

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/278622360>

Uma análise estatística dos indicadores de criminalidade em Salvador

Article · January 2008

READS

31

6 authors, including:



Aloisio Machado

Universidade Estadual de Feira de Santana

24 PUBLICATIONS 98 CITATIONS

[SEE PROFILE](#)



Rogério Hermida Quintella

Universidade Federal da Bahia

38 PUBLICATIONS 35 CITATIONS

[SEE PROFILE](#)



Uma análise estatística dos indicadores de criminalidade em Salvador¹

Jair Sampaio*
 Aloísio Machado da Silva Filho**
 Rogério Quintella***
 Gilney Figueira Zebende****

INTRODUÇÃO

A sociedade contemporânea apresenta uma complexa gama de contradições e conflitos que é evidenciada por meio de fatores como urbanização acelerada, desemprego, má distribuição de renda, serviços precários de educação, saúde, saneamento, habitação, segurança, etc. Estes fatores estão diretamente relacionados aos problemas sociais, cujo interesse constitui a tônica predominante entre estudiosos do desenvolvimento social.

No universo de problemas sociais, talvez seja, hoje, a criminalidade aquela que desperta maior atenção pela

¹ A primeira versão deste texto foi publicada nos Anais do XI Encontro de Modelagem Computacional realizado em novembro de 2008 em Volta Redonda – Rio de Janeiro.

* Estatístico e doutorando em Administração Pública (UFBA). Pesquisador da Superintendência de Econômicos e Sociais (SEI). jairsoaresjr@gmail.com

** Estatístico e mestrando em Modelagem Computacional e Tecnologia Industrial (Senai/Cimatec). Técnico da Superintendência de Estudos Econômicos e Sociais (SEI). aloisioestatistico@yahoo.com.br

*** Professor do Núcleo de Pós-Graduação em Administração (UFBA). rogerio@ufba.br

**** Professor do programa de Mestrado Interdisciplinar em Modelagem Computacional e Tecnologia Industrial (Senai/Cimatec). gfzebende@hotmail.com

ameaça que representa ao desenvolvimento social. Diante dessa nova realidade, a sociedade brasileira parece reconhecer que o aparelho estatal tem falhado no cumprimento de uma das suas funções básicas: oferecer segurança aos seus cidadãos.

Ademais, a segurança pública é também um problema político, visto que os meios necessários para reprimir os crimes envolvem a participação efetiva e integrada dos governos e a aplicação de recursos para esse fim. Felizmente, parece haver consenso entre pesquisadores e governantes que o crescimento acentuado dos crimes no Brasil tem alcançado proporções alarmantes. Esses fatores, aliados à necessidade de monitorar as transformações sociais daí resultantes, fizeram ressurgir o interesse e as possibilidades de aplicação de novas métricas para subsidiar a tomada de decisão por parte de governantes e organismos públicos.

Diante da problemática da criminalidade, mensurar este fenômeno no tempo e no espaço passou a ser uma das preocupações entre os estudiosos no assunto. Nessa perspectiva, este trabalho tem como objetivo medir a correlação de longo alcance em séries temporais de registros oficiais de homicídio doloso e furto ou roubo de veículos, divulgados pela Secretaria de Segurança Pública do Estado da Bahia (SSP-BA), utilizando o *Knowledge Discovery in Databases* (KDD), mais especificamente a função associação e a técnica *Detrended Fluctuation Analysis* (DFA).

KNOWLEDGE DISCOVERY IN DATABASES

O KDD engloba as etapas que produzem conhecimentos a partir de dados relacionados e sua principal característica é a extração não-trivial de informações e conhecimentos implicitamente contidos em uma base de dados. Essas informações e conhecimentos são, usualmente, de difícil detecção por métodos tradicionais de análise, sendo também típicos e potencialmente úteis na tomada de decisões (FRAWLEY; PIATETSKY-SCHAPIRO; MATHEUS; FAYYAD; PIATETSKY-SCHAPIRO; SMYTH, 1992) Assim, enquanto os métodos tradicionais são capazes de tratar apenas as informações explícitas, o KDD é capaz de

detectar informações armazenadas nas bases de dados, transformando-as em conhecimento.

O KDD é composto por um conjunto de etapas que, em geral, podem ser reunidas em três fases: preparação, análise e interpretação (ADRIAANS; ZANTIGE, 1996; BRACHMAN; ANAND, 1996; BRACHMAN et al, 1996; FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; HAN; KAMBER, 2000). Todas essas fases são críticas, sendo usualmente a fase de análise a mais complexa. Ela compreende, entre outras, a etapa de mineração de dados, que tem como objetivo encontrar padrões nos dados armazenados. Esta etapa é freqüentemente confundida na literatura com o próprio processo de KDD (HAN; KAMBER, 2000).

Na literatura brasileira da área, o que se observa é uma quase total ausência de trabalhos com abordagem de KDD em dados públicos. Praticamente toda a pesquisa nacional tem sido desenvolvida com enfoque em mineração de dados e *business intelligence*. Uma exceção é o trabalho em que Soares Junior e Quintella (2003, p.89) descrevem o KDD de forma muito simplificada como "o processo não-trivial para geração de conhecimento a partir da busca sistemática de padrões em grandes volumes de dados". Tal definição confunde-se com o entendimento geral do que é *data mining* (DM).

Como já mencionado, o DM é tratado como uma das etapas da descoberta de conhecimento em bases de dados. Reconhece-se, no entanto, que nem todo processo de *data mining* é conduzido em um contexto de KDD. Han e Kamber (2000, p.8) conceituam *data mining* como "uma etapa na descoberta do conhecimento em bancos de dados que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis".

As técnicas empregadas em DM podem ser divididas em dois grandes grupos: heurísticas e matemáticas. Entre as heurísticas, as redes neurais artificiais são as que mais se destacam, seguidas da inteligência artificial simbolista. Dentro do grupo da matemática, por sua vez, destacam-se a análise estatística e a modelagem matemática.

Os algoritmos de DM mais empregados são comumente divididos em cinco funções: classificação, regressão, associação, clustering e sumarização. Para este estudo serão empregadas as funções associação, regressão e sumarização.

A função associação identifica relações significativas existentes entre os eventos ocorridos em determinada ocasião baseada em modelos de dependência. Cabena e outros (1998) descrevem a função associação como o processo de interconexão de objetos, na tentativa de expor características e tendências. Os principais métodos são regras de associação e característica seqüencial.

A regressão é similar à classificação, diferencia-se desta por objetivar a predição de um valor real em vez de um atributo nominal ou uma categoria. Atualmente, as ferramentas de análise de regressão são encontradas nos diversos níveis de plataformas de computação, até mesmo no popular MS-Excel. Existem, no entanto, outros modelos de regressão mais complexos, envolvendo maior número de variáveis explicativas e relacionamentos não-lineares, e entre eles destacam-se a regressão múltipla (RLM), o probito e a regressão não-linear.

Por fim, a sumarização engloba a organização e o resumo dos dados. Visa, principalmente, a orientar e motivar análises posteriores mais complexas. Pode ser relacionada à estatística, mais especificamente à análise exploratória de dados ou à estatística descritiva. A sumarização utiliza as funções complementares de caracterização e visualização para observar a presença de alguma característica estrutural nos dados.

“A função associação identifica relações significativas existentes entre os eventos ocorridos em determinada ocasião baseada em modelos de dependência”

Como um exercício da aplicação do KDD, elegeu-se neste artigo a temática referente à modelagem e mensuração da criminalidade na cidade de Salvador, utilizando-se da série de 2003 a 2007 dos registros diários oficiais de homicídio doloso (Código Penal, art. 121) e furto ou roubo de veículos (Código Penal, art. 155 e art. 157, respectivamente), divulgados pela Secretaria de Segurança Pública do Estado da Bahia (SSP-BA).

PROCEDIMENTOS METODOLÓGICOS

Nesta seção, descreve-se resumidamente o procedimento utilizado para analisar, utilizando o *Knowledge Discovery in Databases* (KDD), a série temporal de três dos mais representativos indicadores de criminalidade da capital baiana. A série histórica estudada baseou-se em dados oficiais e secundários do Centro de Documentação e Estatística Policial (Cedep), publicados pela Secretaria de Segurança Pública do Estado da Bahia (SSP-BA).

Entre outros objetivos, este artigo pretende, através do KDD, modelar a criminalidade a partir de bases de dados públicas. Para isso, é utilizado um conjunto de indicadores de criminalidade que, presumivelmente, tem um impacto sobre a situação de segurança pública e que serão descritos e discutidos a seguir.

O objeto do estudo aqui descrito é o *Knowledge Discovery in Databases* (KDD), mais especificamente em bases de dados públicas, tendo como recorte sua aplicação ao estudo da criminalidade na cidade de Salvador. O recorte foi escolhido por ser foco de atenção cada vez maior por parte da sociedade civil, dos governos, organizações não-governamentais nacionais e internacionais. Já o recorte geográfico pode ser justificado por tratar-se da terceira maior região metropolitana do País, simultânea e paradoxalmente uma das mais violentas. Observa-se ainda que não há literatura suficiente tratando de aplicações de KDD na área de segurança pública no Brasil. Nos poucos trabalhos existentes, observa-se a falta de uma estruturação específica de fases do processo KDD para bases de dados públicas.

A bibliografia descreve diversas abordagens para o KDD. Para este estudo, optou-se por um modelo híbrido derivado

das propostas encontradas na literatura. Aqui são descritas as duas fases principais: prospecção e mineração de dados. A fase de prospecção compreende as etapas: identificação de objetivos; levantamento (identificação e classificação das fontes existentes e definição do modelo de análise); reunião; seleção e criação das bases de dados; consistência (limpeza ou eliminação de ruído e enriquecimento) das bases de dados e compatibilização das bases de dados. Já a fase de mineração abrange as etapas: transformação dos dados; função; técnicas e algoritmos e avaliação dos resultados.

A fase inicial de prospecção destina-se à delimitação das perguntas de pesquisa, definição dos objetivos, organização da equipe de trabalho e planejamento das atividades a serem executadas.

No contexto deste artigo, o emprego do KDD tem como objetivos: mensurar o fenômeno criminalidade e delimitar e estruturar uma base de dados de porte e relevância social para uso do KDD para fins do estudo da criminalidade na cidade de Salvador.

Na etapa de levantamento foi feita a identificação e classificação das principais fontes de informação públicas visando identificar bases de dados com capacidade para suprir as necessidades da pesquisa. Foi utilizada a série histórica do Centro de Documentação e Estatística Policial (Cedep), disponibilizada pela SSP-BA. Em seguida, foi definido o modelo de análise e seus respectivos indicadores. A partir desta estruturação, buscou-se medir um fenômeno de caráter multidimensional – a criminalidade. Para a modelagem da criminalidade, foram selecionados os indicadores de furtos e roubos de veículos (Código Penal, art. 155 e art. 157, respectivamente) e do número de vítimas de homicídios classificados como intencional (Código Penal, art. 121). As razões para esta opção serão descritas resumidamente a seguir.

Em sintonia com o que é preconizado na literatura, os indicadores de criminalidade foram selecionados a partir das taxas identificadas nos registros existentes no rol disponibilizado pela Secretaria de Segurança Pública da Bahia. A escolha desse conjunto de indicadores considerou alguns critérios pragmáticos defendidos por Tironi e outros (1991), Soares Junior e Quintella (2003) e Trzesniak (1998).

Entre os principais critérios observados destacam-se: relevância, gradação de intensidade, univocidade, padronização, rastreabilidade, estabilidade, representatividade e simplicidade.

Uma vez estabelecidos os objetivos, feita a identificação e a classificação das fontes, definido o modelo de análise a ser utilizado com suas respectivas dimensões, selecionados, avaliados e justificados os indicadores, deu-se início à criação da base de dados efetiva – jazida de dados – que serviu ao processo KDD.

A partir desse ponto, os dados foram trabalhados com o objetivo de estruturar a jazida de dados para atender à mineração de dados. Para atingir este objetivo, foram empregadas as etapas do KDD: reunião, seleção, criação e consistência. O produto destas etapas foi um arquivo, totalizando 5.478 registros, em formato compatível com o conjunto de softwares empregados. A população da pesquisa, portanto, foi constituída de todas as ocorrências policiais registradas na cidade de Salvador entre 1º de janeiro de 2003 e 31 de dezembro de 2007. Por fim, na etapa de compatibilização foram unificadas as diferentes bases de dados originais já consistidas, resultando na jazida de dados. Após a etapa de compatibilização, a jazida de dados passa a ter o formato requerido para as transformações sintáticas e semânticas que compõem o início da mineração de dados.

Após o desenlace do processo de prospecção, dá-se início à mineração de dados. Observa-se aqui que este segundo processo é mais complexo que o primeiro, embora, paradoxalmente, seja o que exige menor tempo para sua execução. No KDD, assim como na pesquisa mineral, freqüentemente dedica-se mais tempo à delimitação da jazida que a sua exploração.

No decorrer da mineração de dados realizada durante esta pesquisa, foram executadas as tarefas de transformação dos dados; escolha da função de mineração; técnica e algoritmo de busca e avaliação dos resultados.

A etapa de transformação consiste na transformação sintática e semântica da jazida de dados. A transformação sintática é aquela que não altera o significado dos dados, visa apenas a atender os requisitos das ferramentas e

algoritmos da mineração utilizados nas etapas subsequentes. Já a transformação semântica busca atender, com o cálculo de indicadores, ao modelo de análise previamente definido.

Após todo o trabalho de prospecção e de delimitação da jazida de dados transformada, dá-se início à escolha da função ou conjunto de funções. A seleção da função determina a maneira como é feita a busca por reconhecimento de padrões e relacionamentos complexos. O sucesso desta seleção, para Diniz e Louzada Neto (2000, p.28), “está diretamente ligado à experiência e intuição do analista”.

Neste artigo, o conjunto de funções selecionadas para esta etapa foi: a) Análise preliminar dos dados pela sumarização; b) Mensuração da tendência na incidência dos crimes para todos os indicadores (variáveis); c) Emprego da função associação para avaliação da persistência ou antipersistência da tendência para cada crime estudado.

Definido o conjunto de funções e seu respectivo encaideamento, partiu-se então para a busca do grupo de técnicas e algoritmos mais apropriados para cada função. Embora pudessem ser utilizados algoritmos de origem heurística, nesta pesquisa optou-se por trabalhar com algoritmos derivados da análise estatística. Descreve-se, a seguir, cada uma dessas etapas.

A sumarização foi aplicada no estudo preliminar dos indicadores selecionados e armazenados na jazida de dados, bem como a orientação quanto à escolha das técnicas para as funções de mineração, associação e regressão. Foi aplicada a função complementar caracterização, optando-se por estatística descritiva por mera questão de disponibilidade de software. Cabe ressaltar que também poderiam ser utilizadas as técnicas de SQL tradicional ou OLAP, entre outras. A função complementar visualização, por sua vez, contribuiu de forma significativa para a interpretação dos resultados.

Para atender à função associação foi escolhido o método de Análise das Séries Temporais (AST), mais especificamente Média Móvel Trimestral (MMT). Com a MMT foi possível identificar a tendência de cada um dos indicadores analisados.

Por fim, foi utilizada a Detrended Fluctuation Analysis (DFA) para obter os escores fatoriais de cada setor censitário, com o objetivo de encontrar o expoente α de correlação de longo alcance. A modelagem da criminalidade através do KDD possibilitou avaliar as correlações de longo alcance das séries temporais dos crimes na capital baiana.

FUNÇÃO ASSOCIAÇÃO DO KDD: ANÁLISE DAS SÉRIES TEMPORAIS

Embora o estudo cronológico (dia a dia) da criminalidade utilizando técnicas estatísticas seja uma iniciativa inovadora, quando comparada às outras abordagens, a Análise de Séries Temporais das estatísticas criminais é necessária e pode ser realizada.

Para Morettin e Tolo (2004, p.1), uma série temporal pode ser definida como um conjunto de observações ordenadas no tempo. Seja Y a variável observada e t a variável tempo, uma série temporal é definida pelos valores: $Y_1, Y_2, Y_3, \dots, Y_n$, nos tempos: $t_1, t_2, t_3, \dots, t_n$. Assim, Y é uma junção de t simbolizada por

$$Y=f(t) \quad (1)$$

Onde:

Y : valor da variável em estudo;

t : período a que ela se refere;

f : regra que relaciona o valor da variável com o período a que ela se refere.

A Análise de Série Temporal (AST) é o procedimento pelo qual são identificados e segregados os fatores relacionados com o tempo que influenciam os valores observados na série. Uma vez identificadas, estas informações podem ser usadas para auxiliar na interpretação e na projeção de valores da série temporal (KAZMIER, 1982, p.329). Para o autor, a abordagem clássica da análise das séries temporais deve contemplar quatro componentes:

$$f(\text{série}) = (\text{tendência} + \text{variações cíclicas} + \text{variações sazonais} + \text{variações irregulares}).$$

As variações cíclicas mostram as oscilações de longo prazo, de subida ou descida, que se observam na tendência. Já as variações sazonais são padrões que variam segundo algum período específico da série, enquanto as variações irregulares recebem tal nomeação porque não obedecem a um padrão uniforme e são causadas por eventos aleatórios de muito curto prazo. Além das propriedades da abordagem clássica, uma outra propriedade pertinente em uma série temporal é a correlação, isto é, se a série de dados possui memória de longa duração ao longo do tempo. É possível mensurar a existência de correlação em séries temporais através de diversas técnicas (TAQUU et al, 1995).

Atualmente, modernas técnicas estatísticas de Análise de Séries Temporais (AST) estão disponíveis e têm sido utilizadas para diagnosticar o comportamento de determinados fenômenos em função do tempo; entre os mais recentes, o método Detrended Fluctuation Analysis (DFA) vem sendo aplicado com sucesso no estudo de séries temporais (PENG et al, 1994).

Denominado análise de flutuações sem tendência, o DFA pode ser aplicado para analisar correlações de longo alcance em séries temporais não-estacionárias. O método DFA é uma evolução da análise R/S de um *random walk* e sua principal vantagem é permitir a identificação de correlações de longo alcance em sinais com tendências polinomiais, que podem mascarar as verdadeiras correlações (KANTELHARDT et al, 2001).

A técnica DFA tem sido aplicada com sucesso em diversas áreas do conhecimento, como por exemplo: seqüência de DNA (BULDYREV et al, 1995) e (PENG et al, 1995); dinâmica do batimento cardíaco (PENG et al, 1998); Correlação de longo alcance em disquetes de computador (ZEBENDE et al, 1998); análise de raios-X estelar oriundos de sistemas binários (MORET et al, 2003); séries temporais em economia e mercados de capitais (LIU et al, 1999), além de diversos outros campos de aplicação.

Para modelar estatisticamente a DFA, conforme orienta Peng et al (1994), é preciso seguir ao menos três passos principais:

Seja uma série temporal do indicador estudado u_i , por exemplo, de homicídios dolosos, com i variando de 1 a N (número total de observações da série).

- I. Inicialmente é preciso calcular o desvio de cada registro em relação à incidência média daquele dia na série completa, ou seja, integrando assim o sinal u_i obtendo

$$y(k) = \sum_{i=1}^k u_i - \langle u \rangle \quad (2)$$

onde $\langle u \rangle$ é valor médio de u_i . Esse procedimento permite eliminar o grande número de oscilações e assim obter uma série estacionária, ou seja, apropriada para aplicar o método. Em seguida, a série integrada $y(k)$ é dividida em intervalos de igual tamanho de amplitude (box) n não sobrepostos.

- II. Para cada amplitude (box) de tamanho n é necessário ajustar um polinômio de grau 1 a $y(k)$, que pode ser denominado como $y_n(k)$, e esta estatística representa a tendência local na amplitude (box) de tamanho n . A série $y(k)$ é subtraída de $y_n(k)$ em cada amplitude (box) de tamanho n . Assim, para cada amplitude (box) de tamanho n a sua raiz quadrática média $F(n)$ é computada, isto é:

$$F(n) = \sqrt{\frac{1}{N_{max}} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (3)$$

É importante destacar que o cálculo descrito anteriormente deve ser repetido sistematicamente para diferentes amplitudes (boxes) de tamanho n .

- III. Por fim, como a função $F(n)$ comporta-se como uma lei de potência, do tipo

$$F(n) \propto n^a$$

em que a é o expoente de correlação de longo alcance, esta relação pode ser linearizada em um gráfico $\log F(n) \times \log(n)$ e é representada por uma reta cuja inclinação a representa o expoente procurado.

Para séries não correlacionadas espera-se que $F(n) \propto n^{1/2}$, o que significa expoente $a=0,50$ representando uma série descorrelacionada, caso típico do movimento

browniano usual. Em contrapartida, quando o expoente $\alpha > 0,50$, pode-se inferir que a série apresenta persistência, e o expoente $\alpha < 0,50$, um comportamento antipersistente (FEDER, 1988, p. 171).

Como é apresentada na próxima seção, a aplicação dos passos da técnica DFA permite medir a correlação de longo alcance dos indicadores de criminalidade que mais afetam a vida cotidiana do soteropolitano.

ANÁLISE DOS RESULTADOS

O aumento da incidência dos crimes pode ser percebido por toda a população e sua distribuição não se manifesta homogeneamente, nem no tempo e nem no espaço. A criminalidade possui especificidades que estão condicionadas ao espaço geográfico e ao tempo em questão. Estudos recentes revelam que a incidência de delitos em áreas periféricas, por exemplo, tendem aos crimes contra a pessoa, enquanto nas áreas centrais, mais abastadas economicamente, aos crimes contra o patrimônio (FELIX, 2002, p.3).

O emprego do KDD e a análise da série temporal dos registros diários selecionados permitem descrever o comportamento do fenômeno social, além de oferecer indícios que podem subsidiar a elaboração de políticas públicas que favoreçam a redução da incidência de

crimes e contribuam com o desenvolvimento social da capital baiana.

Com o objetivo de observar a tendência da série, optou-se pela suavização da série de alguns indicadores com auxílio da Média Móvel Trimestral.

O resultado da aplicação da função Associação do KDD – Média Móvel Trimestral (MMT), apresentado na Figura 1, permite identificar uma tendência de crescimento dos roubos de veículos a partir do segundo trimestre de 2005.

Já para os registros de furto de veículo, a aplicação da MMT sugere uma tendência de diminuição no acontecimento dos crimes (Figura 2). O comportamento da série dos registros de furto e roubo de veículos parece apresentar uma relação de complementaridade, onde a redução do furto é 'compensada' com o aumento do roubo de veículos, resultando em tendência crescente para série acumulada dos dois delitos.

O comportamento do homicídio doloso em Salvador no período em análise também apresentou tendência de crescimento a partir do primeiro trimestre de 2006 (Figura 3), o qual permite inferir, assim como no caso do roubo de veículo, sobre a necessidade de ampliar as políticas e ações governamentais voltadas para combater a incidência desses crimes.

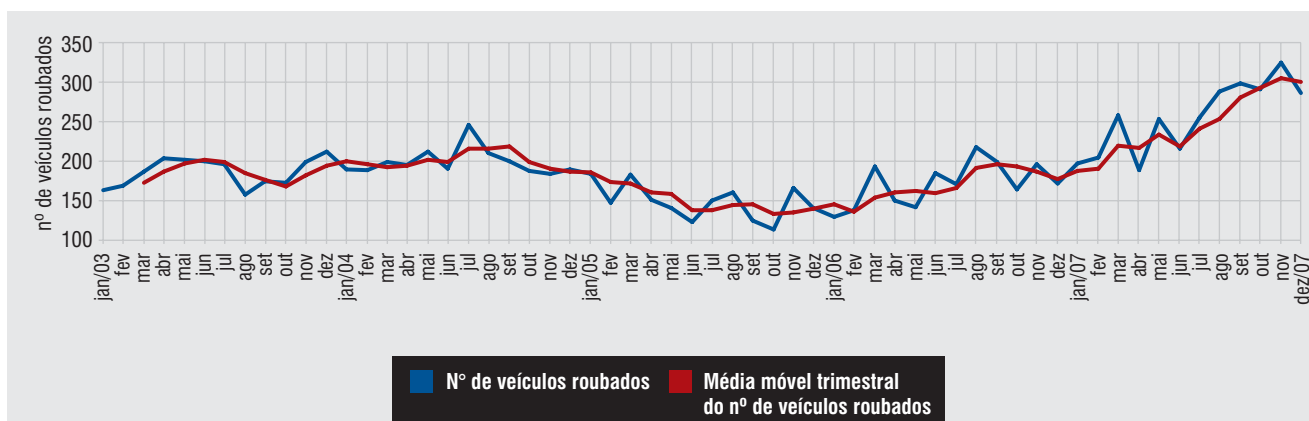


Figura 1
Média móvel trimestral do roubo de veículo em Salvador-BA – jan 2003-jan 2007

Fonte: Cedep

Nota: Cálculos dos autores

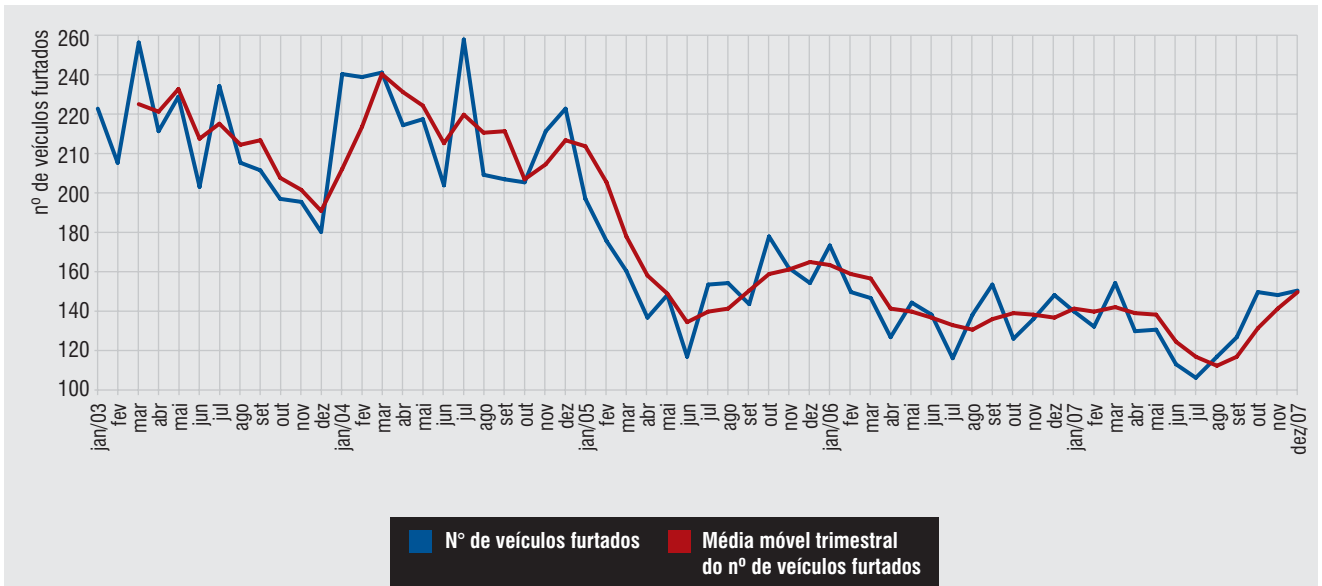


Figura 2
Média móvel trimestral do furto de veículo em Salvador-BA – jan 2003-jan 2007

Fonte: Cedep
Nota: Cálculos dos autores

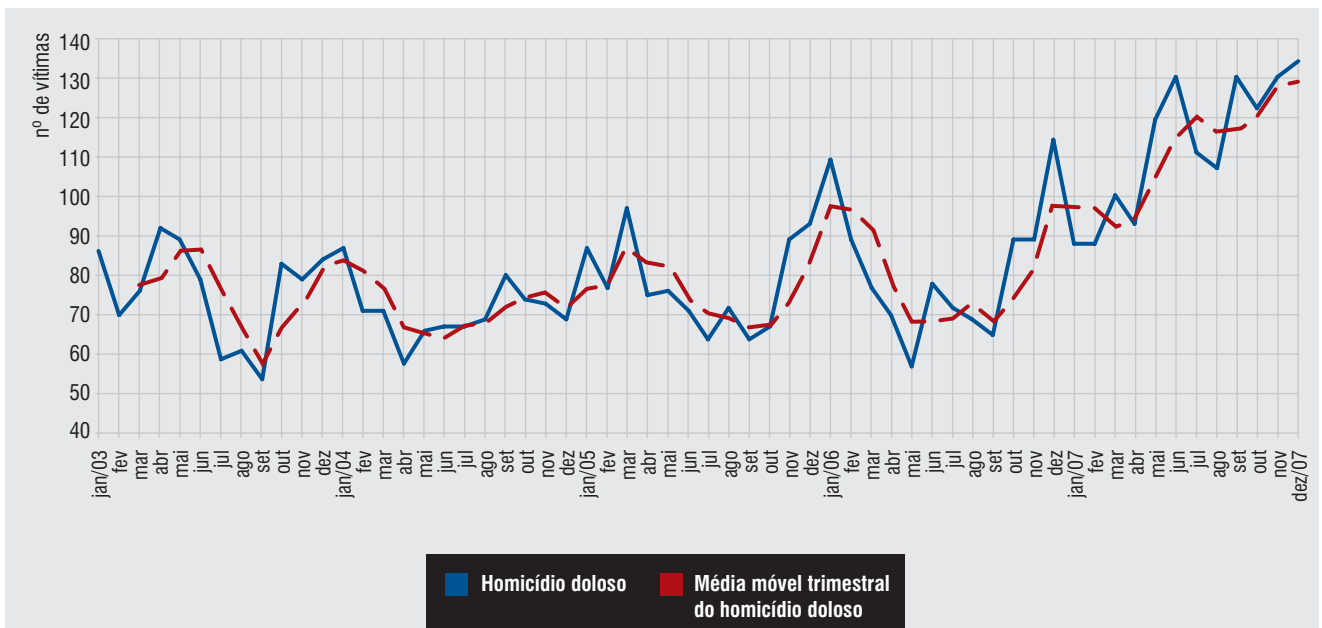


Figura 3
Média móvel trimestral do homicídio doloso em Salvador-BA – jan 2003-jan 2007

Fonte: Cedep
Nota: Cálculos dos autores

Como a análise do comportamento da tendência pode ser complementada pela análise da persistência ou antipersistência da série, neste trabalho optou-se por calcular os expoentes de correlação de longo alcance (α) através da função associação do KDD, mais especificamente do método DFA. A função associação foi utilizada para identificar as propriedades de correlação dos sinais das séries originais dos indicadores de criminalidade estudados. Ou seja, o expoente α , obtido através do DFA, permite avaliar em que medida a tendência observada na série temporal passada implica em manutenção do comportamento no futuro, indicando um efeito de memória de longa duração na série.

Tabela 1
Comportamento do expoente α de alguns delitos de Salvador-Ba – 2003-2007

Delitos	Expoente α	Erro
Roubo de veículo	0,64	0,01
Furto de veículo	0,62	0,01
Homicídio doloso	0,58	0,02

Fonte: Cedep

Nota: Cálculos dos autores

A análise conjunta dos expoentes (α) calculados a partir da série temporal dos indicadores de criminalidade soteropolitanos (Tabela 1) permite avaliar se a série apresenta comportamento persistente, antipersistente ou descorrelacionado.

Grosso modo, a interpretação do expoente α deve considerar as seguintes situações: i) expoente igual a 0,50 revela que a seqüência de dados é descorrelacionada, ou seja, não se pode relacionar o comportamento de uma série passada com o comportamento de uma série futura; ii) expoente α menor que 0,50 indica uma expectativa de inflexão na trajetória da série futura, por exemplo, se a série estudada apresenta um comportamento antipersistente e uma tendência de crescimento, isso implicará em uma tendência de redução futura e vice-versa; e iii) expoente α maior que 0,50, a série é dita persistente e o comportamento de uma série passada tende a continuar no futuro.

Neste estudo, os expoentes α calculados apresentaram persistência ($\alpha > 0,50$) para todas as séries temporais

analisadas. Em se tratando de análise do desenvolvimento social, mais especificamente da criminalidade utilizando DFA, a situação esperada para correlação de longo alcance é um expoente $\alpha > 0,50$, ou seja, comportamento persistente, combinado com uma tendência de decrescimento ao longo da série.

Essa expectativa de comportamento foi constatada apenas nos registros oficiais de furto de veículo. Contudo, a situação observada nos registros de roubo de veículo e homicídio doloso é preocupante e evidencia uma tendência de crescimento do número de delitos aliada a um comportamento persistente da série. Esta constatação suscita ações imediatas do poder público no sentido de reverter tal tendência, sob o risco da sociedade arcar com as conseqüências nefastas que este cenário denuncia.

CONSIDERAÇÕES FINAIS

Foi possível observar que a aplicação do KDD em bases de dados públicas permitiu identificar, mesmo com as limitações impostas pelas bases de dados utilizadas, que, através da relação entre o expoente α obtido a partir do DFA com a tendência da série de indicadores, é possível afirmar que existe uma expectativa de redução do furto de veículo, enquanto o roubo de veículo e o homicídio doloso tendem a manter um comportamento crescente.

Neste artigo, diversas contribuições foram apresentadas ao estudo do KDD; entre as principais destacam-se: elaboração de uma proposta de estruturação e sistematização de etapas para o processo KDD em bases de dados públicas e a modelagem do subdesenvolvimento por meio do fenômeno social da criminalidade.

Como o DFA tem por princípio fornecer uma noção de futuro, o método pode ser também utilizado como ferramenta auxiliar para o planejamento nos órgãos de Segurança Pública. Após aplicação do DFA em séries de alguns indicadores de criminalidade pode-se pensar em propostas futuras de utilização da técnica nestes e em outros indicadores sociais para as demais capitais e regiões do Brasil. Sendo assim, a questão inovadora que este trabalho proporciona está relacionada à possibilidade

de utilização de séries temporais diárias de indicadores de criminalidade, permitindo uma melhor compreensão do fenômeno em estudo.

Os resultados deste estudo demonstram para os analistas de dados e especialistas do domínio que a utilização dos algoritmos disponíveis nos softwares de DM exige, além de uma postura responsável, o conhecimento aprofundado de cada etapa do processo, bem como do domínio que está sendo estudado. Espera-se que este artigo contribua para uma reflexão acerca da forma com que essas bases de dados públicas vêm sendo utilizadas.

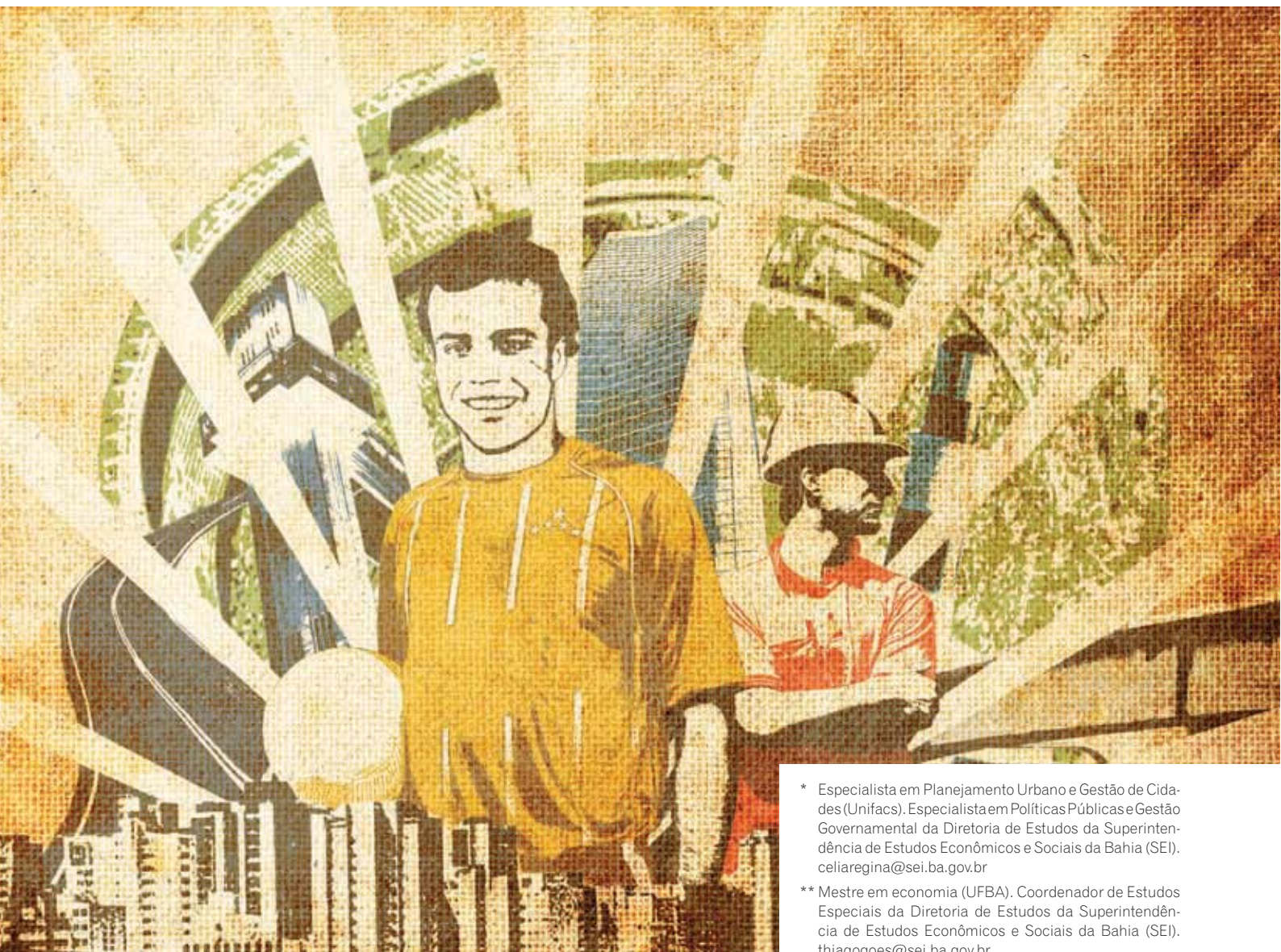
Concluindo, a modelagem de estatísticas diárias dos delitos a partir do KDD pode gerar ações antecipadas por parte da sociedade e seus representantes legais. Espera-se que este trabalho possa oferecer subsídios para formulação de políticas e programas de governo voltados às questões da área de segurança pública, como também contribuir com futuros estudos sobre criminalidade, principalmente porque existem poucos trabalhos desenvolvidos utilizando técnicas estatísticas mais sofisticadas, fornecendo uma nova proposta para avaliar os indicadores de criminalidade no auxílio da gestão da segurança pública.

REFERÊNCIAS

- ADRIAANS, P.; ZANTIGE, D. , *Data mining*. Harlow: Addison-Wesley, 1996.
- BRACHMAN, R. J. et al. *Mining business databases*. Communications of the ACM. ACM Press, New York, v. 39, n. 11, p. 42-48, Nov.1996.
- BULDYREV, S. V. et. al. *Physical Review E*, n. 87, p.10, 1995.
- CABENA P.et al., *Discovering data mining: from concept to implementation*. New Jersey: Prentice Hall, 1998.
- DINIZ, Carlos Alberto R.; LOUZADA NETO, Francisco. *Data Mining: uma introdução*. São Paulo: ABE, 2000.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI/MIT Press, 1996.
- FRAWLEY, W.; PIATETSKY-SHAPIRO, G.; MATHEUS, C. *Knowledge discovery in databases: an overview*. AI Magazine, Menlo Park, CA, p. 213-228, Fall 1992.
- FEDER, Jens. *Fractals*. Plenum Press: New York, 1988.
- FELIX, Sueli Andruccioli. *Geografia do crime: interdisciplinaridade e relevâncias*. Marília:Unesp-Publicações, 2002.
- HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. New York: Morgan Kaufmann, 2000.
- KANTELHARDT, J. W. et. al. Detect long-range correlations with detrended fluctuation analysis. *Physica A*, n. 295, p. 441-454, 2001.
- KAZMIER, L. *Estatística aplicada à economia e administração*. São Paulo: McGraw-Hill, 1982.
- LIU, Y. et. al. Statistical properties of the volatility of price fluctuation. *Physical Review E*, v. 60, p. 1390-1400, 1999.
- MORET, Marcelo A. et. al. Fluctuations analysis of stellar X-Ray Binary Systems. *Physical Review E*, n. 68, p. 41104, 2003.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: Edgard Blucher, 2004.
- PENG, C. K. et. al. On the mosaic organization of DNA sequences. *Physical Review E*, n. 49, p. 1685-1689, 1994.
- _____. Statistical properties of DNA Sequences. *Physica A*, n. 221, p. 180, 1995.
- _____. Scaling and universality in heart rate variability distribution. *Physica A*, n. 249, p. 241, 1998.
- SOARES JUNIOR, Jair Sampaio; QUINTELLA, Rogério Hermida. Sistemas de apoio à decisão e descoberta de conhecimento em bases de dados: uma aplicação potencial em políticas públicas. *Organizações & Sociedade*, Salvador, v. 28, p. 83-98, 2003.
- TAQQU, M. S. et al. Estimators for long-range dependence: an empirical study. *Fractals*, London, v. 3, n. 4, p. 785-798, 1995.
- TIRONI, L. F. et al. Critérios para geração de indicadores de qualidade e produtividade no serviço público. Brasília: IPEA/MEFP, 1991.
- TRZESNIAK, P. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. *Revista de Ciência da Informação*, Brasília, v. 27, n. 2, p. 159-164, maio/ago.1998.
- ZEBENDE, G. F. ; OLIVEIRA, P. M. C. de; PENNA, T. J. P. . Long-range Correlation in Computer Diskettes. *Physical Review E*, v. 57, p. 3311-3314, 1998.

Copa do Mundo de futebol em Salvador: uma **oportunidade de desenvolvimento** urbano e regional

Celia Regina Sganzerla Santana*
Thiago Reis Góes**



* Especialista em Planejamento Urbano e Gestão de Cidades (Unifacs). Especialista em Políticas Públicas e Gestão Governamental da Diretoria de Estudos da Superintendência de Estudos Econômicos e Sociais da Bahia (SEI). celiaregina@sei.ba.gov.br

** Mestre em economia (UFBA). Coordenador de Estudos Especiais da Diretoria de Estudos da Superintendência de Estudos Econômicos e Sociais da Bahia (SEI). thiagogoes@sei.ba.gov.br