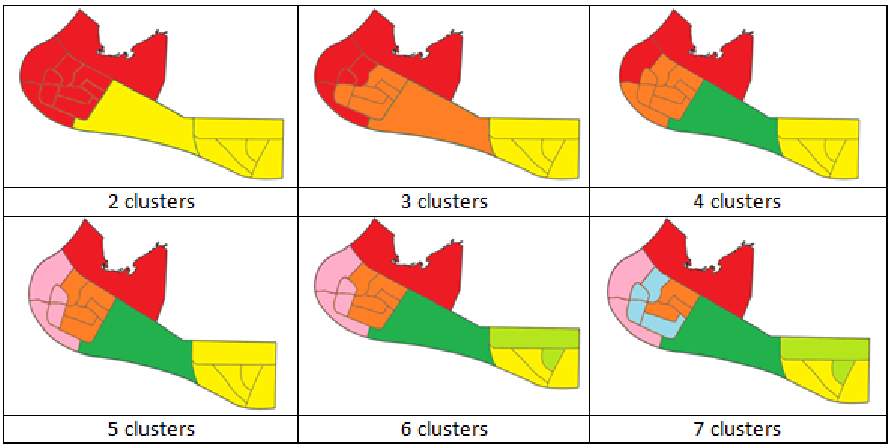

Fig. 6. Application of K-means clustering Algorithm to AlJahra

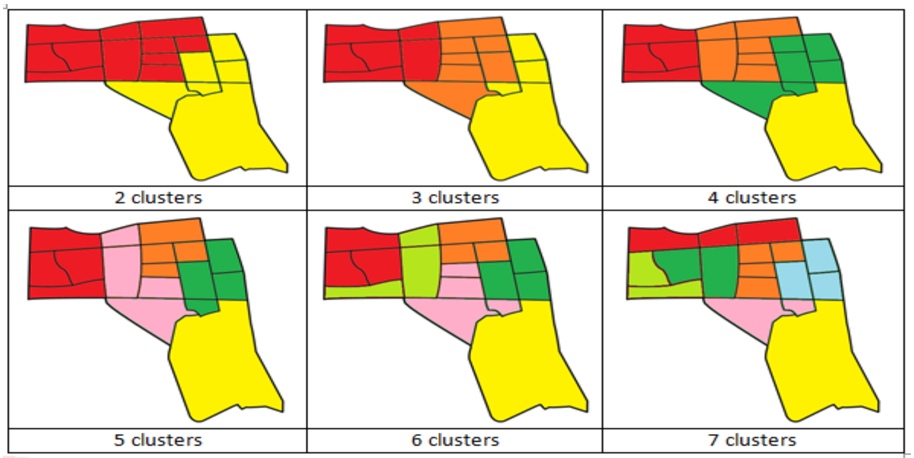

Fig. 7. Application of K-means clustering Algorithm to Farawnyia

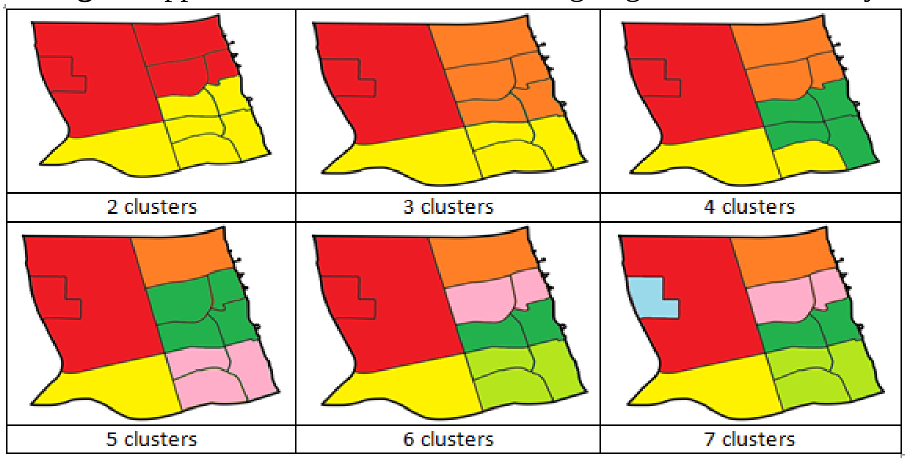

Fig. 8. Application of K-means clustering Algorithm to Mubarak Al-kebeer

3.2.2. Dynamic clustering algorithms

In real cases, sometimes you need the number of clusters to be dynamic according to the problem domain. Both proposed clustering algorithms (Dynamic A and Dynamic B) depend on choosing random points in governorates' cities. The starting city is randomly chosen and adjacent points which construct a hotspot area according to a specific threshold, where the user can change this threshold according to his needs. In Dynamic A, separated clusters are considered, where in Dynamic B overlapped clusters are taken into consideration. Dynamic A algorithm is illustrated in Table 5, where Dynamic B is a modified version of Dynamic A and the overlapping areas are covered as well. We applied these two algorithms to Kuwait governorates. Taking Mubarak El-Kebeer as an example, as shown in Table 6 and illustrated in Fig. 9, we have 4 cities in one cluster.

In order to illustrate the performance of the three algorithms: K-means, dynamic A, and dynamic B clustering algorithms, run time is measured in milliseconds. The performance study is run under a machine of Intel i3 processor and 4 GB RAM, using VB.net 2010. Since we have from 2-7 clusters, we run the three algorithms for the six governorates, as illustrated in Tables 8-13. In each table, we have the number of cities in governorate the capital, 14 cities in Hawalli, 16 cities in Al-Ahmady, 16 cities in Al-Jahra, 17 cities in Al-

JES, Assiut University, Faculty of Engineering, Vol. 43, No. 1, January 2015, pp. 1 – 15

Farawniya, and 11 cities in Mubarak Alkebeer. Results shown in all tables (Table 7-12) prove that K-means algorithm takes the highest time than the other two dynamic algorithms. In addition, dynamic A takes less time than dynamic algorithm. For example, having two clusters and governorate the capital, where the number of cities is 34, K-means algorithm takes 171ms, dynamic A is 80ms, and dynamic is 127ms.

Table 5.
Dynamic A and Dynamic B Algorithm Steps

Dynamic A Algorithm		**Dynamic B Algorithm**	
Input	G: number of governorates	**Input**	G: number of governorates
	m: number of cities		m: number of cities
	h’: hotspot candidate		h’: hotspot candidate
	c: point in a city (can be a cell)		c: point in a city (can be a cell)
	T: Threshold of hot spots		T: Threshold of hot spots
Output	H: hotspot array	**Output**	H: hotspot array
Process	For i=1 to G do	**Process**	For i=1 to G-1 do
	Begin		Begin
	T =0; M=0		T=0; M=0
	h’ = random point in city of G		h’ = random point in city of G
	For j= 1 to m do		For j= 1 to m do
	Begin		Begin
	If c is adjacent to h’		If c is adjacent to h’
	M= M ∪ c		M= M ∪ c
	T = T+ 1		T = T+ 1
	End		End
	For c Є M do		For c Є M do
	If T >=4 then		If ((T >= 4) and
			(overlap(h’), buffer(c,
			200)) =2 then
	H(i) = H(i) ∪ c		H(i) = H(i) ∪ c
	End		End

Table 6.
Clustered Cities in Mubarak Al-Kabeer governorate using dynamic A

City Name (EN)	City Name (AR)	Adjacent Serial	Cluster
Abu Fatera	أبو فطيرة	5	B
Al-Addan	العدان	10	D
Al-Funitees	الفنيطس	8	C
Al-Qurain	القرين	6	B
Al-Qosour	القصور	7	C
Al-Messila	المسيلة	9	C
Sabah Al-Salem	صباح السالم	11	D
Industrial Sabhan	صبحان الصناعية	1	A
Sabahn Central	صبحان الوسطى	2	A
Sabhan South Central	صبحان جنوب الوسطى	3	A
Mubarak Al-Kabeer	مبارك الكبير	4	B

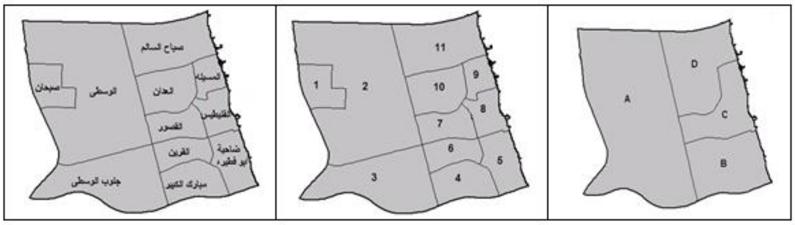

Fig. 9. Mubarak Al-Kebeer Clusters

Table 7.
Comparisons of K-means with dynamic algorithms for two clusters for 6 governorates

	Two Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	171 ms	80 ms	127 ms
Hawalli	14	149 ms	69 ms	111 ms
Al-Ahmady	16	143 ms	75 ms	113 ms
Al-Jahra	16	139 ms	73 ms	113 ms
Al-Farawaniya	17	141 ms	74 ms	112 ms
Mubarak	11	111 ms	40 ms	91 ms

Table 8.
Comparisons of K-means with Dynamic Algorithms for three clusters for 6 governorates

	Three Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	180 ms	101 ms	134 ms
Hawalli	14	139 ms	71 ms	121 ms
Al-Ahmady	16	131 ms	89 ms	126 ms
Al-Jahra	16	134 ms	88 ms	123 ms
Al-Farawaniya	17	134 ms	86 ms	123 ms
Mubarak	11	131 ms	52 ms	129 ms

Table 9.
Comparisons of K-means with Dynamic algorithms for four clusters for 6 governorates

	Four Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	209 ms	120 ms	189 ms
Hawalli	14	189 ms	100 ms	188 ms
Al-Ahmady	16	197 ms	113 ms	160 ms
Al-Jahra	16	199 ms	113 ms	140 ms
Al-Farawaniya	17	170 ms	112 ms	139 ms
Mubarak	11	152 ms	99 ms	129 ms

Table 10.
Comparisons of K-means with dynamic algorithms for five clusters for 6 governorates

	Five Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	217 ms	129 ms	200 ms
Hawalli	14	199 ms	107 ms	161 ms
Al-Ahmady	16	191 ms	119 ms	169 ms
Al-Jahra	16	191 ms	115 ms	161 ms
Al-Farawaniya	17	192 ms	115 ms	171 ms
Mubarak	11	160 ms	107 ms	145 ms

Table 11.
Comparisons of K-means with dynamic algorithms for six clusters for 6 governorates

	Six Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	220 ms	140 ms	201 ms
Hawalli	14	212 ms	120 ms	166 ms
Al-Ahmady	16	213 ms	125 ms	173 ms
Al-Jahra	16	211 ms	121 ms	165 ms
Al-Farawaniya	17	214 ms	122 ms	164 ms
Mubarak	11	180 ms	109 ms	145 ms

Table 12.
Comparisons of K-means with dynamic algorithms for seven clusters for 6 governorates

	Seven Clusters			
	Number of Cities	K-means	Dynamic A	Dynamic B
The Capital	34	229 ms	141 ms	209 ms
Hawalli	14	219 ms	127 ms	179 ms
Al-Ahmady	16	222 ms	131 ms	182 ms
Al-Jahra	16	221 ms	130 ms	179 ms
Al-Farawaniya	17	223 ms	130 ms	181 ms
Mubarak	11	199 ms	111 ms	149 ms

3.3. Crime hot spot identification phase

The objective of this phase is to identify crime hot spots in the six governorates mentioned previously. In this phase, since K-means algorithm takes the highest time to partition the cities within each governorate and its cluster numbers are static it will not be efficient to use it to find crime hot spots in Kuwait. In addition, because dynamic A clustering algorithm takes less time than dynamic B clustering algorithm and both are dynamic algorithms, we used dynamic A clustering algorithm to identify crime hot spots.

Thus, hot spots candidates are entered to a classifier, as shown in Fig.1, to detect the hot spot areas in each city within each governorate. If the user requires to check for an overlap of hotspot areas, dynamic B as it applies overlap of hot spots. Skurichina and Duin [24] suggested the usage of random subspace classifier, which we apply in this paper because of the random way of the unequal distribution of the datasets over the classes and the possible modification in the training phase. Details of random subspace classifier are explained in [24]. Let each training object X_i (i =1, ..., n) in the training sample set X = (X_1, X_2, . . ., X_n) be a p-dimensional vector X_i = (X_{i1}, X_{i2}, . . ., X_{ip}), described by p features (components). In the Random sampling method, one randomly selects r < p features from the p-dimensional data set X. One thus obtains the r-dimensional random subspace of the original p-dimensional feature space. Applying random subspace classifier for Kuwait datasets, 98% accuracy and 99.4% ROC are obtained. In addition, precision (98.7%), recall (98.4%), and F1 (98.28%) measures are computed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP+FP+FN}} \quad (3)$$

4. Results and discussion

The same six governorates mentioned previously are used in this study. Fig. 10 illustrates hot spots of one crime type, which is Act of Disgrace and humiliation. One can notice that this crime is a hot spot in Al-Farawnyia and cold spot in Mubarak Kebeer, as illustrated in Fig. 10.

Fig. 11 shows spatial crime distribution according to crime type. For example, act of discharge and humiliation is distributed as (49) records, adultery (18), aggravated assault (117), bribery (5), counter fitting (19), drug (118), embezzlement (6), fight or resist employee on job (6), forging of official documents (128), 6), weapon, robbery and attempted robbery (388), suicide and attempted suicide (18). Fig. 12 clarifies the time of the year where crimes get higher than other time of the year. For example, March and May have the highest level of crime of the year, then July and August, then April and June, as illustrated in Fig. 12.

5. Conclusions and future work

In this paper, we discussed K-means and dynamic clustering algorithms to identify crime hot spots. K-means algorithm has a static number of clusters and therefore we had to develop dynamic clustering algorithms to handle the situation of varied clusters. The proposed dynamic clustering algorithms are compared using real data of Kuwait governorates. K-means took the highest time, where dynamic A took less time than dynamic B. In addition, we identified the distribution of crime types over Kuwait governorates, the high time of crimes in Kuwait. We plan in the future to handle online maps. In addition, we will apply location-based services with our algorithm to determine monitoring crimes via moving objects tracking.

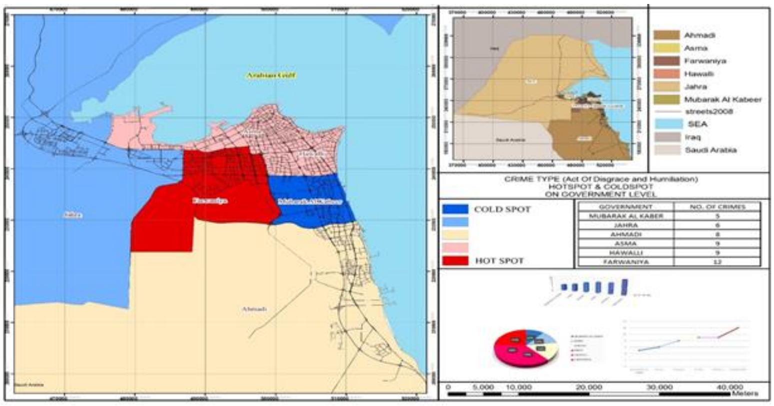

Fig. 10. Crime Hot Spots in Kuwait governorates

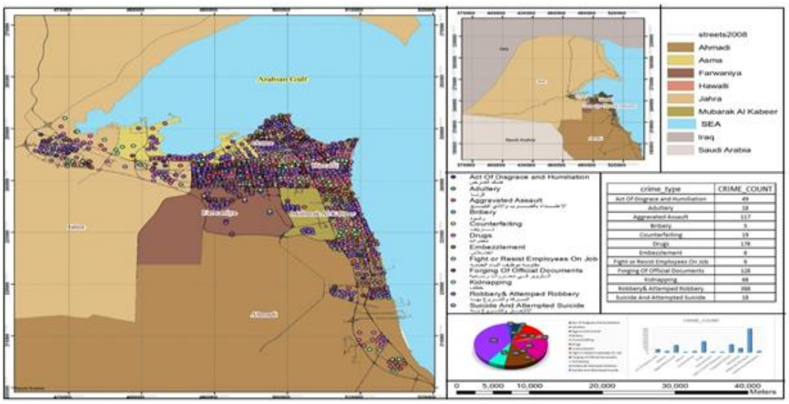

Fig. 11. Spatial Crime Distribution in Kuwait based on Crime Type

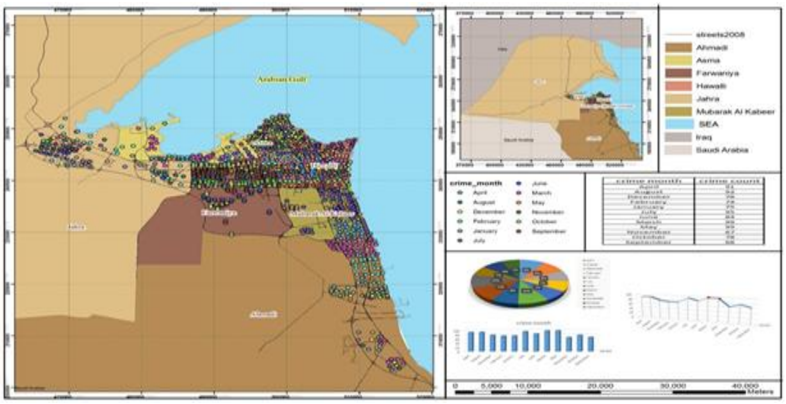

Fig. 12. Spatial Crime Distribution in Kuwait based on Time of the year

REFERENCES

[1] http://researcher.watson.ibm.com/researcher/view_group.php?id=4152 (last accessed: August 9, 2014)

[2] M. Keyvanpour, M. Javideh, and M. Ebrahimi, "Detecting and investigating crime by means of data mining: a general crime matching framework," WCIT 2010, Procedia Computer Science 3 (2011) 872–880.

[3] R. W. Adderley, "The use of data mining techniques in crime trend analysis and offender profiling," Ph.D. Thesis, University of Wolverhampton, Wolverhampton, England , 2007.

[4] H. Chen, W. Chung, J. Xu, G.Wang, Y. Qin, M. Chau, "Crime data mining: A general framework and some examples," Computer, 37(4), 50–56, 2004.
[5] O. de Vel, A. Anderson, M. Corney, G. Mohay, "Mining e-mail content for author identification forensics," SIGMOD Record, 30(4), 55–64, 2001.
[6] G. Wang, H. Chen, H. Atabakhsh, "Automatically detecting deceptive criminal identities," Communications of ACM, 47(3), 70–76, 2004.
[7] P. Phillips and I. Lee, "Areal aggregated crime reasoning through density tracing," In International workshop on spatial and spatio-temporal data mining in conjunction with IEEE international conference on data mining (pp. 649–654). Omaha, NE, USA, 2007.
[8] P.Phillips and I. Lee, "Mining co-distribution patterns for large crime datasets," Expert Systems with Applications," 39, 11556–11563, 2012.
[9] L. Zhu, D. Gorman, and S. Horel, "Hierarchical Bayesian spatial models for alcohol availability, drug hot spots and violent crime," International Journal of Health Geographics, 5: 54, 2006.
[10] H. Liu and D. Brown, "Criminal incident prediction using a point-pattern-based density model," International journal of forecasting, 19(4):603-622, 2003.
[11] Y. Xue and D. Brown, "Spatial analysis with preference specification of latent decision makers for criminal event prediction," Decision support systems, 41(3):560-573, 2006.
[12] M. Smith and D. Brown, "Application of Discrete Choice Analysis to Attack Point Patterns," Information Systems and e- Business Management, 5(3):255-274, 2007.
[13] D. Brown, J. Dalton, and H. Hoyle, "Spatial Forecast Methods for Terrorist Events in Urban Environments," Intelligence and security informatics: Second Symposium on Intelligence and Security Informatics, ISI 2004, June 10-11, 2004; Tucson, AZ, USA Springer-Verlag New York Inc; 426, 2004.
[14]A. Rodrigues, P. Diggle, and R. Assuncao, "Semiparametric approach to point source modelling in epidemiology and criminology," Journal of the Royal Statistical Society: Series C (Applied Statistics), 59(3):533-542, 2010.
[15] S. Guha, R. Rastogi, and K. Shim, "CURE: An Efficient Clustering Algorithm for Large Databases," in ACM SIGMOD Conference, 1998.
[16] R. Ng and J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," in Very Large Data Bases Conference, 1994.
[17] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," in ACM SIGMOD Conference, 1996.
[18] A. Jain and R. Dubes. Algorithms for Clustering Data. Prentice Hall, New Jersey, 1998.
[19] L. Kaufman and P. Rousseuw. Finding Groups in Data- An Introduction to Cluster Analysis, Wiley Series in Probability and Math. Sciences, 1990.
[20] T. Khonen, M. R. Shroeder, and T. S. Huang. Self-Organizing Maps. Springer-Verlag, New York, 2001.
[21] N. Levine, "Crime mapping and the Crimestat program," Geographical Analysis, vol. 38, no. 1, pp. 41–56, Jan. 2005.
[22] Y. Sun, C. Cui, S. Ke, and J. Lu, "Research on Dynamic Clustering Routing Considering Node Load for Wireless Sensor Networks," Scientific Research Journal, Vol. 5 No. 3B, PP. 508-511, 2013.
[23] A. Shafeek and K. S. Hareesha, "Dynamic Clustering of Data with Modified K-Means Algorithm," 2012 International Conference on Information and Computer Networks (ICICN 2012), pp.221-225, 2012.
[24] M. Skurichina and R. P. W. Duin, "Bagging, Boosting and the Random Subspace Method for Linear Classifiers," Pattern Analysis & Applications, Vol. 5, pp.121–135, 2002.

JES, Assiut University, Faculty of Engineering, Vol. 43, No. 1, January 2015, pp. 1 – 15

تطوير خوارزميات التجميع الديناميكى المكانية والزمانية لتحديد بؤر الجريمة فى الكويت

الملخص العربى:

مع زيادة معدل الجريمة على مستوى العالم أصبح من الضرورة بناء وتطوير خوارزميات ذات كفاءة للتعامل مع الوضع الحالى . يعتبر تحديد بؤر الجريمة عن طريق التجميع المكانى والزمانى من أهم النقاط البحثية . من خلال هذا البحث تم اقتراح خوارزمين للتجميع الديناميكى للبيانات المكانية والزمانية لتحديد بؤر الجريمة فى الكويت وتم تطبيق هذه الخوارزميات على أنواع جريمة مختلفة مثل الادمان ، الاذلال ، الزنا ، الهجوم عمدا ، الرشوة ، التصدى ، الاختلاس ، مقاومة الموظفين على العمل ، تزوير الوثائق الرسمية ، الأسلحة ، السرقة ، الانتحار ،سرقة البنوك فى محافظات الكويت : العاصمة ، حوالى ، الأحمدى ، الجهرة ، الفرعونية ، مبارك الكبير . ومن خلال تطبيق خوارزم التصنيف العشوائى المتفرع تم الحصول على النتائج التالية دقة 98.7% واسترداد 98.4% و 98.28% F1 و 99.4% ROC .