

MODELAGEM FUZZY DA VIOLÊNCIA E CRIMINALIDADE

Wellington Jorge Boness dos Santos

Programa de Modelagem Computacional, SENAI Cimatec,
Av. Orlando Gomes 1845, 41.650-010, Salvador, BA, Brasil
wboness@bol.com.br

Hernane Borges de Barros Pereira

Programa de Modelagem Computacional, SENAI Cimatec,
Av. Orlando Gomes 1845, 41.650-010, Salvador, BA, Brasil
&

Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana,
Campus Universitário, 44031-460, Feira de Santana, BA, Brasil
hbbpereira@gmail.com

RESUMO

Muitos pensadores, como Einstein, Freud e Platão olhavam o mundo como uma sucessão de conflitos irremediáveis. Relatos de violência são datados desde os primórdios da existência humana, ganhando proporções com o crescimento das populações. Mudanças na forma de perceber e combater a violência também foram observadas. As responsabilidades pelo controle também migraram do Patriarca para o Governador. Para garantir o exercício dessas responsabilidades, surgiram as necessidades de registro e análise da violência. Este trabalho objetivou mensurar os níveis de violência e a significância dos indicadores frente a essa medição. A escolha dos indicadores é feita em tempo de execução, permitindo a inserção de diversos fatores sociológicos na análise, garantindo a interdisciplinaridade e a transversalidade que o tema exige. Face à imprecisão do conceito de violência, a análise foi realizada através da estatística aplicada e da utilização da lógica difusa na determinação dos graus de violência e da significância dos indicadores.

PALAVRAS CHAVE. Violência e Criminalidade. Modelagem Computacional. Lógica Difusa.

ABSTRACT

Many thinkers, like Einstein, Freud and Plato looked the world like a hopeless conflict succession. Reports of violence was dated from the dawn human existence, getting proportions with population growth. Changes on perceiving and combating violence way were also observed. The responsibilities to control it also migrated from the micro (Patriarch) to macro (Governor). To ensure the exercise of these responsibilities, needs to record and analysis the violence. This study aimed to measure the violence levels and the indicators significance against this measurement. The indicators choice is made at runtime, allowing the insertion of several sociological factors in the analysis, ensuring that the interdisciplinary and cross-cutting theme demands. Like the violence concept is vagueness, the analysis was performed using Applied Statistics and the fuzzy logic was used in determining of the violence degrees and the indicators significance.

KEYWORDS. Violence and criminality. Computational Modeling. Fuzzy Logic.

1. Introdução

A violência é uma problemática que possui registros datados dos primórdios da concepção do homem. Krug et al. (2002) comenta que a violência é enxergada como parte da experiência humana e os seus impactos podem ser vistos, sob várias formas, em todas as partes do mundo.

Por muito tempo coube ao patriarca a responsabilidade pelo controle da violência. O crescimento populacional e a organização das sociedades culminaram na inserção do Estado, e seus representantes, como agentes competentes para prevenir, controlar, julgar e punir atos de violência. O “*Código de Hamurabi*”, que dentre outras coisas objetivava a implantação da justiça na Terra, reflete influência do Estado na vida das classes sociais da época. Baseava-se na “*lex talionis*” como normatização do ordenamento jurídico, imputava ao Governador a responsabilidade maior pela manutenção da ordem e o controle da violência, porém atribuía a todos da sociedade o ônus do resultado da violência. Este código é o primeiro registro do tratamento da violência de forma interdisciplinar.

Algumas observações sobre as origens da violência eram focadas no indivíduo. Por exemplo, Einstein et al. (1988) afirmam que o homem possui internamente um desejo por ódio e destruição, permanecendo latentes em tempos normais, vindo a aflorar em circunstâncias incomuns; Raggio (1992), por sua vez, afirma que a agressividade é algo vital ao ser humano. Recentemente, as pesquisas evidenciam que os fatores individuais e/ou biológicos interagem com outros fatores, como os familiares, comunitários, econômicos, culturais e outros externos para criar uma situação onde a violência comumente ocorre.

A formação dos grandes centros urbanos vem mostrando mais ocorrências delituosas que as cidades e localidades de menor porte, salvo pontuais exceções. Essa constatação é feita com base em visões quantitativas (e.g. registros de ocorrências delituosas que estes locais possuem, tanto em números absolutos quanto em relativos). Da mesma forma, o surgimento e a manutenção de organizações criminosas acontecem nos grandes centros urbanos com mais frequência que nos lugarejos mais modestos e longínquos. Os motivos que levam a essa característica perpassam o simples aumento do contingente populacional, enveredando pelas dimensões política, econômica, social, cultural, etc. Fato é que os sentimentos e ações individuais, anteriormente tratados pela família, agora são atos coletivos de violência, cuja responsabilidade de controle e regulação estende-se ao Estado.

Em 1996, na 49ª Assembléia Mundial da Saúde, a violência foi declarada como importante e crescente problema de saúde pública, cujo foco foi direcionado para as consequências da violência para os indivíduos, as famílias, as comunidades e os continentes. A ausência de mecanismos eficazes para a coleta e aferição das informações que mantêm os diversos setores responsáveis pelas ações de controle e combate à violência os impede de alcançar o seu mister. Um efeito colateral dessa falha é o confronto desses órgãos de segurança com entidades defensoras de direitos, as quais alegam o monopólio do conhecimento. A eficácia perpassa pela consciência da impossibilidade de se analisar a violência de uma única maneira, tomando-a como um fenômeno único.

Com o intuito de gerar indicadores de abrangência nacional, foi realizado no país um mapeamento da violência e criminalidade entre os anos de 1998 e 2008. Como limitação do trabalho, foi relatada a não tentativa de realização de um diagnóstico das causas pela impossibilidade de estabelecer critérios de análise pela diferença entre os municípios. A determinação do ajuste dos índices de violência foi feita pela relação entre o número absoluto de ocorrências de morte e a quantidade de habitantes (100 mil).

Lobo e Fernandez (2003) afirmam que há uma significativa influência de outros indicadores, como a educação, a eficiência da polícia, a concentração de renda, a densidade demográfica, o grau de urbanização, bem como as rendas municipais e estaduais nos índices de violência. Para satisfazer a esses critérios para a composição dos indicadores, torna-se necessária uma profunda mudança na coleta, persistência e nos critérios de socialização dos dados

disponíveis. Deve-se buscar mecanismos de identificação e qualificação de novos indicadores, além de utilizar outras formas de análise para otimizar os resultados. Este trabalho objetivou mensurar os níveis de violência e a significância dos indicadores frente a essa medição, propondo um modelo de análise e combate à violência e criminalidade.

2. Indicadores de violência

É evidente o crescimento dos fatos de violência e criminalidade nos diversos segmentos da sociedade atual. A evolução do contingente populacional determina um aumento substancial das ações de planejamento nos diversos setores (e.g. econômico). Como o crescimento espacial, social e econômico acontece de forma desordenada, os setores responsáveis pelas atividades básicas, como a saúde, o saneamento, a educação e a urbanização não conseguem executar as suas incumbências de forma satisfatória, culminando no aparecimento de problemas sociais de toda monta.

Apesar da missão constitucional, o provimento da segurança é realizado pelos órgãos do Estado, especificamente as Polícias e Guardas, nos seus respectivos níveis municipal, estadual e federal. É patente que as ações de contenção da violência não podem se limitar à atuação desses órgãos, senão através da comunhão desses órgãos e entidades afins. Assim, a captação das informações para análise posterior deve ser originária desses diversos segmentos, de forma a maximizar a capilaridade da coleta. Com base nos registros oficiais de medição da violência, percebe-se a impossibilidade confessa da determinação das suas origens e causas, sendo necessária a criação de novas soluções para tal atividade.

Problemas relativos às diversas origens dos dados, a ausência de registro em vários setores e a falta de integração entre os órgãos constituem desafios a serem superados. A extensão territorial é outro fator complicador que, aliado à desigualdade social, contribuem com a dificuldade de implantação de políticas de difusão da informação.

Estas questões influenciam negativamente na identificação e na construção de indicadores que permitam a utilização de métodos e modelos quantitativos capazes de relacionar áreas do conhecimento, de forma inter e/ou transdisciplinar, visando a formulação de estratégias mais técnicas de atuação.

Para se estabelecer a forma de coletar os dados é necessário definir quais os dados a serem coletados. Violência e criminalidade envolvem várias áreas, torna-se necessária a busca de insumos em todas elas. Pela própria diversidade existente entre essas áreas, nota-se a dificuldade de inter-relacionar as informações provenientes de fontes específicas empiricamente. Em contrapartida, emerge a possibilidade da inserção de modelos (e.g. estatísticos) que permitam estabelecer relações de afinidade entre as áreas consideradas. Tais mecanismos conduzem à produção do conhecimento de forma transversal, o que favorece a identificação de fatores indicação e medição da violência e criminalidade.

Analogamente, pela natureza multivariada dos dados, aumenta-se a possibilidade de inconsistência dos resultados decorrentes dos modelos matemáticos e estatísticos, sobretudo no que se refere à qualificação e quantificação dos fenômenos abstratos, como é o caso da violência. Assim, emerge a viabilidade da utilização de modelos computacionais (e.g. lógica difusa, algoritmos genéticos, *Detrended Fluctuation Analysis*, *Detrended Cross-Correlation Analysis*, etc.) como mecanismos de solução para as limitações técnicas anteriores.

A estratégia utilizada para garantir a robustez dos dados foi a de inserir possíveis indicadores sociais no bojo das coletas de dados dos órgãos de segurança. Assim, na operacionalização diária dos registros de ocorrências delituosas e assistenciais realizadas pela polícia, especificamente a Polícia Militar da Bahia, dados candidatos para a seleção de indicadores de violência são mantidos e disponibilizados para a análise.

Cada campo do Registro de Ocorrências Policiais (ROP) foi modelado como registro de uma base de dados, a fim de permitir a manipulação de forma automatizada. Através de um software (Figura 1), o banco de dados foi alimentado e disponibilizado para uso pelo modelo

proposto. O ROP é fruto de um trabalho da Polícia Militar, com o intuito de registrar as informações julgadas suficientes para a Instituição. No decorrer dos trabalhos, percebeu-se que outros dados candidatos poderiam ser apensados, a fim de promover a transversalidade (e.g. grau de instrução dos envolvidos).

The image shows a web-based form titled "Lançamento de Ocorrências". It is a structured data entry interface for police incidents. The form is organized into several sections: 1. General information: Hora, Origem da solicitação (CENTEL), Decorrente de Operação Policial (Não). 2. Location: Local (Rua, Complemento, Bairro: Centro), CEP, Município (Feira de Santana), Tipo do Local (Via Pública), Complemento, Tipo do Local. 3. Incident details: Ponto de Referência, Latitude, Longitude, Data, Hora Ocorrência, Horário no Local, Tipo de Ocorrência (Consumado). 4. Involved parties: Envolvidos. 5. Additional information: Informações Complementares, Apreensões. 6. Vehicle type: Padrão de viatura utilizada na ocorrência (Viatura Operacional, Viatura Administrativa, Operação Especial). 7. Incident result: Resultado da Ocorrência (Flagrante Lavrado). 8. Officer information: Dados do PM que registrou a ocorrência (Tipo de Policiamento: Radiopatrulhamento, Coronel PM, Matricula, OPM, Companhia, Pelotão: 1º Pel). 9. Action buttons: Aceita, Limpa, Sair.

Figura 1. Tela de lançamento dos Registros de Ocorrências Policiais (ROP).

O primeiro passo foi a identificação dos tipos de variáveis a serem trabalhadas. As variáveis aleatórias foram as bases para a criação dos indicadores e, considerando que na observação de seus valores, mostraram-se enumeráveis e contáveis, concluiu-se que o espaço amostral é finito, logo variáveis aleatórias discretas. No caso dos horários, um agrupamento permitiu uma melhor caracterização dos elementos, sendo que foram separados por hora (espaçamento padrão de 01 hora) e por turno: madrugada (00:01 a 05:59), manhã (06:00 a 11:59), tarde (12:00 a 17:59) e noite (18:00 a 00:00).

O segundo passo foi montar as distribuições. Pela literatura, se pela associação de cada ponto do espaço amostral à sua correspondente probabilidade de ocorrer, obtêm-se a distribuição das probabilidades do experimento e se, além disso, associarmos a esses mesmos pontos uma variável aleatória indicadora, obteremos uma distribuição que pode ser entendida como a distribuição das probabilidades das variáveis aleatórias associadas a pontos do espaço amostral (FONSECA e MARTINS, 1982).

Como a distribuição obtida no processo de associar cada indicador candidato à sua probabilidade de ocorrer irá mostrar como a probabilidade total se divide pelos diversos resultados possíveis do experimento, ela poderá ser totalmente definida tanto pelas frequências simples quanto pelas acumuladas. Através disso, foram colhidos e tratados os dados por indicador. O tratamento dos tipos de ocorrência, por exemplo, desta forma possibilitou a montagem das séries e o seu tratamento no nível individual e do grupo.

Cada tipo de distribuição foi identificado e tratado de acordo com o seu comportamento. Nos casos em que o número de opções foram iguais a dois, utilizou-se a distribuição de *Bernoulli*. Nos casos em que a probabilidade de fracassos em situações não foram mensuradas tenha ocorrido (e.g a ausência de RG, não identificação de vítima ou autor, etc.), a modelagem foi realizada considerando a distribuição de *Poisson*. O modelo permite analisar situações como: (1) o número de registro das ocorrências de um determinado registro não realizado por algum motivo, ou (2) o número de atendimentos não realizados por uma guarnição de serviço.

O passo seguinte foi o de calcular as medidas de posição e dispersão para cada distribuição. Para a determinação dos níveis de significância, originalmente, pensou-se em eleger o melhor indicador (i.e. Indicador Padrão) em tempo de execução. O cálculo das estatísticas foi realizado para cada um dos indicadores candidatos, o que permitiu estabelecer uma comparação entre eles. Pela totalização dos valores (positivos e negativos) da correlação, aplicada a regressão e extraídos os valores do erro, o indicador que possuiu o maior valor agregado foi eleito o

Indicador-Padrão.

As medidas de posição foram calculadas a partir da identificação das séries, feitas através de um *flag* no vetor, decorrente de um algoritmo de conversão de rótulos. No caso dos horários, entretanto, a média foi calculada por

$$\bar{X} = PM_o + h \frac{\sum (d_i * x_i)}{\sum x_i} \quad (1)$$

onde PM é o ponto médio da classe de maior frequência e $d_i = \frac{(PM - PM_o)}{h}$.

A genérica é $\bar{X} = \frac{1}{n} * \sum_{i=1}^n P(x_i)$, onde $P(x_i)$ corresponde aos dados da ocorrência específica recolhidos da base de dados e totalizados pelo indicador-padrão. Nas amostras muito pequenas foi inserido um coeficiente de variação para ajudar a validar, ao invés de utilizar valores pré-calculados, provendo assim os ajustes necessários: $CV = (\sigma / X) * 100$. Adicionalmente, a fim de ajustar possíveis distorções entorno da média, introduziu-se o erro padrão médio $\sigma_x = \frac{\sigma}{\sqrt{n}}$. Com isso, nos casos em que o valor do desvio indicou que a média não foi a melhor representante da amostra (geralmente em amostras pequenas), o ajuste foi realizado. Observou-se que o problema diminuiu com o aumento das amostras, culminando numa diminuição do erro (i.e. melhoria da média).

Após a extração das medidas de posição, há uma eleição do indicador-padrão. Isto é feito através da comparação estabelecida entre todos os indicadores candidatos. Inicialmente é calculado o valor da covariância: $Cov_{xy} = \sum [(P_{x1}(i) - \bar{X}_1) * (P_{x2}(i) - \bar{X}_2)]$, onde será medida a significância de cada indicador frente aos demais. O grau de aproximação ou afastamento será dado com a equação seguinte: $\rho_{xy} = \frac{Cov_{xy}}{\sigma_x * \sigma_y}$, que varia $-1 \leq \rho_{xy} \leq 1$. Para a seleção, computacionalmente foi criado um vetor bidimensional do tamanho total dos indicadores, onde foram mantidos os valores estatísticos calculados. Cada indicador terá sua significância com base nos valores acumulados (ponderados) de sua correlação. A escolha do mais significativo considerou a maior proximidade da unidade (positiva).

Para a caracterização dos graus de violência, dividiu-se a abrangência em duas partes: a geral e a local. Esta estratégia permitiu delinear a função de pertinência do fenômeno observado (Equação 2). Esta função baseou-se nos aspectos quantitativos das ocorrências totais (delituosas) e nas ocorrências específicas do tipo penal (TP) selecionado. Assim, considerando-se que “ n ” é o número total de registros e “ H ” é o número de ocorrências de um tipo (ou grupo) penal (e.g. homicídios), o grau de pertinência do TP “ i ” no conjunto das ocorrências desse TP é a seguinte:

$$p_H(i), \text{ com } i = (1, 2, 3 \dots n) \quad (2)$$

A determinação do grau de violência do TP será obtida (1) da área de abrangência da seleção e (2) de qualquer setor no interior desta. Com isso, têm-se os resultados generalizados e específicos, respectivamente, do TP referenciado, o que permite a visibilidade do grau de violência, tanto em um caráter ordinário (geral), quanto sintomático.

As equações 3 e 4 para o cálculo desses graus de violência são as seguintes:

$$\mu_{VG}(i) = \left(\frac{\sum_{i=1}^n p_H(i)}{\sum_{i=1}^n p(i)} \right), \text{ com } i = \{1, 2, \dots, n\} \quad (3)$$

$$\mu_{VE}(i) = \left(\frac{\sum_{i=1}^h p_{HL}(i)}{\sum_{i=1}^n p_H(i)} \right), \text{ com } i = \{1, 2, \dots, n\} \text{ e } h = \{1, 2, \dots, \text{tam}(\mu_H)\} \quad (4)$$

que são os graus da violência do tipo referenciado e dos locais especificados na seleção, dentro do conjunto de ocorrência do TP em questão. O grau de pertinência será estabelecido pelo seguinte critério:

1. não violento, se: $\mu_V(i) = 0$;
2. com certo grau de violência, se: $0 < \mu(i) < 1$;
3. totalmente violento, se: $\mu_V(i) = 1$.

Para fins de qualificação do grau de violência com base no resultado através da função de pertinência, a Tabela 1 fará a representação da devida correspondência, qual seja:

Tabela 1. Pertinência dos graus de violência.

| Valor | 0 | 0 – 0,1 | 0,1 – 0,3 | 0,3 – 0,4 | 0,4 – 0,6 | 0,6 – 0,8 | 0,8 – 1 | 1 |
|--------------------------|--------------|----------------------|----------------|----------------|--------------------|---------------------|----------------|---------------------|
| Grau de Violência | Não violento | Muito pouco violento | Pouco violento | Violência leve | Violência moderada | Violência acentuada | Muito violento | Totalmente violento |

De posse da importância dos indicadores e do grau de violência, calculou-se a significância dos indicadores frente ao TP selecionado. O cálculo do índice geral foi obtido através da média ponderada dos indicadores selecionados, uma vez que cada indicador possuía o seu grau de significância individual. Esse índice é dado pela Equação 5.

$$\mu_v(x) = \sum_{i=1}^n \rho_{xy}(i) * \mu_{VG}(i), \quad (5)$$

onde “n” é o número de indicadores selecionados. A média ponderada nesse modelo incide tanto no indicador com maior significância, quanto no possuidor de maior frequência. Estabelecida a primeira entrada do controlador difuso.

3. Modelagem Difusa

Os conjuntos clássicos possuem uma função característica ou discriminante baseada na lógica booleana natural, definindo claramente os elementos que são pertencentes ou não ao seu escopo. A sua função de aceitação serve bem para as situações em que os valores são granulados, bem como com variação não contínua, porém muito pouco para fenômenos que possuem uma mudança contínua no seu conjunto de valores.

“Um conjunto *Fuzzy* codifica a imprecisão ou flocosidade associadas a um fenômeno através da sua superfície. A forma da curva, na verdade, representa a semântica do conceito real.” (COX, 1994, p. 33). Desta forma, o conceito de pertinência cresce do simples {pertence, não pertence} para algo como {..., pertence muito pouco, pertence pouco,..., pertence muito,...}. Esses novos valores retratam melhor o modelo de fenômenos da vida real.

O conceito de Violência é algo que, por si só, cria polêmicas nas comunidades que se propõem a discuti-lo. Muitos aspectos devem ser considerados na sua caracterização, tornando-a cada vez mais complexa. Os indicadores mais comuns levam à análise através dos números de certos delitos (e.g. roubo, homicídio, etc.), bem como aspectos sócio-econômicos (e.g. alfabetização, desemprego e pobreza).

É notável a dificuldade de caracterização da violência, e ainda mais difícil a sua mensuração e gradação. Alguns indicadores quantificam o grau de violência através da relação entre esses indicadores e a população da área considerada. Daí, questiona-se a abrangência dessa quantificação, uma vez que os fatores considerados apenas fazem parte de um todo, não

amparado plenamente na equação.

Pensando nessa visível dificuldade, a inclusão da lógica difusa, como veículo promotor do mister de mensurar e graduar a violência, tornou-se viável. O cálculo dos níveis de violência foi baseado em aspectos tanto quantitativo, quanto qualitativo. O processo consiste em quantificar e inserir os valores em conjuntos nebulosos e tratá-los pelos controladores difusos.

3.1. Controlador Difuso

Um controlador difuso é, basicamente, composto por quatro elementos: a base de regras, o mecanismo de inferência, o fuzzificador e o defuzzificador. A base de regras contém uma descrição lingüística de como o controlador deverá funcionar, construída a partir da experiência do especialista na área, convertida em uma quantificação de lógica difusa (KLIR & YUAN, 1996). O mecanismo ou máquina de inferência serve para interpretar e aplicar o conhecimento e, conseqüentemente, otimizar as decisões de como melhor retratar as opiniões dos especialistas (RUAN & HUANG, 2000). O fuzzificador e o defuzzificador são, respectivamente, as interfaces que convertem as informações de entrada e saída do mecanismo de inferência, a fim de aplicar as regras e exibir as conclusões decorrentes do processo.

A principal função de um controlador *difuso* é a de permitir automatizar o controle do sistema baseado nas informações do especialista na solução de questões que ele resolveria com a sua experiência (McNEILL & THRO, 1994). Estas experiências são contadas como formas de tomada de decisão. Assim o especialista indica a informação de entrada a ser utilizada na tomada de decisão.

A determinação das variáveis de entrada do sistema é um processo de vital importância, pois definirá o sucesso do resultado de saída. Assim, a sua definição deverá ser baseada na decisão do melhor direcionamento à solução do problema. Esta definição nem sempre pode ser obtida diretamente. No modelo atual, em primeira instância, a obtenção das entradas se deu em decorrência da análise estatística dos dados. Em um momento posterior, a entrada corresponde à saída do primeiro controlador.

O passo seguinte foi o da definição da base de regras a ser utilizada. Conceitualmente, temos uma descrição lingüística, em linguagem natural, que relata como melhor controlar o mecanismo. Ela “detém as variáveis lingüísticas, os valores lingüísticos, as suas funções de aceitação associadas, bem como o conjunto de todas as regras lingüísticas” (PASSINO, 1998, p. 36). Esta descrição foi realizada pelo relato de especialistas em segurança pública (i.e. oficiais da PM/BA) e foi fragmentada de modo a permitir a identificação de valores e conseqüentes formações de variáveis lingüísticas, o que gerou uma linguagem que possibilitou mostrar como controlar o processo de decisão, a partir das entradas, oferecendo uma saída estabilizada.

O processo consistiu na quantificação do comportamento do sistema, a partir das descrições lingüísticas dos oficiais, persistindo sob a forma de variáveis. Como as regras são formadas por variáveis lingüísticas, suas designações serão regras lingüísticas. Assim, gerou-se o conjunto de regras a partir dos valores quantificados nas variáveis e de descrições lingüísticas, culminando em um conjunto de abstrações que significam diferentes situações ou níveis de eventos atendidos em uma situação do mundo real, enfrentados por esses oficiais, no dia-a-dia do exercício da sua atividade.

A função de aceitação (*membership function*), responsável pela quantificação das descrições, permitiu a realização dos ajustes, pois cada correlação lingüística e o seu respectivo valor quantificado (i.e. grau de certeza) pode ser validado na função de pertinência. Isto permitiu a visualização dos resultados de forma contínua, completando as lacunas do modelo probabilístico e possibilitando determinar resultados de situações “pouco violentas”, “não tão violentas”, etc. Cada valor lingüístico possui a sua função de aceitação, variando em função do tempo, permitindo, assim, que os ajustes fossem feitos visualmente, sob a forma de gráficos (Figura 2).

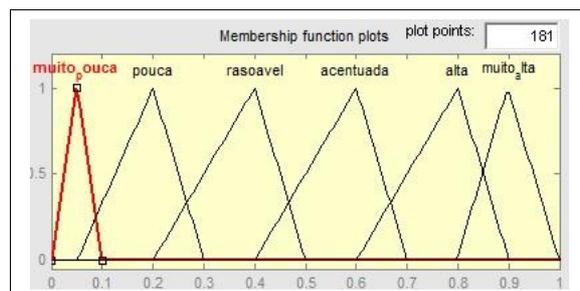


Figura 2. Função de aceitação .

3.2. Arquitetura do Controlador

A arquitetura geral do modelo previu a construção de dois controladores difusos. A escolha das variáveis de entrada seguiu o critério de significância estatística, elegendo aquelas que possuíam maior impacto nas saídas do sistema. Na maioria dos testes, o horário foi eleito como indicador-padrão e a partir dele foram calculadas as correlações necessárias.

Os dois controladores difusos desenvolvidos são: (1) Grau de violência, destinado a determinar o grau de violência do local considerado, a partir dos resultados estatísticos extraídos das séries geradas pelos valores das ocorrências policiais registradas nos ROP, permitindo calcular os graus de violência de toda região (geral), ou analisar certos indicadores selecionados (local ou específica). Duas aplicabilidades decorrem desse resultado: a possibilidade de mensuração dos níveis de violência em uma região espacial e a definição do nível de violência para análise da significância dos indicadores utilizados. (2) Grau de Significância dos Indicadores, que objetiva determinar qual o nível de violência do tipo penal considerado, bem como a significância de cada indicador frente a ele. Como é previsto na literatura, a determinação das variáveis de entrada pode ser de forma direta ou calculada a partir de critérios identificados e seguros. Neste caso, buscou-se extrair dos dados existentes das ocorrências registradas em certo período e local, informações estas necessárias para identificar e quantificar as variáveis de entrada, feito a partir da identificação de certas variáveis lingüísticas, cujo conteúdo serviu para a formação das regras e a conseqüente base de regras do sistema.

Para o primeiro controlador, as variáveis de entrada são Grau de violência Geral e o Grau de Violência Específico do tipo penal considerado. Foi definido o de critério de pertinência e criada a tabela de pertinência dos graus de violência (Tabela 01), cuja função foi inserida para a utilização da base de regras (Figura 3). A determinação dos valores de cada item da tabela, bem como a relação entre cada grupo de regra com a respectiva correspondência de valor foram realizadas pelo especialista em segurança pública.

```
25. If (mvg is alta) and (mve is muito_pouca) then (violencia is Alta) (0.2)
26. If (mvg is alta) and (mve is pouca) then (violencia is Alta) (0.4)
27. If (mvg is alta) and (mve is acentuada) then (violencia is Alta) (0.8)
28. If (mvg is alta) and (mve is alta) then (violencia is Alta) (1)
29. If (mvg is alta) and (mve is muito_alta) then (violencia is Alta) (0.8)
30. If (mvg is muito_alta) and (mve is muito_pouca) then (violencia is Muito_Alta) (0.2)
31. If (mvg is muito_alta) and (mve is muito_pouca) then (violencia is Muito_Alta) (0.4)
32. If (mvg is muito_alta) and (mve is razoavel) then (violencia is Muito_Alta) (0.6)
33. If (mvg is muito_alta) and (mve is acentuada) then (violencia is Muito_Alta) (0.8)
34. If (mvg is muito_alta) and (mve is alta) then (violencia is Muito_Alta) (0.9)
35. If (mvg is muito_alta) and (mve is muito_alta) then (violencia is Muito_Alta) (1)
```

Figura 3. Excerto da base de regras da violência geral.

Para a montagem da base de regras foram estabelecidos os níveis aceitáveis que cada função poderia assumir e a relação entre elas. Inicialmente, criou-se uma quantidade de regras candidatas e, após os ajustes, encontrou-se 89 regras finais, que permitiram quantificar das descrições lingüísticas dos graus de violência.

A função de aceitação (μ), responsável pela tradução dessas regras, permite a

observação visual dos graus de certeza com que cada regra foi modelada. Foram construídas duas funções de aceitação para o primeiro controlador: um para a modelagem da violência geral (μ_{vg}) e outro para a local (μ_{ve}).

O segundo controlador foi destinado a medir o grau de influência que os indicadores possuem frente ao tipo penal considerado. Para este controlador, as variáveis de entrada são a saída do primeiro controlador e o índice geral de violência (Equação 5). Uma correlação é calculada entre os indicadores e o grau de violência do local. De forma análoga ao primeiro, as variáveis de entrada deste segundo controlador também decorrem da determinação processual das informações candidatas a qualificar e quantificar as ocorrências. Uma nova situação ocorre na medida em que o uso da saída do primeiro controlador tornou-se uma variável de entrada para o segundo. Logo, o novo controlador possui as características de detecção direta e pré-calculada das variáveis de entrada. A função de aceitação é semelhante à do primeiro controlador, diferindo apenas na mudança do especialista na formação da base de regras: estatística. Como saída para este controlador, tem-se o valor difuso do nível de significância do tipo penal analisado: quão violento é o local frente àquele tipo penal.

Os processos de fuzzificação e defuzzificação dos dois são semelhantes. A utilização do desenvolvimento de um controlador tipo *Mamdani* se deu pelo fato de tentar estabelecer um misto de filosofias no trabalho. Como a topologia *Sugeno* favorece uma modelagem voltada para um mapeamento mais matemático e a escolha dos indicadores já foi inteiramente baseada na relevância estatística, buscou-se utilizar a estrutura que aproveita as descrições linguísticas de forma mais intuitiva, uma vez que os conceitos sociais foram introduzidos *a posteriori*. Além disso, o modelo permite a inserção de indicadores candidatos a variáveis de entrada (eleitos estatisticamente), o que culminaria em um aumento significativo na dificuldade de realização de ajustes manuais em um controlador do tipo *Sugeno*.

3.3. Aplicação do Controlador

O modelo foi aplicado em um sistema computacional que objetivou integrar diversas tecnologias, a fim de oferecer uma solução para a medição e o mapeamento da violência e criminalidade. A fonte dos dados foi o registro de ocorrências policiais de uma Companhia Independente da Polícia Militar do Estado da Bahia, situada no centro da Região Metropolitana de Salvador, no período de 2007 a 2009. Foram registradas cerca de mil e duzentas ocorrências no período, cujas informações foram modeladas e disponibilizadas para a utilização.

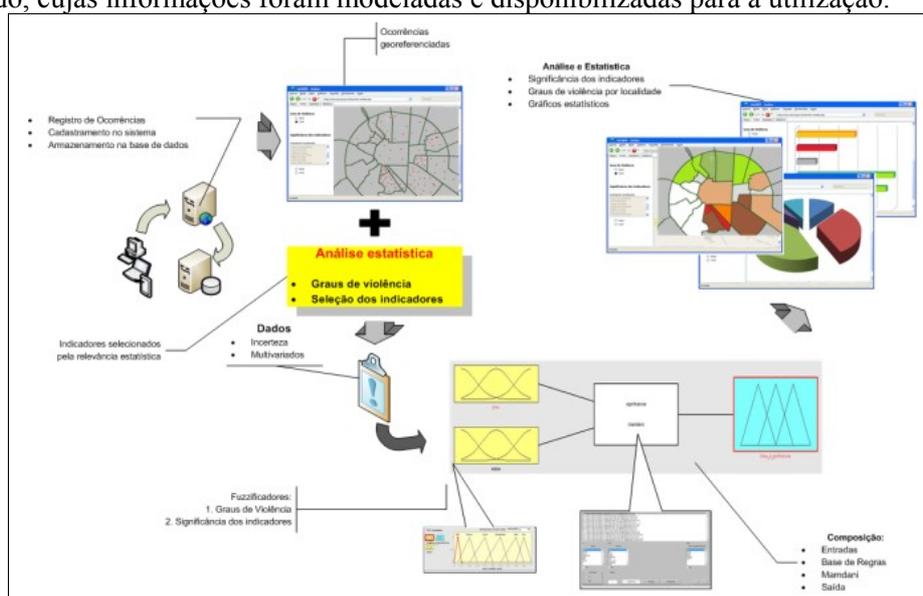


Figura 4. Esquema de funcionamento do sistema computacional que desenvolvemos para a Secretaria da Segurança Pública/BA.

O processo consistiu na coleta, lançamento, modelagem e análise dos dados registrados, com o intuito de prove indicadores significativos de violência, provendo transversalidade nas áreas do conhecimento, a fim de garantir a amplitude necessária para a mensuração dos níveis de violência e representá-los de forma georeferenciada.

O SACRIME é o fruto de uma pesquisa desenvolvida para a Secretaria da Segurança pública do Estado da Bahia através de um convênio firmado com a Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). A proposta foi a de georeferenciar a violência e a criminalidade. O processo consiste no seguinte: (1) Os dados do ROP são lançados e armazenados num servidor. O lançamento pode ser realizado por um usuário previamente cadastrado a partir de um computador com acesso à internet. (2) Os dados são utilizados pelo modelo para a composição das séries estatísticas, eleição do indicador-padrão e distribuição espacial das ocorrências no mapa. (3) O primeiro controlador difuso calcula o grau de violência da região selecionada e serve como uma das entradas do segundo que calcula o índice de significância dos indicadores. (4) A saída pode ser formatada de forma a prover informação visual do grau de violência dos locais selecionados (e.g. ruas, bairros, cidades, etc.), como gráfico indicador dos níveis de violência ou como gráfico indicativo dos níveis de significância dos indicadores frente ao tipo penal considerado.

4. Resultados

A Figura 5 representa graficamente o processo de eleição do Indicador-Padrão, a partir dos indicadores candidatos. Observa-se que há uma oscilação, no decorrer do processo, devida à relação de significância do indicador candidato frente aos demais. Há um confronto particular de cada item selecionado. A oscilação indica, o grau de aproximação ou afastamento decorrente do cálculo da correlação entre eles.

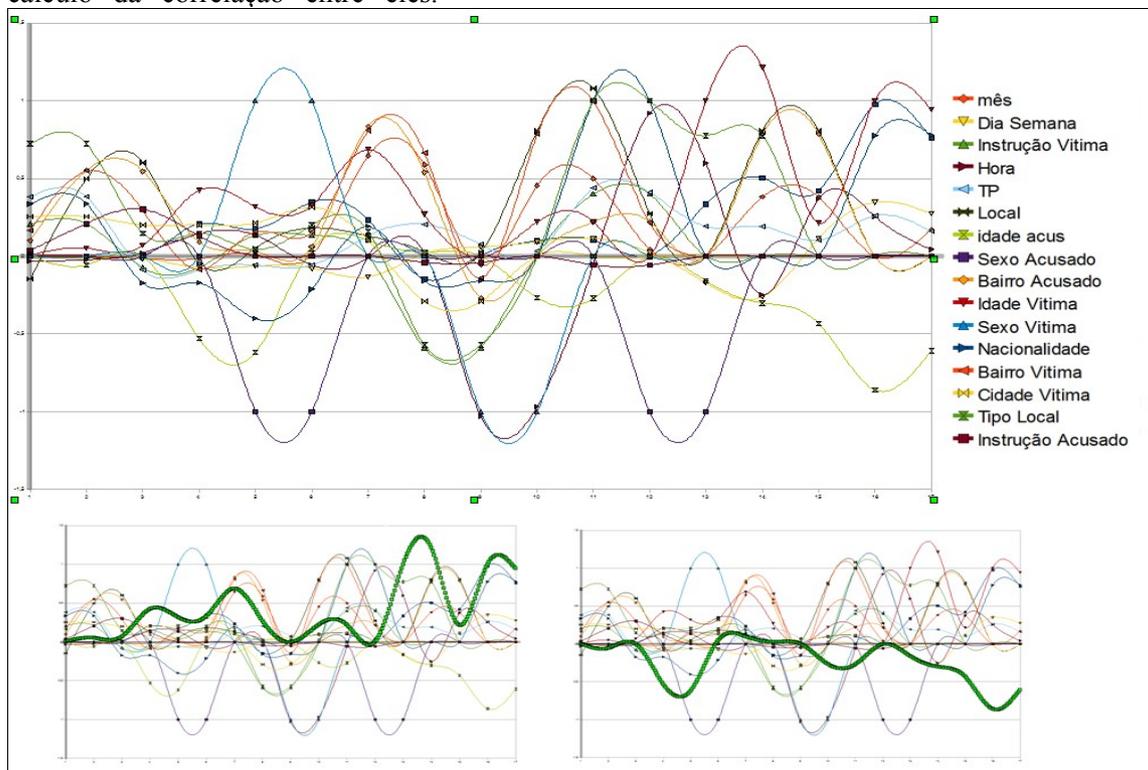


Figura 5. Processo de seleção do Indicador-Padrão, a partir dos indicadores candidatos. As curvas em destaque representam o melhor (Horário da Ocorrência) e pior (Sexo da Vítima) indicadores.

A análise considerou as ocorrências policiais compreendidas entre 01/09/2007 e 18/05/2010, num total de 3228 ocorrências. O gráfico representa a relação entre o acumulado da significância do indicador (Eixo Y) e a correlação entre o indicador referenciado e aos demais indicadores (Eixo X). O indicador eleito foi o de horário das ocorrências, seguido do sexo da vítima. Comparando a evolução gráfica do melhor e do pior indicador observa-se que os níveis de significância do melhor e do pior indicador evoluem inversamente. Analogamente, o gráfico da Figura 6 demonstra como ocorre o espalhamento entorno da média do índice acumulado das significâncias dos melhores indicadores e como eles evoluem entre si.

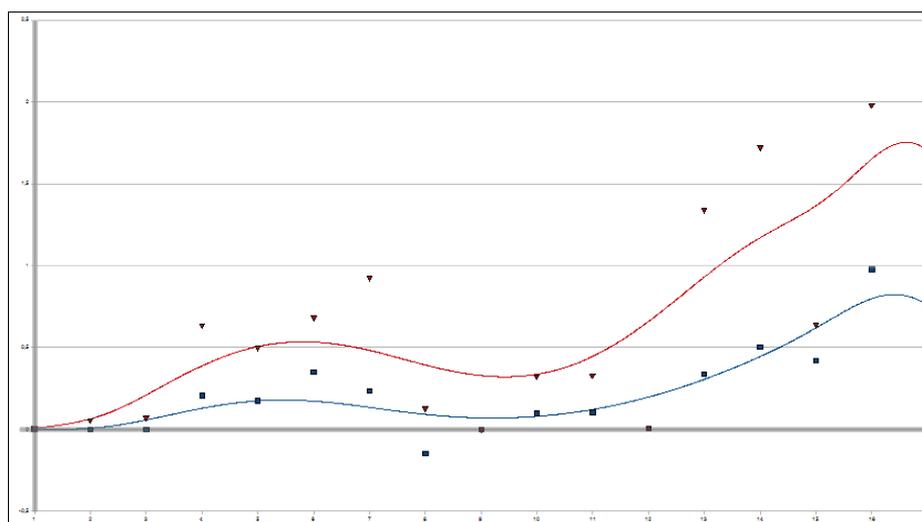


Figura 6. Correlação e espalhamento entorno dos melhores indicadores (i.e. Horário das Ocorrências; Sexo da Vítima).

A escolha do Indicador padrão é vital para o processo, pois os graus de significância geral e específico, entrada para a entrada do controlador difuso, é calculado a partir dele. A Figura 7 representa a estrutura do segundo controlador que possui como entradas a saída do primeiro (graus de violência) e o índice de violência específico da região considerada. O modelo da máquina de inferência escolhida foi o *Mamdani* e como saída um conjunto difuso que representa o grau de significância do indicador frente àquele tipo penal pré-selecionado.

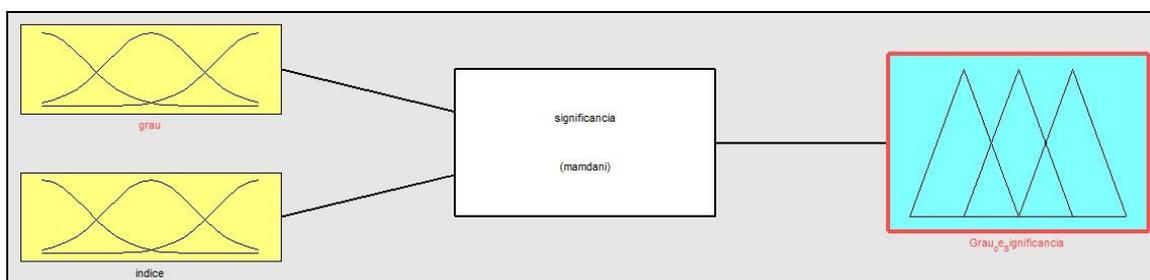


Figura 7 - Estrutura do controlador responsável pela significância dos indicadores.

5. Conclusões

O principal objetivo da pesquisa foi o de produzir um conjunto *fuzzy* que melhor retratasse uma quantificação e qualificação real dos níveis de violência e da significância de seus indicadores, a partir do conceito de violência. A literatura aponta definições de locais violentos a

partir de certa quantidade de ocorrência por n habitantes. Como lacuna para estas qualificações citam-se, dentre outras, as ausências dos fatores culturais e sócio-econômico envolvidos. Ademais, não há uma delineada quantificação da característica. A utilização dos conceitos de lógica nebulosa permitiu quantificar a violência e, igualmente, inserir outros fatores importantes na sua caracterização, promovendo a interdisciplinaridade e transversalidade que o tema exige.

O resultado deste trabalho é a tradução dos registros das ocorrências policiais em níveis e importâncias relevantes sob o aspecto da violência. O rol de avaliação dos níveis de violência, antes reduzido a número e local excluía indicadores como datas e horários das ocorrências, permitiram prover recursos de análise que ampliaram a observação. A sua significância pode acentuar ou amenizar a questão do número de violência de um determinado local.

A saída do primeiro controlador permitiu verificar o nível de violência do local considerado, segundo o tipo penal escolhido. Como as ocorrências foram georeferenciadas, o mapeamento das ocorrências foi substituído por um gráfico de níveis de violência que indica o grau de criminalidade do local frente às demais regiões. Além disso, a análise propicia avaliar uma região tanto de forma isolada quanto de forma globalizada. Assim, fatos pontuais não serão tão representativos quanto pareceriam se fossem consideradas ocorrências isoladas.

A saída do segundo controlador traduz os níveis de violência segundo a sua significância no contexto. Esse resultado permite avaliar cada fato delituoso da região observada e verificar nível de significância no contexto de todas as ocorrências existentes. Como consequência, pode-se determinar quais indicadores são mais relevantes, permitindo iniciar o processo de inferência.

O modelo mostrou-se bastante flexível, permitindo a inclusão de outros indicadores candidatos a qualquer tempo, submetendo-os à prova de significância individual e coletiva. Esta situação permite a inclusão de novos elementos na fase de registro ou investigação das ocorrências, objetivando prover aos órgãos de interesse um conhecimento mais apurado das possíveis causas da violência e o seu estágio atual no âmbito considerado.

Como contribuição, o trabalho permite iniciar o processo de inferência, a fim de possibilitar a previsão de acontecimento de certos delitos, além de viabilizar o planejamento das ações de combate à violência e criminalidade baseando-se no conhecimento (não somente nos dados) da região considerada.

Referências

- Cox, E.**, *The fuzzy systems handbook: a practitioner's guide to building, using, and maintaining fuzzy systems*. Academic Press Limited: USA, 1994.
- Einstein A., Born M., e Born H.**, *Correspondance: 1916-1955*. Seuil: Paris, 1988.
- Fonseca, J. S. e Martins, G. A.**, *Curso de Estatística*, 3ª Ed. Atlas: São Paulo, 1982.
- Klir, G. J. e Yuan, B.**, *Fuzzy Sets, Fuzzy Logic and Fuzzy Systems*. World Scientific, 1996.
- Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B. e Lozano, R.**, *World report on violence and health*. World Health Organization: Geneva, 2002.
- Lobo, L. F. e Fernandez, J. C.**. *A Criminalidade na região Metropolitana de Salvador*. XXXI Encontro Nacional de Economia da ANPEC, 2003.
- McEntire M.** *The Blood of Abel: the Violent Plot in the Hebrew Bible*. Mercer University Press: Macon, 1999.
- McNeill, F. M. e Thro, E.** *Fuzzy Logic: a practical approach*. AP Professional: London, 1994
- Passino, K. M. e Yurkovich, S.**, *Fuzzy control*. Addison Wesley: California, 1998.
- Raggio, V.**, *Concepção materialista da história, psicanálise e violência*. Vozes: Petrópolis, 1992.
- Ruan D. e Huang C.**, *Advances in Fuzzy Mathematics and Engineering: Key selectec papers by Lotfi A. Zadeh*. Beijing Normal University Press: Beijing, 2000.