Crime Forecasting Using Data Mining Techniques

Chung-Hsien Yu[1], Max W. Ward[1], Melissa Morabito[2], and Wei Ding[1]
[1]Department of Computer Science, [2]Department of Sociology,
University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125
{csyu, ding}@cs.umb.edu, maxwward@gmail.com, melissa.morabito@umb.edu

***Abstract*—Crime is classically "unpredictable". It is not necessarily random, but neither does it take place consistently in space or time. A better theoretical understanding is needed to facilitate practical crime prevention solutions that correspond to specific places and times. In this study, we discuss the preliminary results of a crime forecasting model developed in collaboration with the police department of a United States city in the Northeast. We first discuss our approach to architecting datasets from original crime records. The datasets contain aggregated counts of crime and crime-related events categorized by the police department. The location and time of these events is embedded in the data. Additional spatial and temporal features are harvested from the raw data set. Second, an ensemble of data mining classification techniques is employed to perform the crime forecasting. We analyze a variety of classification methods to determine which is best for predicting crime "hotspots". We also investigate classification on increase or emergence. Last, we propose the best forecasting approach to achieve the most stable outcomes. The result of our research is a model that takes advantage of implicit and explicit spatial and temporal data to make reliable crime predictions.**

Categories and Subject Descriptors—H.2.8 [DATABASE MANAGEMENT]; Database Applications – Data Mining; Spatial Databases and GIS

General Terms—Experimentation

Keywords—Classification; Spatial Data Mining; Crime Forecasting

I. INTRODUCTION

Crime is neither systematic nor entirely random. It ebbs and flows with cycles of human behavior, but particular places are crime attractors. It is critical to identify the spatial and temporal patterns for a better understanding of crime events and to theorize their correlates. Using maps and time series data, practical crime prevention solutions can be developed that correspond to specific places and times. Spatial data mining is a uniquely qualified field to enable the analysis necessary to develop effective crime forecasting.

It is only within the last few decades that the technology necessary to make spatial data mining a practical solution for wide audiences of law enforcement officials has become affordable and available. Now it is quite reasonable, and common in fact, for larger police departments to have adequate computer hardware, data analysis software, and mapping tools that enable visualization of dense spatial data. It is also a recent development that the quantities of quality data needed to see the patterns in crime events over the course of significant socioeconomic cycles has been available. That too is now something that most police departments have available. Furthermore, technologists skilled in spatial data mining are now emerging.

The United States National Institute for Justice (NIJ) has sponsored research into crime mapping and forecasting for a number of years including a recent award to the University of Massachusetts at Boston. As part of the University's NIJ funded research, this paper focuses primarily on property crime which is identified as a research priority for the police department in a particular United States city in the Northeast. The goal of the project is to explore a methodology for reliably predicting the location, time, and/or likelihood of future residential burglary. Herein, you will find a review of the preliminary results of the University's collaboration with the city's police department. Due to the sensitivity of the data, the city will not be named.

First, we discuss how to generate architected data sets from original crime records. The architected data sets contain the aggregated counts of different types of crimes and related events as categorized by the city's police department. Spatial and temporal information pertaining to the crime data is embedded in these architected data sets. Second, several sophisticated data mining classification techniques are chosen to perform the crime forecasting. Finally, we analyze which classification approach is potentially the best method for predicting whether residential burglary will happen. We call an affirmative prediction a "hotspot". Along with occurrence prediction, we also explore predicting whether the crime will emerge or increase at certain locations. We say that this location is "heating up". Our experimental results demonstrate that through numerous combinations of classification methods and data feature sets, the best forecasting approach can be determined. Furthermore, our research provides a valuable look at the nature of how crime patterns organize and originate as well as how they exist in space and time.

This research paper is made up of the following consecutive sections: Data Generation, describes the data set; Approach Architecture, details the feature construction and data manipulation; Experimental Results, explores our analysis; Conclusion, reviews our research findings; and Deployment, discusses the motivation for our research and its intended use.

II. DATA GENERATION

The data used for this research was gleaned from a variety of city agencies. Each original data entry is a record for an individual crime or related event. Each record contains the type of event, the location in longitude and latitude, and the time and date of the incident. Before any data mining can begin, a preprocessing is needed to make it suitable for classification.

A. Data Grid

A police-department requirement for the deployment of this crime prediction model is that it forecast residential burglary over space and time. Accordingly, the model classifies burglaries monthly across a uniform grid. The grid divides the city into checkerboard-like cells. Within each cell, data is aggregated into six categories including Arrest, Commercial Burglary, Foreclosure, Motor Vehicle Larceny, Residential Burglary, and Street Robbery. Each cell is populated on a monthly basis.

Two resolutions of data were researched. One measured 24-by-20 square grid cells, the other 41-by-40. The cells in the 24-by-20 grid measure approximately one-half mile square. In the 41-by-40 grid, the distance is just over one-quarter mile square. In both cases, each monthly data set is a matrix of the six previously mentioned categories. The finer resolution allows the grid to be interrogated with a more detailed eye toward the spatial information inherent in the dataset. Conversely, the lower resolution has the effect of generalizing the spatial knowledge.

B. Empty Grid Cells

Empty grid cells have to be removed from the datasets because they have a detrimental yet counter intuitive side effect. They improve the performance of the classifiers. It is easy for any given classifier to correctly predict that nothing will happen in an empty grid cell. This "intelligence" is truly artificial.

An empty grid cell is defined as lacking any count in that cell in any of the investigated categories over the entire timeframe being analyzed. The majority of empty grid cells have two explanations. One, the boundaries of the city aren't rectangular like the grid being used is, and two, there are many areas within the city limits such as airport runways, bodies of water, and public open spaces where these events simply don't happen. The result is empty grid cells that have to be removed.

III. APPROACH ARCHITECTURE

The approach for our forecast model is classification. Mathematically, classification is the process of learning a function f that maps each attribute of a set $X=\{x_1, x_2, \ldots, x_n\}$ to a predefined class label y. A chosen algorithm works routinely to develop a model from the set of labeled data input. Classification creates the model that best relates the attribute set to the class label of the data. In the primary classification of this study, if the attribute set of category counts maps to a grid cell with at least one residential burglary, that set is labeled a "hotspot". Under an alternate classification, if the attributes map to a cell with an increasing number of burglaries, they are said to be "heating-up".

The attribute set used in this study is based on the Broken Windows Theory [1]. Accordingly, related categories of events are used to describe Residential Burglary. As previously noted, these categories are Arrest, Commercial Burglary, Foreclosure, Motor Vehicle Larceny, Residential Burglary, and Street Robbery. An explanation for how those events may affect Residential Burglary can be described by the following two factors:

1) Social norms and conformity: This sociological and social-psychological term has been defined as "the rules that a group uses for appropriate and inappropriate values, beliefs, attitudes and behaviors. The customary rules of behavior that coordinate our interactions with others" [2]. Hence, we add Street Robbery, Motor Vehicle Larceny, and Commercial Burglary for the feature set.

2) Social signaling and signal crime: Signal crime is based on the theory that certain crimes may act as a "signal" to a neighborhood that it is at risk [3]. Examples commonly given are bus shelters, foreclosures, and drug dealers. Hence, we add Foreclosure and Offender Arrest data for the feature set.

In criminological theory, offenders monitor other people and their environment to find opportunities for criminal activity. A disordered environment which has other crime incidents residential mobility is more likely to send the signal that this is a place to conduct crimes without being caught.

Considering this rationale, we use the basic attribute set and grid structure previously described to build a complex group of features out of the original crime data. These features become the modified attribute set employed in our classifications. They are designed to leverage and maximize the spatial and temporal qualities of the data set.

B. Feature Construction

1) Leveraging Temporal Knowledge (The t-Month Approach): The basic premise here is that a Residential Burglary that happened in one month can be described by events that came before it. In particular, a Residential Burglary is described by previous counts of Arrests, Commercial Burglaries, Foreclosures, Motor Vehicle Larcenies, Residential Burglaries, and Street Robberies. For instance, February's Residential Burglaries can be described by the events that happened in January, March can be described by February, and so on.

We can use events in January and February as training data, and then rely on March as test data. In the training process, we assume January events predict Residential Burglaries in February. For each area R_i (one of those grid cells), the six attributes of each set $X_i = \{x_1, x_2, \ldots, x_6\}$ (the set of event counts) in January will be used as training features, and Residential Burglary in February will be used as the training label y_i. Similarly, test data is constructed of the six attributes in March, and the test labels for evaluation of the classification are the Residential Burglaries that happen in April.

Fig. 1 illustrates a simplified training sample that has three crimes as its features. Each row indicates the count of three different crimes on the same 2-by-2 grid. Let *crime 1* be used as the class label. For the top left grid cell, R_1, red cells are its three features, $X_1 = \{4, 2, 7\}$, and the green cell is its label, $y_1 = 1$. Similarly we can have training samples for the other three grid cells at top right, bottom left, and bottom right: $R_2 = \{ X_2 = \{3, 1, 8\}, y_2 = 2\}$, $R_3 = \{ X_3 = \{2, 0, 4\}, y_3 = 3\}$, and $R_4 = \{X_4 = \{0, 0, 1\}, y_4 = 0\}$. In our study, we define a hotspot as a grid cell that has at least one incident of Residential Burglary. In this example R_1, R_2 and R_3 are hotspots while R_4 is not. In a t-month based feature set where $t = 1$, R_1 is represented by a vector X_1 with label y_1.

A t-month vector, when $t \geq 2$, is achieved by concatenating previous month's vectors. In Fig. 2, the 2-month based feature set for cell $R1$ in February, call it X_{21}, consists of twelve features from January and February (highlighted in green) with a label from March (highlighted in yellow). Therefore, $R_1 = \{X_{21} = \{4, 2, 7, 1, 4, 1\}, y_{21} = 0\}$.

The t-month approach has the effect of decreasing the training sample size as the number of concatenated month increases. With twelve months of data, twelve sets of 1-month-based vectors can be produced; eleven sets of 2-month-based vectors can be produced, and so on. The total sets of t-month-based vectors, S, can be calculated with this function, $S = Y - t + 1$, where Y is the total months of available data. Later on in our experiments, these vectors sets are used to form training datasets as well as test datasets for classification.

2) Maximizing Spatial Knowledge (Neighborhood Averaging): Spatial autocorrelation, a primary tenant of the study of spatial data, indicates "characteristics at proximal locations appear to be correlated, either positively or negatively." [4] Recognizing that events in each grid cell may be influenced by neighboring cells, an eight-neighborhood average, known as the Moore neighborhood,[5] is used in our study to store one grid's spatial neighborhood knowledge. We count how many neighbors a grid has, calculate the total count of each event in the neighborhood, including the grid in question, and calculate the average count for the neighborhood. The average is calculated in the standard way by dividing the total count by the number of neighbors plus 1.

Figure 1. Examples of how to construct features class labels using a 2-by-2 data grid.

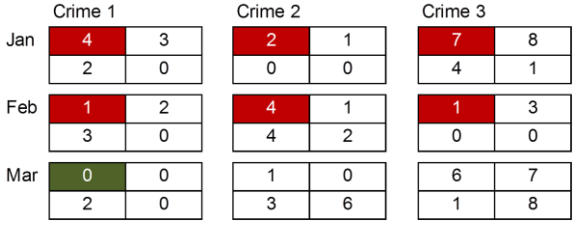

Figure 2. Examples of how to construct features class labels using a 2-by-2 data grid.

C. Balancing Data

One challenge in crime prediction, similar to other rare event prediction, is that hotspots and cold spots are unbalanced. That is cold spots are much more prevalent than hotspots. In our dataset, this is especially true with the higher resolution 41-by-40 grid. This has the result of confusing the necessary measures of precision, recall, and $F1$. In particular, the $F1$ score of hotspots is far lower than the $F1$ score of cold spots because the classifiers are well trained on cold spots. The calculation on $F1$ score in our study is defined as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \; where$$

$$Precision = \frac{TP}{TP + FP} \; and$$

$$Recall = \frac{TP}{TP + FN} \; and$$

$$TP = the\ number\ of\ True\ Positives\ (true\ hotspots\ predictions),$$

$$FP = the\ number\ of\ False\ Positives\ (false\ hotspots\ predictions), and$$

$$FN = the\ number\ of\ False\ Negatives\ (false\ cold\ spot\ predictions)$$

To resolve this issue, we adjust the weight of hotspots and cold spots. By increasing the weight of hotspots based on the ratio between hotspots and cold spots, the data set can be balanced before the classification process. The weight function is defined by the following:

$$w = \begin{cases} C/H, & for\ hotspots \\ 1, & for\ coldspots \end{cases}, where$$

$$C = Total\ number\ of\ cold\ spots, and$$

$$H = Total\ number\ of\ hotspots$$

For example, if there are 100 hotspots and 400 cold spots in one training data set, the weight of hotspots will be set as 400/100 = 4, and the weight of cold spots will be set as 1. Therefore, the weight of each data entry in the training dataset is reset before the whole training dataset is used to train the classifier. As a result, a misclassified hotspot will result in greater penalty compared to a misclassified cold spot. This guides the chosen classifier to focus on hotspot classification.

It should be noted that even though this increases the *F*1 score of hotspot classification, the overall accuracy of the prediction is decreased because of the resulting misclassification of cold spots. We accept the trade off on accuracy to have a higher *F*1 score on hotspots as the police department understandably has a greater vested interest in knowing that an area will be a hotspot.

IV. EXPERIMENTAL RESULTS

Our experiments include several classifiers: One Nearest Neighbor (1NN) and a location constrained variation, Decision Tree (J48), Support Vector Machine (SVM) with radial basis function as the kernel type, Neural Network (Neural) with 2-layer network, and Naive Bayes (Bayes) [6].

A. 1NN as a Baseline

We employ a traditional 1NN and a variation with a location constraint. These classification methods establish a baseline upon which subsequent methods are measured. In the traditional approach, the algorithm finds the most similar data vector entry using Euclidean distance. The formula is defined as follows:

Let X_i *be a vector with* p *features such that* $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. *Let* n *be the total number of input samples and let* d *be the distance between each sample. The Euclidean distance* d *between two grid cells,* X_i *and* X_l, *where (*i, l = *1, 2, … ,* p*) is defined as:*

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \cdots + (x_{ip} - x_{lp})^2}$$

In the standard approach, the algorithm searches the entire dataset, without regard for location or time, for the first grid cell with the shortest Euclidean distance. When the search is location constrained, time differences are still allowed, but the algorithm limits the search to the grid with the same location.

Fig. 3 clearly shows the location constrained algorithm gives a better result in precision, recall, *F*1, and overall accuracy, where

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ grid\ cells}$$

The experiment illustrated in Fig. 3 makes use of a 1-month based feature set, and the 24-by-20 data grid. The classifiers are trained on 11 months of data, and tested on one.

It can be said that 1NN makes a simple assumption, namely that similar circumstances must result in similar outcomes. The location constrained 1NN, however, is not so naive. It capitalizes on the well-known notion that spatial data is heavily influenced by its place. Separate the notion and our analysis of this particular dataset from its mathematical explanation, and what a location constrained 1NN algorithm tells you is what any law enforcement official might: neighborhoods well known for residential burglary in the past are likely to experience residential burglary in the future. Where the 1NN establishes naive baseline, the location constrained version establishes a wise one that far more sophisticated algorithms find difficult to overcome.

B. Performance Comparison

The following tables (Table I, Table II) detail the experimental results of four different classifiers using *t*-month-based features where *t* = 1, 2, … , 10. As mentioned previously, there are *S* sets of *t*-month-based vectors produced during the feature construction process. In these tables, the 24-by-20 data grid is used. The test set is always the last of those *S* sets of vectors, and the remainders are used as the training set.

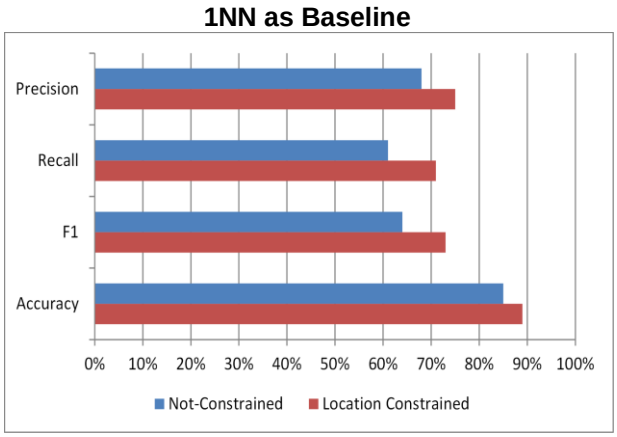

Figure 3. 1NN results with and without the location constraint.

TABLE I. ACCURACY (%) ON CLASSIFICATIONS USING DIFFERENT CLASSIFIERS WITH *T*-MONTH-BASED FEATURES, WHERE $T = 1, 2, \ldots, 10$.

Accuracy	Classification Methods			
	SVM	J48	Neural	1NN
1-month	87.70%	86.25%	**87.91%**	84.58%
2-month	87.91%	**88.33%**	87.50%	85.20%
3-month	87.08%	86.25%	**89.16%**	82.91%
4-month	85.83%	86.25%	**88.95%**	84.58%
5-month	**86.25%**	**86.25%**	85.62%	85.41%
6-month	85.62%	86.04%	**88.12%**	86.87%
7-month	86.25%	85.20%	**88.54%**	87.50%
8-month	85.41%	**86.87%**	20.41%	86.25%
9-month	84.37%	85.83%	**90.62%**	87.08%
10-month	84.37%	85.00%	**88.75%**	87.70%

TABLE II. *F*1 ON CLASSIFICATIONS USING DIFFERENT CLASSIFIERS WITH *T*-MONTH-BASED FEATURES, WHERE $T = 1, 2, \ldots, 10$.

F1	Classification Methods			
	SVM	J48	Neural	1NN
1-month	70.64%	67.00%	**71.84%**	62.62%
2-month	71.56%	71.42%	**71.69%**	65.36%
3-month	71.02%	67.00%	**72.34%**	60.95%
4-month	69.36%	65.62%	**71.35%**	62.62%
5-month	**70.00%**	67.32%	69.60%	63.91%
6-month	69.33%	68.24%	**71.64%**	67.69%
7-month	70.53%	65.02%	**72.36%**	68.42%
8-month	**69.82%**	68.65%	32.50%	64.51%
9-month	67.81%	68.22%	**74.86%**	69.00%
10-month	68.35%	63.26%	**73.00%**	69.10%

J48, Neural, and SVM classifiers consistently outperform the naive 1NN approach. Moreover, Neural often performs slightly better than J48 and SVM. This shows the strength of neural networks when modeling complex systems.

Another interesting finding which is notable not because of its success, but rather its failure, is that the Neural approach fails dramatically at the 8-month point. This suggests that the algorithm may be picking up on some kind of seasonal pattern that it does not quite know what to do with. Indeed, the data shows that the eighth month in the data set experiences a low point in residential burglary.

Introducing the Naïve Bayes classifier to our experiment, we make an assumption similar to that made when using the location constrained 1NN. We claim that where residential burglary has occurred once there is a higher chance of it happening again. Our experiments show that our assumption is valid. Using the Naive Bayes classifier yields better results than Neural Networks. The results are illustrated in Fig. 4. A 9-month-based feature set is used in experiment illustrate by Fig. 4 because it yielded the highest accuracy and *F*1 in previous experiments. The 24-by-20 data grid is used in this experiment.

A "Leave-One-Month-Out" (LOMO) approach is adopted this time. Instead of running the classification only once on one set of training and test data, the LOMO approach is used to run the classification on S - 1 sets. Recall that the number of vector sets used in the t-month approach can be expressed as $S = Y - t + 1$, where Y is the total months of available data and t is the number of months concatenated in the feature vectors. The LOMO approach works as follows. When $t = 9$ and $Y = 12$, we have $S = 4$. This means that there are 4 sets of vectors, s_i, where $i = 1, 2, \ldots, 4$, with 9-month-based features produced during the feature construction. With $S = 4$, three groups of training and test data can be paired as following: {s_1 as training, s_2 as test}, {s_2 as training, s_3 as test}, and {s_3 as training, s_4 as test}. As a result, three

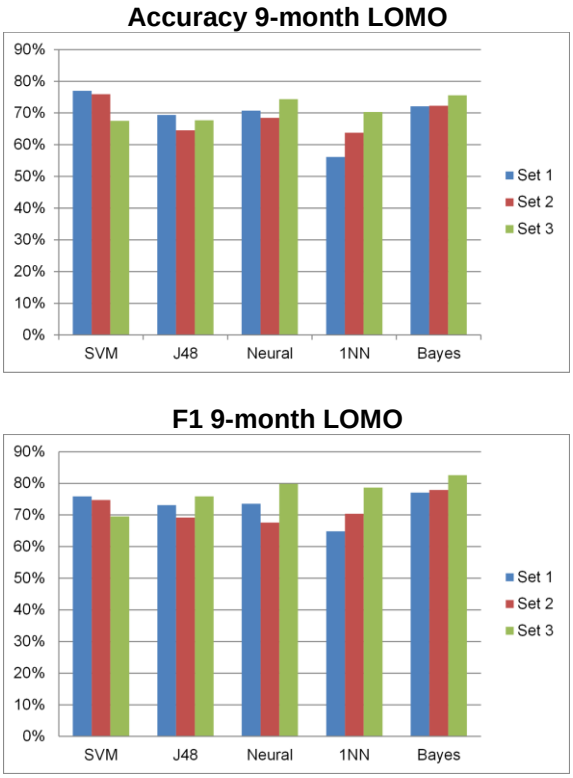

Figure 4. Accuracy and *F*1 scores using SVM, J48, Neural, 1NN, and Bayes.

classification runs can be performed with these three pairs of datasets.

C. Classification Comparison

Two final classifications, predicting hotspots and increases, are performed on the two investigated dataset resolutions, 24-by-20 and 41-by-40. The classifications use two feature sets, one that incorporates the calculated

neighborhood spatial feature set and one that does not. With two different feature sets, and two different classifications we have the following four different scenarios:

41 - ORG - HOT, 41 - NB - HOT, 41 - ORG - INC, 41 - NB - INC

In these abbreviated labels, 41 refers to the 41-by-40 data set, ORG and NB differentiate between the original feature set and the one with explicit spatial features added, and HOT and INC indicate the two classifications predicting hotspots and increases. The labels can be seen in results obtained from experimentations on these four scenarios, depicted in Fig. 5. For simplicity of demonstration, we average the accuracy and F1 score yield from the five different classifiers, 1NN, J48, Neural, SVM, and Naïve Bayes.

Regarding the 41-by-40 data set, the charts indicate that the accuracy of predicting increases is a little higher than predicting hotspots. When considering *F*1 scores, predicting increases on average is about 10% lower than predicting hotspots. This indicates many more non-increase grid cells than increase cells in the data set. The ratio is about 5 to 1. As a result, even when balancing the data, the classifiers used in our experiment are affected. The higher accuracy on predicting increases is due to the classifiers being more accurate at detecting the larger number of non-increase grid cells. We also find that using spatial features helps achieve higher accuracy when predicting hotspots. Surprisingly, these features do not help predict increase, except when using 1-month-based features.

Using the 24-by-20 data set we recognized the same observation which is that adding spatial features improves the prediction only when fewer month features are concatenated in the *t*-month approach. We assume that this is simply a natural characteristic of the classifiers (i.e. adding more features in the training data makes the classification process more complicated and less reliable). Consequently, the results can't be improved upon once the number of features reaches a certain threshold. A possible solution to this issue would be to select a smaller group of the most valuable features. More investigation is needed to identify which features will prove most useful.

D. Ensemble Learning (Voting)

In an attempt to stabilize our results as well as to further improve classifier performance, the voting method is adopted. The basic idea of voting is to use multiple classifiers trained on the same data set. Then, the label of an instance is decided by majority vote from the classifiers. For example, if two out of three classifiers predict one grid as a hotspot, then the grid is labeled as a hotspot.

In our experiment, a heterogeneous set of classifiers are selected to perform the voting: SVM, Neural, and Bayes. These classifiers exhibit a varying ability to detect subtleties in the data. As expected, the voting effect does stabilize the outcomes somewhat. Fig. 6 illustrates the results.

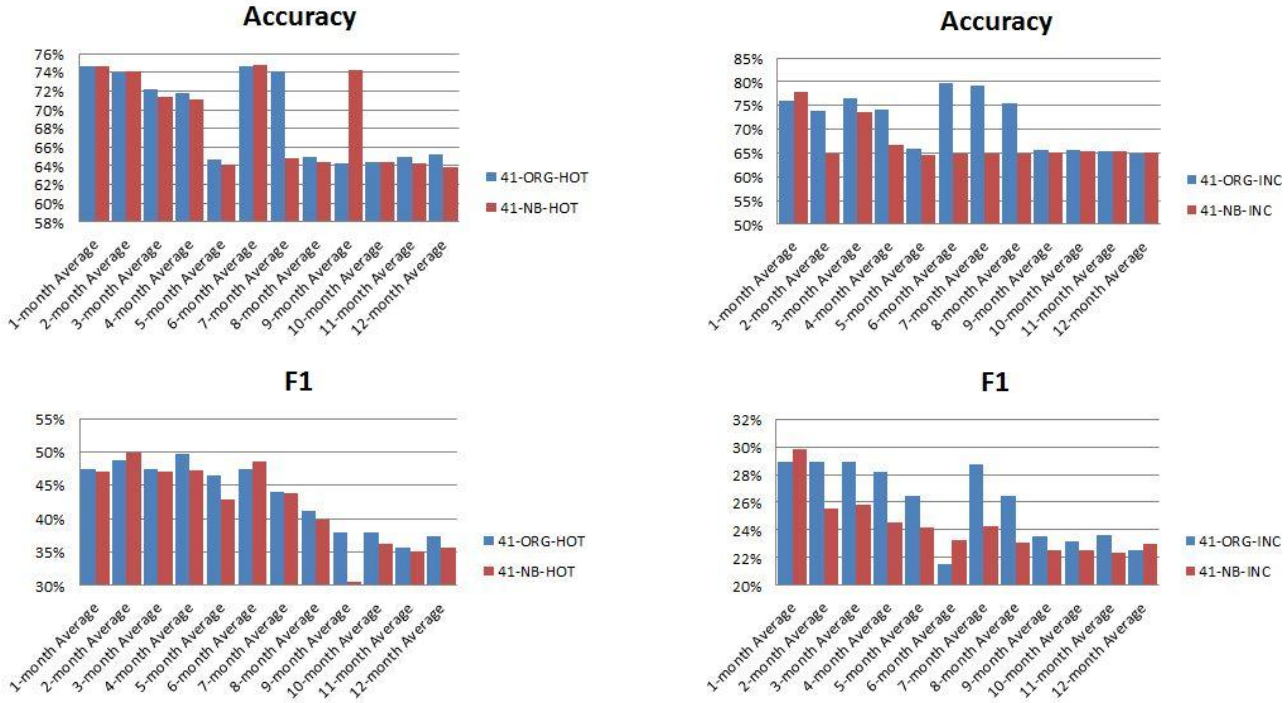

Figure 5. Accuracy and *F*1 scores using the 41-by-40 grid dataset. (ORG-without spatial features. NB-with spatial feature. HOT-predict hotspots. INC-predict heating up hotspots.)

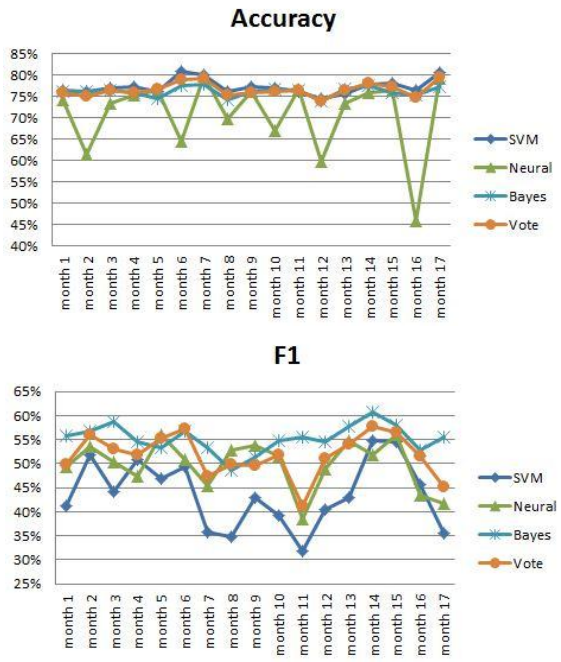

Figure 6. Accuracy and F1 scores of using voting with 41-by-40 grid data, 6-month based features, and the LOMO approach to make 17 predictions. The charts show that voting stabilizes the outcomes somewhat.

V. CONCLUSION

A. Location Constraint and Spatial Knowledge

Three results in our experiments point to the value of leveraging the spatial knowledge inherent in the crime data set. The first and most obvious is the success of the simple 1NN classifier modified with a location constraint. Finding the most similar circumstance within the same neighborhood proved more effective than finding it within the entire city. The second indicator is the success and stability of the probability-based Naive Bayes classifier. The basic logic of the location constrained 1NN is not unlike that of Naive Bayes: namely that what has happened in a particular place in the past is likely to recur. The third result pointing to spatial knowledge is in the 24-by-20 grid data. Our success measures are consistently higher when using the lower resolution data set. We believe this is specifically due to each grid cell exhibiting a broader spatial knowledge. The challenge of future research will be to locate the optimum point at which spatial knowledge is most keen.

B. Classification

An overall observation on every classification method employed that is particularly interesting is that the more complex algorithms don't vastly improve upon a simple and straight forward, location-constrained, nearest neighbor approach. See Fig. 7 for a summary

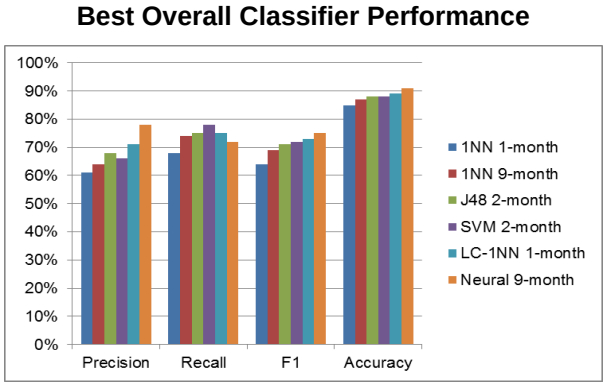

Figure 7. Overall classification results. Best performance of classifier using different training sets.

of several of our previously conducted classifications. The chart legend is ordered by descending *F*1 score. Our location-constrained nearest neighbor (LC-1NN 1-month) approach using a 1-month based feature set is second only to the Neural Network approach (Neural 9-month) using a 9-month based feature set.

C. Grid Size

After the experiments on two different data grids, 24-by-20 and 41-by-40, we concluded that the 24-by-20 grid consistently gives us better results than 41-by-40 grid (Fig. 8). Better performance in coarse resolution indicates that insufficient information can be collected

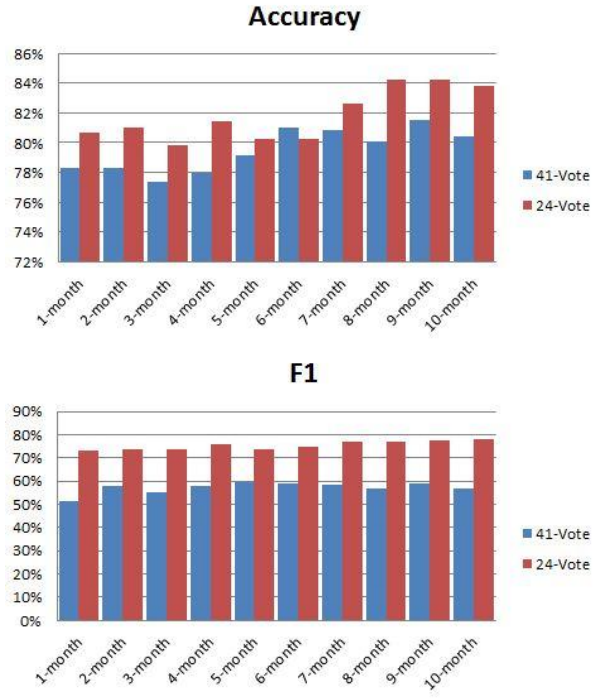

Figure 8. Accuracy and *F*1 scores of using two different grid sizes. The charts show that the 24-by- 20 grid always achieves better results than the 41-by-40 grid.

on a finer granularity. The finer grid represents the crime counts of a small area resulting in crime counts that are much lower than in the larger grids. The lower counts appear to inhibit the classifiers.

D. Predicting Hotspot Increases

Predicting that burglary will increase in the next month is somehow more difficult than predicting a hotspot using either data resolution. In this city, where crime patterns appear to be fairly stable, the initiative of predicting an increase would be more useful than just knowing where the crime will happen. Future research might focus on improving the increase prediction.

VI. DEPLOYMENT

Currently, we are in the process of preparing to deploy this system for the police department. Work conducted until this point was funded by a planning grant which enabled us to apply data mining techniques to the historical data. With additional funding, we will be able to incorporate these techniques into the routine work of the crime analysts within the department.

A major barrier faced in preparing our experiments was that all of the data used were not actually housed in the police department. We had to collect and combine data from a variety of city agencies. To be able to use current data to make future predictions, we will have to automate systems that bring the data from these other agencies together and then classify it for the police department promptly. Unfortunately, data from local governments is often not readily accessible or well organized.

Once these data sharing systems are in place, we have developed a plan with the police department to disseminate the results of the prediction models. Command staff will use the predictions from these data as tools for resource allocation decisions which are made on a weekly and sometimes daily basis depending on city needs. Specifically, predictions will be provided to police executives to make staffing decisions during weekly deployment meetings. Given the low violent crime rate of this city, command staff are able to make deployment decisions based on intelligence. They have been unable to use the same intelligence to allocate resources to address property crime. Property crime comprises more than 70 percent of crimes reported to the police but receives little media or research attention. Accordingly, at the urging of the police department, our analyses began with residential burglary. In the future, we intend to expand our work to include motor vehicle theft as well violent crimes, such as street robbery and assault.

ACKNOWLEDGMENTS

The work is partially supported by the DOJ award #2009-DE-BX-K219.

REFERENCES

[1] George Kelling and Catherine Coles. Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities, ISBN: 0-684-83738-2.

[2] Steven N. Durlauf and Lawrence E. Blume (Eds), 'Social Norms' in New Palgrave Dictionary of Economics, Second Edition, London: Macmillan, 2011.

[3] Martin Innes (2003). Crime as a Signal, Crime as a Memory, Journal for Crime, Conflict and the Media, vol 1, pp 15-22.

[4] De Knegt; H.J.; F. van Langevelde; M.B. Coughenour; A.K. Skidmore; W.F. de Boer; I.M.A. Heitkönig; N.M. Knox; R. Slotow; C. van der Waal and H.H.T. Prins (2010). Spatial autocorrelation and the scaling of species–environment relationships. Ecology 91: 2455–2465. doi:10.1890/09-1359.1

[5] Weisstein, Eric W. "Moore Neighborhood." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/MooreNeighborhood.html

[6] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, and Addison Wesley. Introduction to Data Mining (2006) ISBN: 0-321-321136-7