

# Empirical Discriminative Tensor Analysis for Crime Forecasting

Yang Mu<sup>1</sup>, Wei Ding<sup>1</sup>, Melissa Morabito<sup>2</sup>, Dacheng Tao<sup>3</sup>,

<sup>1</sup> Department of Computer Science, University of Massachusetts Boston, 100 Morrissey Boulevard Boston, MA 02125, US

{YangMu, Ding}@cs.umb.edu

<sup>2</sup> Department of Sociology, University of Massachusetts Boston, 100 Morrissey Boulevard Boston, MA 02125, US

Melissa.Morabito@umb.edu

<sup>3</sup> Department of Computer Science, University of Technology Sydney, Broadway NSW 2007, Australia

Dacheng.Tao@uts.edu.au

**Abstract.** Police agencies have been collecting increasing amount of information to better understand patterns in criminal activity. Recently there is a new trend on using the data collected to predict where and when crime will occur. Crime prediction is greatly beneficial because if it is done accurately, police practitioner would be able to allocate resources to the geographic areas most at risk for criminal activity and ultimately make communities safer. In this paper, we discuss a new four-order tensor representation for crime data. The tensor encodes the longitude, latitude, time, and other relevant incidents. Using the tensor data structure, we propose the Empirical Discriminative Tensor Analysis (EDTA) algorithm to obtain sufficient discriminative information while minimizing empirical risk simultaneously. We examine the algorithm on the crime data collected in one Northeastern city. EDTA demonstrates promising results compared to other existing methods in real world scenarios.

**Keywords:** Discriminative tensor analysis, tensor least square, crime forecasting.

## 1 Introduction

Using machine learning and data mining technologies to reliably and accurately predict where and when crime will occur is greatly beneficial. If it is done accurately, practitioners would be able to allocate resources to the geographic areas most at risk for criminal activity and ultimately make communities safer.

Many aspects will influence the crime prediction task:

1. The spatial area. For example, a residential area is more likely to have residential burglaries than an industrial area.
2. The temporal dimension. For example, university vacation period or holidays may be more vulnerable for certain crimes than other time periods.

---

3. The influence from other crime relevant events or other crime types. For example, increasing arrested offenders may call for better security management in this area to reduce the future criminal events.

It is desirable to combine these factors to better predicate a crime. A simple solution is to quantize all the information in these three factors and use one single vector to represent the data. The disadvantages of this representation are obvious: 1) the dimensionality of the features is always much higher than the number of data which can easily be haunted by the small sample size problem [1] and 2) the geometry structure among these three aspects cannot be well preserved.

In this paper, we present a fourth-tensor representation to describe a crime. A fourth-order tensor encodes: longitude, latitude, time and other relevant events. The tensor representation can rarely occur the small sample size problem because each order of this tensor have lower dimension and have much less chance to be greater than the sample size. In addition, the geometry structure is well preserved in tensor form.

There are existing algorithms which accept tensor as input directly instead of vectorizing the data. Sun et al. proposes the Dynamic Tensor Analysis (DTA) and Streaming Tensor Analysis (STA) methods to cope with the increasing data number problem. In addition, Sun also generalizes the basic tensor analysis problem to higher order which minimizes the reconstruction error between the input tensor and the desired output core tensor. Similar ideas for second- and third-order tensors are proposed in [2, 3, and 4]; Yan et al, generalize the Linear Discriminant Analysis (LDA) to tensor generalization form to accept tensor data directly. Tao et al, propose a framework for tensor convex optimization [14].

Our previous work [5] proposed a three-way DLA (TWDLA) method to reveal the discriminative information from third-order tensor features. (1) It accepts three-way data as input directly so the structure information can be well preserved. (2) It models the local geometry over every modality of the input third-order tensors, so the spatial relations of input tensors within a class can be preserved. (3) The margin between classes is maximized over each modality of the tensor to achieve the best discriminative performance. (4) It avoids the under sampling problem.

However, TWDLA and other discriminative methods [2, 3, and 7] omit the empirical reconstruction error while extracting the discriminative information to reconstruct the data. The low reconstruction error reflects the reconstruction with low chance to have a mistake. The desired core tensor is the indicator tensor which incorporates the ground truth information and is expected to be generated from an optimal algorithm. The reconstruction procedure is a standard least square problem which minimizes the Frobenius norm between the projected core tensor and the core tensor. This optimization procedure is an explicit way to minimize the empirical risk.

In this paper, we propose a new method denoted as Empirical Discriminative Tensor Analysis (EDTA) which calculates the discriminative projection matrix and considers the empirical reconstruction error simultaneously. The discriminative component is a higher order generalization for TWDLA. The empirical reconstruction error minimization component is obtained by minimize the reconstruction error. EDTA simultaneously optimizes these two components and makes the final projection encapsulate the discriminative information and also minimize the empirical

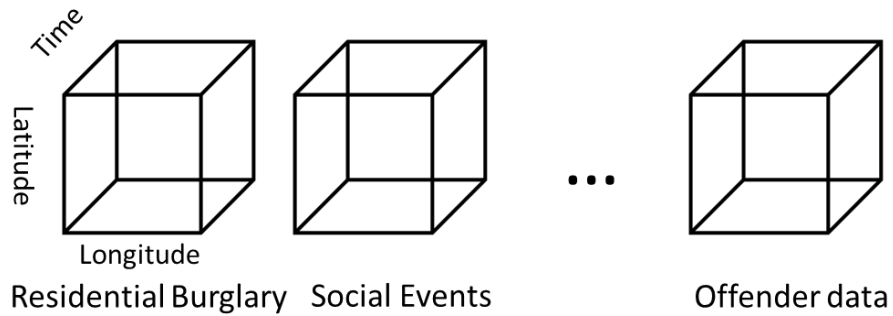
reconstruction risk. A similar way by combining these aspects were proposed to find the sparse optimal solution on vector approaches [8].

The rest of the paper is organized as follows. Section 2 introduces crime problem and data representation. We give a brief introduction of tensor algebra in Section 3. Our proposed Empirical Discriminative Tensor Analysis is introduced in Section 4. Our experimental results and conclusions are presented in Sections 5 and 6 respectively.

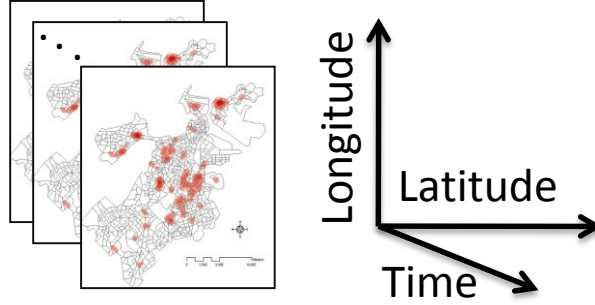
## 2 Problem Formulation and Data representation

In this paper, we aim to predict residential burglaries. Each residential burglary is encoded by spatial and temporal information, including longitude, latitude, and time. We also collect other relevant geo-coded events selected by domain scientists that are believed to be associated with crime tendency, including construction permits, foreclosure, mayor hotline inputs, motor vehicle larceny, social events, and offender data. Crime data is rasterized into small grid cells because it is infeasible to make precise longitude and latitude coordinate predication. The number of residential burglaries for a specific grid cell is the summation of all the crimes happened inside this grid cell. We aggregate the data by month and perform monthly prediction because daily crime data is too few and cannot provide sufficient features from the crime data collected in this Northeastern city. Therefore, the ultimate objective of this crime forecasting task is to predict whether a grid cell will have high residential burglaries for a given month.

As illustrated in Figure 1, we design a data structure of fourth-order tensor to capture features of each residential burglary with respect to three aspects, spatial aspect, temporal aspect, and other relevant events.



**Figure 1.** An example of residential burglary in a fourth-order tensor. The figure depicts fourth-order tensor which is represented by a row of elements. Each element in this fourth-order tensor is a third-order tensor of relevant events, such as residential burglary, social events, offender data, etc. Each relevant event in a third-order tensor is presented by longitude, latitude, and time.



**Figure 2.** The residential burglary third-order tensor example. Each map refers to a residential burglary map in different time. The combination of these maps by time forms a three order tensor.

Figure 1 explains general structure of fourth-order tensor. In a fourth-order tensor, different relevant features in third-order tensors are organized together without losing the geo-spatial structure. Figure 2 further explains the internal structure of third-order tensor using residential burglary as an example. The third-order tensor contains the crimes information of the whole city area among all the collected time periods. The crime in one grid cell at a specific month can always be influenced by the neighboring area and the relevant events in recent months. Therefore, the representation of one crime in one grid cell is a small tensor extracted from the big tensor. The small tensor has smaller spatial and temporal dimension size, because only neighboring area and recent data could influence the crimes of this grid cell.

According to the sociological study [10, 11, and 12], the number of crimes happened in one area always related to  $t$  previous month information and its surrounding area. Waldo Tobler [13] also pointed out that, the first law of geography is "Everything is related to everything else, but near things are more related than distant things." , A naive model to predict the crime is just copy the previous month data or average the neighbor area data. We denoted the surrounding area data of the  $t$  previous month with different relevant events as the *fourth-order tensor features* to predict a crime incident.

### 3 Tensor Algebra

In this section, we give a brief introduction of tensor algebra definitions which are related to this paper. Suppose  $\mathbf{X}$  is an  $M$  order tensor and  $\mathbf{X} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$ . The size of  $i$ th mode of  $\mathbf{X}$  is  $L_i$ . Then we can have the following definition.

**Definition 1 (Mode-d Matrix unfolding).** The mode-d matrix unfolding of  $\mathbf{X}$  is the set of vectors in  $\mathbb{R}^{L_d}$  obtained by keeping the index  $i_d$  fixed and varying the other indices. Therefore, the mode-d matrix unfolding of an  $M$ th-order tensor is a matrix  $\mathbf{X}_{(d)} \in \mathbb{R}^{L_d \times \bar{L}_d}$ , where  $\bar{L}_d = \prod_{i \neq d} L_i$ . We denote the mode-d matrix unfolding operation as  $\text{mat}_d(\mathbf{X})$ .

Normally, Mode-d Matrix unfolding is applied to convert a tensor to matrix along d mode. For example, we may convert a fourth-order crime tensor along the time mode, and then each row vector of the converted matrix encodes all the value at a specific time.

**Definition 2 (Tensor Contraction).** The contraction of a tensor is obtained by summing over all values of the repeated indices for the equating two indices. In general, for tensors  $\mathbf{X} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M \times L'_1 \times L'_2 \times \dots \times L'_M}$  and  $\mathbf{Y} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M \times L''_1 \times L''_2 \times \dots \times L''_M}$ , the contraction on  $\mathbf{X}$  and  $\mathbf{Y}$  is:

$$\begin{aligned} & \llbracket \mathbf{X} \otimes \mathbf{Y}; (1:M)(1:M) \rrbracket \\ &= \sum_{l_1=1}^{L_1} \dots \sum_{l_M=1}^{L_M} \mathbf{X}_{l_1 \times l_2 \times \dots \times l_M} \mathbf{Y}_{l_1 \times l_2 \times \dots \times l_M} \end{aligned}$$

Noted that, only orders consisted in  $(1:M)$ , the same indices shared by two tensors, can be used for contraction.

If the contraction operation is conducted on all indices excluding  $d$  on  $\mathbf{X}$  and  $\mathbf{Y}$  in  $\mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$ , the procedure is denoted as:

$$\begin{aligned} & \llbracket \mathbf{X} \otimes \mathbf{Y}; (\bar{d})(\bar{d}) \rrbracket \\ &= \sum_{l_1=1}^{L_1} \dots \sum_{l_{i-1}=1}^{L_{i-1}} \sum_{l_{i+1}=1}^{L_{i+1}} \dots \sum_{l_M=1}^{L_M} \mathbf{X}_{l_1 \times \dots \times l_{i-1} \times l_{i+1} \times \dots \times l_M} \mathbf{Y}_{l_1 \times \dots \times l_{i-1} \times l_{i+1} \times \dots \times l_M} \\ &= \text{mat}_d(\mathbf{X}) \text{mat}_d^T(\mathbf{Y}) \end{aligned}$$

**Definition 3 (Mode-d product).** The mode-d product of  $\mathbf{X} \times_d U$  of tensor  $\mathbf{X} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$  is an  $L_1 \times \dots \times L_{d-1} \times L'_d \times L_{d+1} \times \dots \times L_M$  tensor and matrix  $U \in \mathbb{R}^{L'_d \times L_d}$ . The procedure is denoted as:

$$\mathbf{X} \times_d U = \sum_{l'_d} (\mathbf{X}_{l_1 \times \dots \times l_{d-1} \times l_d \times l_{d+1} \times \dots \times l_M} U_{l'_d \times l_d})$$

Mode-d product changes the dimension of Mode-d in  $\mathbf{X}$  by multiplying it with a matrix.

To simplify the notation in this paper, we denote

$$\mathbf{X} \times_1 U_1 \times_2 U_2 \dots \times_M U_M \triangleq \mathbf{X} \prod_{k=1}^M \times_k U_k$$

**Definition 4 (Frobenius Norm).** The Frobenius norm of a tensor  $\mathbf{X} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$  is given by

$$\begin{aligned} \|\mathbf{X}\|_{\text{Fro}} &= \sqrt{\llbracket \mathbf{X} \otimes \mathbf{X}; (1:M)(1:M) \rrbracket} \\ &= \sqrt{\sum_{l_1=1}^{L_1} \dots \sum_{l_M=1}^{L_M} \mathbf{X}_{l_1 \times l_2 \times \dots \times l_M}^2} \end{aligned}$$

Frobenius norm, sometimes is also called the Euclidean norm when applied on vector or matrix.

## 4 Empirical Discriminative Tensor Analysis

Suppose there are  $N$  samples which come from  $N$  grid cell. Each sample is an  $M$  order tensor,  $\mathbf{X}_i \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$ . We want to learn a set of projection matrices  $U_k \in \mathbb{R}^{L_k \times L_k'}$ ,  $k = 1, 2, \dots, M$ ,  $L_k > L_k'$  and  $L_k'$  is the intrinsic dimensions for the  $k$ th order. Intrinsic dimensions are the needed dimensions to represent the event and incorporate the sufficient discriminative information for classification. Then the projected core tensor  $\mathbf{Y}_i = \mathbf{X}_i \prod_{j=1}^M U_j$ , where  $\mathbf{Y}_i \in \mathbb{R}^{L_1' \times L_2' \times \dots \times L_M'}$ , is the projected core tensor which not only contains sufficient discriminative information but also minimizes the empirical risk. The desired core tensor  $\mathbf{Z}_i \in \mathbb{R}^{L_1' \times L_2' \times \dots \times L_M'}$  incorporates ground truth label information which is the indicator tensor.  $\mathbf{Y}_i$  is expected to have as much discriminative information as possible. The difference between tensors  $\mathbf{Y}_i$  and  $\mathbf{Z}_i$  stands for the empirical reconstruction error.

### 4.1 Discriminative tensor analysis

The projected core tensor is desirable to have enough discriminative information for classification. This also means same class tensors which correspond to the samples in the same class are expected to be as close as possible and the different class tensors should be further from each other. We use Frobenius Norm in Definition 4 of the subtraction between two tensors to represent the distance of the two tensors. For any sample  $\mathbf{X}_i$ , the projected core tensor  $\mathbf{Y}_i$  is expected to have small distance between tensors in the same class and have large distance between tensors in different classes. Then our expectation of the core tensor can be described using the following objective function:

$$\min \left( \sum_{p=1}^{k_1} \|\mathbf{Y}_i - \mathbf{Y}_{i^p}\|_{\text{Fro}}^2 - \beta \sum_{q=1}^{k_2} \|\mathbf{Y}_i - \mathbf{Y}_{i_q}\|_{\text{Fro}}^2 \right) \quad (1)$$

where  $\beta$  is the tradeoff between two parts.  $i^p$  is the  $p$ th element of  $\mathbf{Y}_i$ 's  $k_1$  nearest neighbors in the identical class and  $i_q$  is the  $q$ th element of  $\mathbf{Y}_i$ 's  $k_2$  nearest neighbors in the different class. The nearest neighbors of  $\mathbf{Y}_i$  forms a local patch with index matrix  $F_i = [i, i^1, i^2, \dots, i^{k_1}, i_1, i_2, \dots, i_{k_2}]$ . Eq.(1) states that for any projected core tensor  $\mathbf{Y}_i$ , the distance between its  $k_1$  nearest neighbors in the identical class is minimized while the distance between its  $k_2$  nearest neighbors in the different class is maximized.

Eq. (1) is equal to:

$$\min \sum_{p=1}^{k_1+k_2+1} \sum_{q=1}^{k_1+k_2+1} L_i(F_i(p), F_i(q)) \llbracket \mathbf{Y}_{F_i(p)} \otimes \mathbf{Y}_{F_i(q)}; (1:M)(1:M) \rrbracket \quad (2)$$

where  $L_i \in \mathbb{R}^{(k_1+k_2+1) \times (k_1+k_2+1)}$  encapsulates the geometry structure and discriminative information of the local patch of input tensor  $\mathbf{X}_i$ . Define the coefficient vector  $\omega_i = \left[ \overbrace{1, 1, \dots, 1}^{k_1}, \overbrace{-\beta, -\beta, \dots, -\beta}^{k_2} \right]$ , then

$$L_i = \begin{bmatrix} \sum_{j=1}^{k_1+k_2} (\omega_i)_j & -\omega_i^T \\ -\omega_i & \text{diag}(\omega_i) \end{bmatrix}.$$

$F_i(p)$  is the  $p$ th element of index matrix  $F_i$ . According to the tensor contraction definition in Definition 2, Eq. (2) is equivalent to:

$$\begin{aligned} & \min \text{tr} \left( U_k^T \left( \sum_{p=1}^{k_1+k_2+1} \sum_{q=1}^{k_1+k_2+1} L_i(F_i(p), F_i(q)) \llbracket \mathbf{Y}_{F_i(p)} \otimes \mathbf{Y}_{F_i(q)}; (\bar{k})(\bar{k}) \rrbracket \right) U_k \right) \\ &= \min \text{tr} \left( U_k^T \left( \sum_{p=1}^{k_1+k_2+1} \sum_{q=1}^{k_1+k_2+1} L_i(F_i(p), F_i(q)) \text{mat}_k(\mathbf{Y}_{F_i(p)}) \text{mat}_k^T(\mathbf{Y}_{F_i(q)}) \right) U_k \right) \\ &= \min \text{tr} (U_k^T L_i^k U_k) \end{aligned} \quad (3)$$

where,  $L_i^k = \sum_{p=1}^{k_1+k_2+1} \sum_{q=1}^{k_1+k_2+1} L_i(F_i(p), F_i(q)) \text{mat}_k(\mathbf{Y}_{F_i(p)}) \text{mat}_k^T(\mathbf{Y}_{F_i(q)})$ .

And  $U_k$  is the projection matrix of order  $k$ .

Eq. (3) gives the optimal solution of a local tensor patch because the objective function defined in Eq. (1) only consider the core tensor  $\mathbf{Y}_i$  and its nearest neighbors. The ultimate goal is to make all the core tensors have the same hallmark as Eq. (3). Therefore, we consider other core tensors as well by summing over all local optimizations together. The whole alignment is given according to Eq. (3):

$$\min \text{tr} (U_k^T L_k U_k) \quad (4)$$

where  $L_k = \sum_{i=1}^N L_i^k$  and it stores the discriminative information and geometry structure.

## 4.2 Empirical error minimization

Indicator matrix [9] indicates the class label of samples by a matrix. Using the indicator matrix can easily transform the empirical reconstruction error minimization problem to least square problem. If directly using class label as the indicator matrix, the problem would be further converted to a standard regression problem which is a special case of least square problem. In this paper, we use a ground truth indicator tensor which has the same size as the core tensor. Then, the empirical reconstruction error corresponds to the distance between the core tensor and the indicator tensor.

The indicator tensor  $\mathbf{Z}_i$  is defined by retrieving the principle components of each order. Because the desired core tensors in the same class are expected to be as close as possible, we use the average value of the tensors in the same class to stand for them. Indicator tensor  $\mathbf{Z}_i$  has the same size with  $\mathbf{Y}_i$ .

$$\begin{aligned}
& \min \sum_{i=1}^N \|Z_i - Y_i\|_2^2 \\
& = \min \sum_{i=1}^N \left\| Z_i - X_i \prod_{j=1}^M \times_j U_j \right\|_{\text{Fro}}^2
\end{aligned} \tag{5}$$

There is no close form solution for Eq. (5), [6] shows Eq.(5) is equivalent to:

$$\begin{aligned}
& \max \text{tr} \left( U_k^T \left( \sum_{i=1}^N \sum_{j=1}^N \text{mat}_k(Y_i) \text{mat}_k^T(Y_j) \right) U_k \right) \\
& = \min \text{tr} (U_k^T A U_k)
\end{aligned} \tag{6}$$

where  $A = \sum_{i=1}^N \sum_{j=1}^N \text{mat}_k(Y_i) \text{mat}_k^T(Y_j)$  and the constant item is ignored in optimization without loss of generality.

### 4.3 Final optimization

Our proposed empirical discriminative tensor analysis (EDTA) method tries to optimize both the discriminative objective function and the empirical error. Therefore, when combining the discriminative and empirical estimation objection function from Eq. (4) and Eq. (7), we can get

$$\begin{aligned}
& \min \text{tr} (U_k^T L_k U_k) - \lambda \text{tr} (U_k^T A U_k) \\
& = \min \text{tr} (U_k^T (L_k - \sigma A) U_k) \\
& = \min \text{tr} (U_k^T L_k^* U_k)
\end{aligned} \tag{8}$$

where  $L_k^* = (L_k - \sigma A)$ , and we expect  $U_k^T U_k = I$  which means the projection is orthogonal.

Solutions of Eq. (8) are given by using standard eigen-decomposition:

$$L_k^* \mathbf{u}_k = \lambda \mathbf{u}_k. \tag{9}$$

The column vectors ordered according to the increasing order of eigenvalues are the solutions of Eq. (9).

It is necessary to point out that  $\llbracket Y_{F_i(p)} \otimes Y_{F_i(q)}; (\bar{k})(\bar{k}) \rrbracket$  in Eq. (3) and  $\text{mat}_k(Y_i)$  in Eq. (6) are relied on all the projection matrices  $U_i$ ,  $i \neq k$ . All the projection matrices should be initialized before use.

For any projection matrix  $U_k$ , the optimization function is convex, such as Eq. (8). The alternating projection provides the final solution for the problem defined in Eq. (8).

The alternating projection procedure is illustrated in Table 1.



---

*Input:* Training tensor  $\mathbf{X}_i \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_M}$ , the desired core tensor  $\mathbf{Z}_i \in \mathbb{R}^{L_1' \times L_2' \times \dots \times L_M'}$ , and the maximum number of training iteration T.

*Output:* The projection matrix  $U_k \in \mathbb{R}^{L_k \times L_k'}$ ,  $k = 1, 2, \dots, M$ ,  $L_k > L_k'$ , constrained by  $U_k^T U_k = I$ , and the output desired core tensor  $\mathbf{Y}_i \in \mathbb{R}^{L_1' \times L_2' \times \dots \times L_M'}$

---

*Initialization:* Set  $U_k \in \mathbb{R}^{L_k \times L_k'}$ ,  $k = 1, 2, \dots, M$  to 1.

Step1. **FOR** t = 1 **TO** T{

Step2.       **FOR** k = 1 **TO** M{

                  Calculate  $U_k$  according to Eq. (8)

                  } // For loop in Step 2.

                  Check convergence: the train stage of EDTA converges if

Step3.        $Err(t) = \sum_{j=1}^M \|U_j^T U_j - I\| \ll \varepsilon$ .

                  } // For loop in Step 1.

Step4.  $\mathbf{Y}_i = \mathbf{X}_i \prod_{j=1}^M U_j$

---

**Table 1 Alternating Projection Optimization Procedure for EDTA**

## 5 Experiments

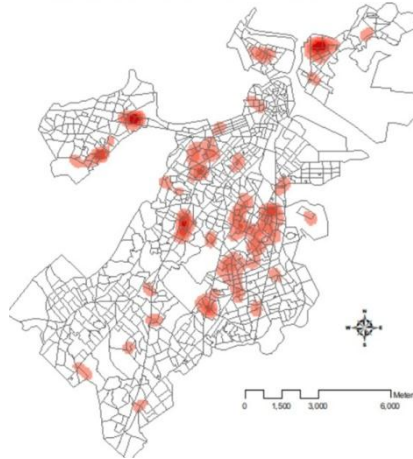
We evaluate the EDTA algorithm on Crime Forecasting in the Northeast city (Figure 3). The crimes happened in 2006 are the pre-knowledge. We select data in 2006 as training samples to test the crime incidents in 2007. We used 20 by 20 grid cells to represent the whole city area. For one grid cell, we expect to predict whether this area will take place any crimes. The grid cell with more than 1 crime is denoted as hot spot and the grid cell with 0 or 1 crime is denoted as cold spot. Thus the crime forecasting problem is a binary classification problem in this project.

We compare the EDTA method with other representative methods. three-way DLA (TWDLA) [5] extracts the discriminative information and accepts third-order tensors as input. We generalize it to accept fourth-order tensors in order to make a fair comparison. We denoted the variation of this algorithm as Four-Way DLA (FWDLA). Offline Tensor Analysis (OTA) is proposed in [6] which minimizes the distance between the projected core tensors with the desired core tensors

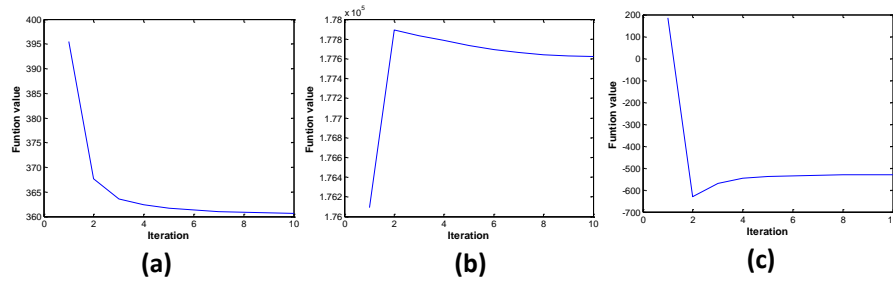
We set the tradeoff value  $\beta$  to 0.1 which gives less weight for the second term of Eq. (1), because tensors in different classes in our crime data always have large distance, and then we eliminate this influence.  $\sigma$  in Eq. (8) is automatically decided by the algorithm to make the trace of  $A$  as the same as the trace of  $L_k$ . The parameter settings of other algorithms are the same as those in their original papers.

Figure 4. shows the convergence status for our proposed EDTA. Figure 4(a) is the objective function value of Eq. (1) , Figure 4 (b) is the objective function value of Eq. (5) and Figure 4 (c) shows the objective function value of Eq. (8). We expect to minimize the function in Eq. (1) and maximize the function in Eq. (5) simultaneously. From Figure 4, we observe that in each optimization procedure, the projection matrices learned from objective function Eq. (8) optimize Eq. (1) and Eq. (5) as well. The derivative of the function value shown in these figures is the convergence rate.

Figure 5. shows the comparison results for 2007 Residential Burglary prediction. FWDLA only considers the discriminative information while OTA only consider reconstruction error. From this figure, we observe that the EDTA could achieve a better result on 12 months' prediction by considering both aspects which is consistent with our assumption.



**Figure 3.** The ground truth of Residential Burglary crime in Boston area. The darker area means higher number of crimes.



**Figure 4.** Function value versus the Iteration of EDTA. It shows the convergence status. (a) is the function value of Eq. (1). (b) is the function value of Eq. (5). (c) is the final function value of Eq. (8).

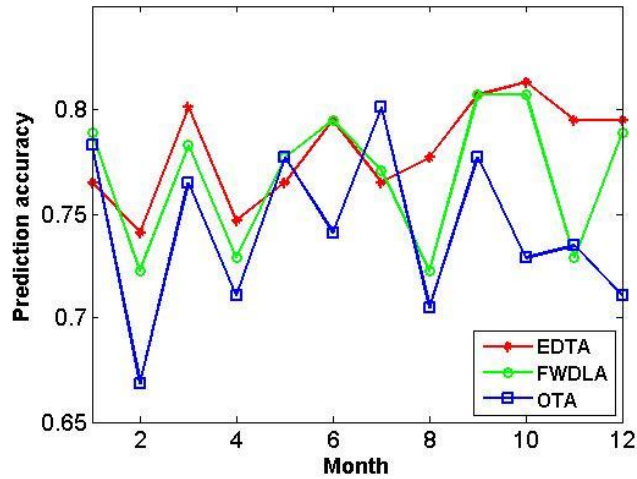


Figure 5. Methods comparison on residential burglary prediction for 12 months in 2007.

## 6 Conclusion

Crime forecasting problem has great impact on sociology and has not been well studied. We studied this problem with machine learning technologies. In this paper, we propose a fourth-order tensor representation to describe a crime which can better represent the crime by preserving the geometry structure, reducing the small sample risk and reduce the conflicts risk. We further propose an Empirical Discriminative Tensor Analysis (EDTA) method to accept tensor data as input directly. EDTA can reveal the discriminative information and minimize the empirical reconstruction risk simultaneously. The discriminative information is obtained by minimize the Frobenius norm of tensors in same class while separate the tensors in different class. Empirical risk minimization is obtained by a standard least square form to minimize the Frobenius norm between transformed tensors and the desired indicator tensors. The experimental results conducted on Boston Residential Burglary data show that EDTA outperforms the vector based method and the methods which learn the discriminative information and minimize the empirical risk respectively.

## Acknowledgement

The work was funded by the National Institute of Justice (No.2009-DE-BX-K219).

## References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition, second edition. Boston: Academic Press, (1990).
2. He X, Cai D, Niyogi P.: Tensor subspace analysis. *Advances in Neural Information Processing Systems*. 2006;18:499.
3. Xu D, Yan S, Zhang L, Zhang H.: Concurrent Subspaces Analysis. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005;00(c):203-208.
4. Kroonenberg P, Leeuw J.D.: Principal component analysis of three-mode data by means of alternating least squares algorithms
5. Mu Y, Tao D, Li X, Murtagh F. Biologically Inspired Tensor Features. *Cognitive Computation*. 2009;1(4):327-341.
6. Sun J, Tao D, Faloutsos C. Beyond streams and graphs: dynamic tensor analysis. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM; 2006:374–383.
7. Yan S, Xu D, Yang Q, et al. Discriminant Analysis with Tensor Representation. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005:526-532.
8. Zhou T, Tao D, Wu X. Manifold elastic net: a unified framework for sparse dimension reduction. *Data Mining and Knowledge Discovery*. 2010;22(3):340-371.
9. Ye J. Least squares linear discriminant analysis. *Proceedings of the 24th international conference on Machine learning - ICML '07*. 2007:1087-1093.
10. Deadman, D.: Forecasting residential burglary. *International Journal of Forecasting*: 19(4) 567-578 (2003)
11. Harries, Richard. "Modelling and predicting recorded property crime trends in England and Wales—a retrospective." *International Journal of Forecasting*, 2003: 557-566.
12. Harries, Gorr and Richard. Introduction to crime forecasting. *International Journal of Forecasting*, 2003: 551-555.
13. Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240.
14. Tao D, Li X, Wu X, Hu W, Maybank SJ. Supervised tensor learning. *Knowledge and Information Systems*. 2007;13(1):1-42.