

Andreia Malucelli^IAltair von Stein Junior^{II}Laudelino Bastos^{III}Deborah Carvalho^{IV}Marcia Regina Cubas^VEmerson Cabrera Paraíso^I

Classificação de microáreas de risco com uso de mineração de dados

Classification of risk micro-areas using data mining

RESUMO

OBJETIVO: Identificar, com o auxílio de técnicas computacionais, regras referentes às condições do ambiente físico para a classificação de microáreas de risco.

MÉTODOS: Pesquisa exploratória, desenvolvida na cidade de Curitiba, PR, em 2007, dividida em três etapas: identificação de atributos para classificar uma microárea; construção de uma base de dados; e aplicação do processo de descoberta de conhecimento em base de dados, por meio da aplicação de mineração de dados. O conjunto de atributos envolveu as condições de infraestrutura, hidrografia, solo, área de lazer, características da comunidade e existência de vetores. A base de dados foi construída com dados obtidos em entrevistas com agentes comunitários de saúde, sendo utilizado um questionário com questões fechadas, elaborado com os atributos essenciais, selecionados por especialistas.

RESULTADOS: Foram identificados 49 atributos, sendo 41 essenciais e oito irrelevantes. Foram obtidas 68 regras com a mineração de dados, as quais foram analisadas sob a perspectiva de desempenho e qualidade e divididas em dois conjuntos: as inconsistentes e as que confirmam o conhecimento de especialistas. A comparação entre os conjuntos mostrou que as regras que confirmavam o conhecimento, apesar de terem desempenho computacional inferior, foram consideradas mais interessantes.

CONCLUSÕES: A mineração de dados ofereceu um conjunto de regras úteis e compreensíveis, capazes de caracterizar microáreas, classificando-as quanto ao grau do risco, com base em características do ambiente físico. A utilização das regras propostas permite que a classificação de uma microárea possa ser realizada de forma mais rápida, menos subjetiva, mantendo um padrão entre as equipes de saúde, superando a influência da percepção particular de cada componente da equipe.

DESCRITORES: Bases de Dados como Assunto. Bases de Dados Factuais. Bases de Conhecimento. Inteligência Artificial. Indicadores Ambientais. Riscos Ambientais. Mapa de Risco.

^I Programa de Pós-Graduação em Informática. Pontifícia Universidade Católica do Paraná (PUC-PR). Curitiba, PR, Brasil

^{II} Secretaria Estadual de Saúde do Paraná. Curitiba, PR, Brasil

^{III} Programa de Pós-Graduação em Computação Aplicada. Universidade Tecnológica Federal do Paraná. Curitiba, PR, Brasil

^{IV} Instituto Paranaense de Desenvolvimento Econômico. Curitiba, PR, Brasil

^V Programa de Pós-Graduação em Tecnologia em Saúde. PUC-PR. Curitiba, PR, Brasil

Correspondência | Correspondence:

Andreia Malucelli
A/C PPGIa
R. Imaculada Conceição, 1155
Prado Velho
80215-901 Curitiba, PR, Brasil
E-mail: malu@ppgia.pucpr.br

Recebido: 18/5/2009
Revisado: 29/9/2009
Aprovado: 14/10/2009

ABSTRACT

OBJECTIVE: To identify, with the assistance of computational techniques, rules concerning the conditions of the physical environment for the classification of risk micro-areas.

METHODS: Exploratory research carried out in Curitiba, Southern Brazil, in 2007. It was divided into three phases: the identification of attributes to classify a micro-area; the construction of a database; and the process of discovering knowledge in a database through the use of data mining. The set of attributes included the conditions of infrastructure; hydrography; soil; recreation area; community characteristics; and existence of vectors. The database was constructed with data obtained in interviews by community health workers using questionnaires with closed-ended questions, developed with the essential attributes selected by specialists.

RESULTS: There were 49 attributes identified, 41 of which were essential and eight irrelevant. There were 68 rules obtained in the data mining, which were analyzed through the perspectives of performance and quality and divided into two sets: the inconsistent rules and the rules that confirm the knowledge of experts. The comparison between the groups showed that the rules that confirm the knowledge, despite having lower computational performance, were considered more interesting.

CONCLUSIONS: The data mining provided a set of useful and understandable rules capable of characterizing risk areas based on the characteristics of the physical environment. The use of the proposed rules allows a faster and less subjective area classification, maintaining a standard between the health teams and overcoming the influence of individual perception by each team member.

DESCRIPTORS: Databases as Topic. Databases, Factual. Knowledge Bases. Artificial intelligence. Environmental Indicators. Environmental Risks. Risk Map.

INTRODUÇÃO

A descentralização como princípio para construção do Sistema Único de Saúde inclui estratégias para mudança de modelo assistencial, dentre elas, a gestão em saúde a partir de uma abordagem territorial. Neste domínio, o território é entendido não apenas como espaço geográfico estanque, mas como território-processo, um espaço social no qual os homens, com suas características pessoais, se associam com outros homens, em movimentos sociais de transformação do seu território.⁶

Como base para elaboração de ações, as equipes de saúde realizam um processo de apropriação e análise de dados referentes às condições da comunidade de seu território de atuação, denominado de territorialização. Esse processo consiste na coleta sistemática de dados demográficos, socioeconômicos, político-culturais, epidemiológicos e sanitários, utilizados para construção de mapas básicos ou temáticos. Além de iniciar ou fortalecer o vínculo entre a equipe de saúde e a

comunidade, este processo identifica as delimitações de pequenos espaços de formato assimétrico, denominados de microáreas.^{8,10}

As microáreas são definidas como uma subdivisão de pequena extensão do território da Unidade Básica de Saúde, na qual seus habitantes possuem uma condição de vida homogênea, que pode determinar riscos à saúde.⁶

Os riscos de uma microárea podem ser classificados em diferentes níveis, dependendo das características que expõem os moradores aos riscos ou determinam o desenvolvimento de agravos à saúde.

O reconhecimento das microáreas de risco é fundamental para estabelecer prioridades a serem trabalhadas pelas equipes de saúde, assim como para planejar ações adequadas aos reais problemas da comunidade.⁶ Para tanto, são utilizadas fontes de dados primárias, que podem ser fruto de entrevistas com informantes-chaves,

moradores da região; e fontes secundárias, oriundas de diversas bases de dados de sistemas de informação das secretarias municipais, ou de outros órgãos governamentais ou não governamentais.⁹

Uma técnica sugerida para coleta dos dados que identifica microáreas de risco é a estimativa rápida, a qual propõe etapas de levantamento de informações; preparo de questionários; conhecimento do território; formulação de hipóteses de subdivisões do território em microáreas; e identificação de informantes-chaves da comunidade para validar as informações coletadas.⁶

O resultado efetivo do processo de territorialização e, por consequência, a delimitação das microáreas ainda é fruto de análises subjetivas do conjunto de dados. Atualmente, as microáreas de risco são delimitadas pela equipe de saúde, apoiadas pelos agentes comunitários de saúde (ACS), conhecedores dos problemas locais, uma vez que os vivenciam por serem moradores da região.

Nessa perspectiva, a utilização de estratégias de análise da situação de saúde em áreas com condições de vida similares pode auxiliar na identificação e priorização de problemas de saúde. Da mesma forma, pode contribuir para a adoção de estratégias de intervenção intersetorial, capazes de modificar as condições de vida além das ações relacionadas com o “cuidado de saúde” propriamente dito.⁷

Perante este contexto e pela importância das informações decorrentes da análise dos dados de um território, este é um campo em que a área da computação pode oferecer suporte por meio de técnicas e ferramentas de manejo de dados, dentre elas, o processo de descoberta de conhecimento em base de dados – KDD (do inglês, *Knowledge Discovery from Database*).

O KDD é um processo que busca identificar padrões, associações, modelos ou informações relevantes, que permanecem ocultos em bases, repositórios e outras formas de armazenamento de dados. Permite identificar padrões válidos, novos, potencialmente úteis e compreensíveis e envolve diversas áreas da ciência como aprendizado de máquina, base de dados, estatística, reconhecimento de padrões, visualização, entre outras.²

Atualmente, o KDD é aplicado em diversas áreas como administração, análise de marketing e medicina.³ No entanto, para que os padrões identificados constituam, de fato, fonte para formar novos conhecimentos aplicáveis para o suporte à decisão, é importante que sejam interessantes, úteis e compreensíveis ao potencial gestor.

O KDD é composto das seguintes etapas: pré-processamento, mineração de dados e pós-processamento. A etapa de pré-processamento é considerada de grande importância e tem como objetivo adequar as bases para

extrair padrões. Após a etapa de pré-processamento, inicia-se a fase de mineração de dados (*data mining*), a qual é considerada como a etapa central na descoberta do conhecimento e envolve a escolha e aplicação da ferramenta e algoritmo a ser utilizado. Dentre os possíveis algoritmos a serem utilizados nesta etapa estão os indutores de regras, os algoritmos genéticos, entre outros.⁴ Por fim, ocorre o pós-processamento, no qual os resultados obtidos são analisados e interpretados. Nesta fase os padrões descobertos são avaliados no sentido de verificar se satisfazem o critério necessário para constituir um elemento importante para o apoio à decisão.

Considerando que a análise para classificar microáreas de risco se constitui em um processo subjetivo de manipulação de dados e que a área da informática possui técnicas que podem tornar essa manipulação objetiva, o presente estudo teve por objetivo identificar, com o auxílio de técnicas computacionais, regras referentes as condições do ambiente físico capazes de contribuir para a classificação de microáreas homogêneas de risco.

MÉTODOS

Estudo exploratório em três etapas, desenvolvido na cidade de Curitiba, PR, em 2007.

Etapa 1 – identificação de atributos para a classificação de microáreas, cuja listagem inicial foi obtida a partir de revisão de literatura. Os grupos de atributos foram submetidos à validação por oito especialistas da área da saúde coletiva, sendo cinco enfermeiros e três médicos, que os classificaram em: atributos irrelevantes, importantes ou essenciais. Os critérios para a seleção dos especialistas foram: profissionais com atuação na saúde pública por, no mínimo, dois anos; vinculados à área acadêmica e com titulação mínima de mestre.

Etapa 2 – construção da base de dados. A partir dos atributos considerados essenciais pelos especialistas foi elaborado um questionário para coleta de dados a ser aplicado para os ACS da rede municipal de saúde. Foram excluídos apenas os ACS que, no momento da coleta, estavam de férias, licença ou ausentes de suas atividades. Os dados foram organizados em planilha eletrônica formando uma base de dados com 531 registros referentes ao ambiente físico das microáreas, representando uma amostra de 46,2% do total de microáreas em Curitiba.

Etapa 3 – aplicação do processo de KDD. Esta etapa seguiu as fases de pré-processamento por meio das tarefas de limpeza, seleção e transformação dos dados. Para a etapa de mineração de dados foi utilizada a ferramenta *Waikato Environment for Knowledge Analysis* (WEKA).⁸ Por se tratar de um problema de classificação, foi utilizado o algoritmo J48, que apresenta

⁸ The University of Waikato. WEKA Version 3.5. [computer program]. [citado 2007 mar 02]. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>

Tabela 1. Conjunto de atributos para classificar uma microárea em relação ao ambiente físico.

Atributo
Equipamentos sociais
1. Escola pública
2. Creche pública
3. Projetos públicos para crianças
4. Igreja (qualquer religião)
5. Policiamento
6. Unidade de saúde
7. Hospital
Comércio
8. Farmácia
9. Mercado
10. Supermercado
11. Mercearia
12. Bar
Hidrografia
13. Rios
14. Cavas
15. Córregos
16. Açudes
Terreno
17. Terreno inclinado
18. Terreno baixo
19. Terreno úmido constantemente
20. Terreno seco
21. Terreno plano
22. Terreno área de aterro
23. Terreno ondulado
24. Terreno úmido ocasionalmente
Infra-estrutura
25. Distribuição de energia elétrica
26. Iluminação pública
27. Abastecimento de água
28. Rede de esgoto
29. Coleta de lixo
30. Malha viária
31. Barreiras de acesso da comunidade
32. Transportes
Comunidade
33. Distribuição das moradias
34. Tipos de moradias
35. Área da construção das casas
36. Microárea considerada violenta
37. Existência de Indústrias poluentes
38. Vazios urbanos
39. Plantações
40. Clima

Continua

Tabela 1 continuação

Atributo
Vetores
41. Vacas
42. Gatos
43. Cachorros
44. Pombos
45. Cavalos
Áreas de lazer
46. Parques
47. Cancha para esporte
48. Praças

o resultado na forma de árvore de decisão, podendo ser transformada num conjunto de regras no formato: “SE... ENTÃO...”.

A avaliação na etapa de pós-processamento foi realizada sob as perspectivas do desempenho computacional e da qualidade do conjunto de regras. Para a avaliação do desempenho computacional foram consideradas as medidas de cobertura e acerto, sendo estas entendidas como:

cobertura: indica o número de exemplos cobertos pelas condições das regras. Uma alta cobertura com alta taxa de acerto pode indicar uma regra de senso comum.

taxa de acerto: apresenta a porcentagem de casos classificados corretamente em relação à cobertura, indicando a credibilidade da regra e foi calculada utilizando a seguinte expressão:

$$\text{Taxa de acerto} = \frac{(\text{cobertura} - \text{erro})}{\text{cobertura}} * 100$$

Nesta expressão o erro é dado pelos casos classificados de forma incorreta pelo algoritmo.

Para avaliação sob a perspectiva da qualidade, foram consideradas as medidas de compreensibilidade e o quão interessante é a regra para especialistas que não estiveram envolvidos na etapa I. Para avaliar a compreensibilidade foi considerado o tamanho da regra, ou seja, o número de condições por regra.

Para avaliar quão interessante são as regras, foram analisadas por três especialistas da área de saúde coletiva, selecionados pelos seguintes critérios: atuantes (assistenciais) na área de saúde pública por mais de dois anos e titulação mínima de especialista em saúde coletiva ou saúde da família. Os especialistas atribuíram para cada regra um dentre três valores possíveis: irrelevante (incompatível com a realidade); confirma seu conhecimento (confirmam o que já se sabe); e interessante (apresentam padrões condizentes com a realidade,

Tabela 2. Medidas para avaliação do desempenho do conjunto de regras obtido na etapa de mineração de dados. Curitiba, PR, 2007.

Avaliação do desempenho	Média (DP)	Mínimo	Máximo	Intervalo
Quantidade de condições por regra	5,74 (2,11)	2	12	10
Cobertura	7,84 (15,00)	1	111	110
Taxa de acerto	91,6% (14,00)	33,3%	100%	66,7%

mas até então desconhecidos). A estimativa do “quão interessante foi a regra” foi elaborada com base no valor atribuído pelos especialistas, em que, quanto maior a mediana, mais interessante é a regra.

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Pontifícia Universidade Católica do Paraná e pelo Comitê de Ética em Pesquisa da Secretaria Municipal de Saúde de Curitiba.

RESULTADOS

A partir de uma lista de 49 atributos (Tabela 1), os especialistas incluídos na etapa de identificação dos atributos classificaram 41 como essenciais e oito como irrelevantes. O conjunto de atributos envolve as condições de infra-estrutura, hidrografia, solo, área de lazer, características da comunidade e vetores. Foram considerados irrelevantes os atributos: supermercado, mercearia, bar, praças, terreno ondulado, clima, terreno úmido ocasionalmente e hospital.

Informações coletadas pelo questionário compuseram a base de dados, referente a 46,2% de um total de 1.149 ACS de Curitiba. Em alguns questionários foram assinalados múltiplos valores para um mesmo atributo, dificultando assim a classificação correta da microárea.

Para melhorar o desempenho do processo de mineração de dados, alguns atributos necessitaram de transformação dos valores. Como exemplo, o atributo “distribuição de energia elétrica” poderia ser assinalado duplamente: “regularizada” e “clandestina”, desta forma o atributo passou a ser denominado “distribuição de energia elétrica regularizada”, com as opções “sim”, “não” e “parcialmente”.

Esta transformação permitiu aumentar a taxa de acerto do classificador, de 87,5% para 88,7%; diminuir o número de regras geradas, de 130 para 79; diminuir o número de regras não cobertas pelos exemplos na base de dados, de 57 para 10; e apresentar regras de mais fácil compreensão devido ao valor do atributo no antecedente da regra ser mais objetivo.

As 68 regras obtidas foram expressas no seguinte formato:

Antecedente: → *SE* <condição>

Conseqüente: *ENTÃO* <conclusão>

Os resultados estatísticos referentes à cobertura, a taxa de acerto e quantidade de condições por regra representando as medidas de desempenho e qualidade consideradas para avaliar o conjunto composto de 68 regras são apresentados na Tabela 2.

Considerando a mediana do valor atribuído à relevância da regra, o conjunto foi dividido em dois, denominados de Conjunto A, composto de 37 regras, consideradas inconsistentes – valor da mediana igual a um; e Conjunto B, composto por 31 regras que confirmam o conhecimento dos especialistas – valor da mediana igual a dois (Tabela 3). A Tabela 4 apresenta uma comparação entre o Conjunto A (regras consideradas inconsistentes) e o Conjunto B (regras que confirmam o conhecimento do especialista). Apesar de o número máximo da amostragem para a quantidade de condições por regras ser igual a 12, o Conjunto A possui 66,7% de regras próximas à média de 5,89 (DP=2,4). Em média, a quantidade de condições por regra no Conjunto A é discretamente maior em relação ao Conjunto B.

Em relação ao conjunto total de regras, a média (5,74; DP=2,11) da quantidade de condições por regra manteve-se próxima ao conceito de regra prática, ou seja, 73,5% se mantiveram entre quatro e sete condições. A média da taxa de acerto para o conjunto de regras foi de 91,6%, com (DP=14,00), mostrando um resultado satisfatório para o conjunto obtido.

As regras obtidas também foram utilizadas para identificar os atributos que melhor diferenciam as microáreas em baixo, médio ou alto risco. Assim, os atributos posicionados nas cinco primeiras condições e com maior frequência foram considerados como os mais importantes (Tabela 5).

Dos atributos apresentados como essenciais pelos especialistas na etapa de identificação, seis não apareceram como condições nas regras, sendo eles: policiamento; gato; cachorro; pombos; vagas suficientes em creches; e iluminação pública. Logo, estes podem ser considerados com menor peso para a classificação das microáreas.

Tabela 3. Regras que confirmam o conhecimento dos especialistas (etapa de identificação dos atributos). Curitiba, PR, 2007.

Identificação	Descrição do antecedente	Conseqüente
Regra 1	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemente = não E Abastecimento de água por rede pública = sim E Córrego = contaminado E Terreno plano = sim E Projetos públicos para crianças = presente	Baixo_risco (7.0/2.0)
Regra 2	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemente = não E Abastecimento de água por rede pública = sim E Córrego = contaminado E Terreno plano = sim E Projetos públicos para crianças = ausente	Médio_risco (9.0/1.0)
Regra 3	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemene = não E Abastecimento de água por rede pública = sim E Córrego = contaminado E Terreno plano = não	Baixo_risco (15.0/1.0)
Regra 4	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemente = não E Abastecimento de água por rede pública = sim E Córrego = não_contaminado	Baixo_risco (3.0/1.0)
Regra 5	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemente = não E Abastecimento de água por rede pública = sim E Córrego = ausente	Baixo_risco (111.0/17.0)
Regra 7	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = legalizada E Terreno úmido constantemente = sim E Córrego = contaminado	Médio_risco (6.0/2.0)
Regra 9	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = não_legalizada E Terreno área de aterro = não E Indústrias poluentes = presente	Baixo_risco (5.0/1.0)
Regra 10	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = não_legalizada E Terreno área de aterro = não E Indústrias poluentes = ausente	Médio_risco (11.0/2.0)
Regra 11	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = não_legalizada E Terreno área de aterro = sim	Baixo_risco (5.0)
Regra 12	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = sim E Área da construção das casas = em_processo_de_legalização E Terreno inclinado = não	Baixo_risco (3.0)
Regra 14	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = parcialmente E barreira de acesso à unidade de saúde = não E Distribuição de energia elétrica regularizada = sim E Distribuição das moradias organizadas = sim E Malha viária de asfalto = sim	Baixo_risco (8.0)
Regra 15	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = parcialmente E arreira de acesso à unidade de saúde = não E Distribuição de energia elétrica regularizada = sim E Distribuição das moradias organizadas = sim E Malha viária de asfalto = parcialmente	Médio_risco (3.0/1.0)
Regra 18	SE microárea é considerada violenta = não E Destino do lixo coletado pela prefeitura = parcialmente E barreira de acesso à unidade de saúde = não E Distribuição de energia elétrica regularizada = sim E Distribuição das moradias organizadas = não	Médio_risco (4.0)
Regra 23	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = sim E Córrego = contaminado	Médio_risco (44.0/4.0)
Regra 25	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = sim E Córrego = ausente E Igreja (qualquer religião) = presente E Cancha para esporte = presente	Médio_risco (21.0/6.0)
Regra 29	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = sim E Córrego = ausente E Igreja (qualquer religião) = presente E Cancha para esporte = ausente E Destino do lixo coletado pela prefeitura = sim E Terreno plano = sim E Vazios urbanos com presença de lixo = sim	Baixo_risco (4.0)
Regra 34	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = não	Médio_risco (3.0/1.0)
Regra 36	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = parcialmente	Médio_risco (42.0/6.0)
Regra 39	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = parcialmente E Plantações = sem_utilização_de_agrotóxicos E Vazios urbanos com presença de lixo = sim	Médio_risco (14.0/3.0)

Continua

Tabela 3 continuação

Identificação	Descrição do antecedente	Conseqüente
Regra 40	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = ausente E Malha viária de asfalto = parcialmente E Plantações = com_utilização_de_agrotóxicos	Médio_risco (4.0)
Regra 44	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = contaminado E Farmácia = ausente E Distribuição das moradias organizadas = parcialmente	Alto_risco (4.0)
Regra 45	SE microárea é considerada violenta = razoável E Ratos = sim E Cava = contaminado E Farmácia = ausente E Distribuição das moradias organizadas = não	Alto_risco (2.0)
Regra 49	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = encanado E Terreno área de aterro = não E Rio = contaminado, E barreira de acesso à unidade de saúde = não	Médio_risco (21.0/1.0)
Regra 50	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = encanado E Terreno área de aterro = não E Rio = contaminado E barreira de acesso à unidade de saúde = sim E Mercado = presente	Médio_risco (2.0)
Regra 51	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = encanado E Terreno área de aterro = não E Rio = contaminado E barreira de acesso à unidade de saúde = sim E Mercado = ausente	Alto_risco (2.0)
Regra 61	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = misto E Terreno área de aterro = não E Parque ou praça = ausente E Área da construção das casas = não_legalizada E Destino do lixo coletado pela prefeitura = sim	Alto_risco (3.0)
Regra 63	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = misto E Terreno área de aterro = não E Parque ou praça = ausente E Área da construção das casas = em_processo_de_legalização	Alto_risco (8.0)
Regra 64	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = misto E Terreno área de aterro = sim	Alto_risco (10.0)
Regra 66	SE microárea é considerada violenta = sim E Cava = ausente E Rede de esgoto = a_ceu_aberto	Alto_risco (3.0/1.0)
Regra 67	SE microárea é considerada violenta = sim E Cava = não_contaminado	Médio_risco (1.0)
Regra 68	SE microárea é considerada violenta = sim E Cava = contaminado	Alto_risco (9.0)

DISCUSSÃO

A identificação de áreas homogêneas de risco auxilia na priorização de ações coletivas, com ênfase na prevenção de doenças, direcionadas aos espaços territoriais onde as iniquidades são maiores, o que resulta num maior impacto sobre as condições de risco.¹

Assim, alguns dos atributos indicados como essenciais podem ser alterados por políticas públicas, dentre os quais, os equipamentos sociais e as áreas de lazer. Outros atributos são modificáveis com ações

intersetoriais e com a contribuição da comunidade, como a existência de vetores. Assim, um sistema de informação que monitore as condições destes atributos pode colaborar para o planejamento de ações em nível local, regional e central. Ainda, permite apontar algumas condições a serem melhoradas por meio do incentivo à própria comunidade.

O fato de a quantidade de condições por regra no Conjunto A ser discretamente maior em relação ao Conjunto B pode indicar que a complexidade das regras

Tabela 4. Comparação entre o desempenho dos conjuntos de regras inconsistentes e confirmadas pelos especialistas. Curitiba, PR, 2007.

Medida	Quantidade de condições por regra		Cobertura		Taxa de acerto	
	A	B	A	B	A	B
Média	5,89	5,55	3,84	12,61	93,7%	88,9%
Desvio-padrão	2,40	1,71	2,65	21,23	14,4	13,2
Intervalo	10	8	11	110	66,7%	33,3%
Mínimo	2	2	1	1	33,3%	66,7%
Máximo	12	10	12	111	100,0%	100,0%

A: Regras inconsistentes

B: Regras confirmadas pelos especialistas

Tabela 5. Relação entre o atributo e sua posição na regra. Curitiba, PR, 2007.

Atributo	Posição na condição da regra
Microárea é considerada violenta	1
Cava	2
Destino do lixo coletado pela prefeitura	2
Ratos	2
Área da construção das casas	3
Barreira de acesso à unidade de saúde	3
Rede de esgoto	3
Distribuição de energia elétrica regularizada	4
Farmácia	4
Malha viária de asfalto	4
Terreno área de aterro	4
Terreno baixo	4
Terreno úmido constantemente	4
Indústrias poluentes	4
Porco	4
Terreno inclinado	4
Rio	5
Abastecimento de água por rede pública	5
Cancha para esporte	5
Cavalos	5
Córrego	5
Distribuição das moradias organizadas	5
Igreja (qualquer religião)	5
Parque ou praça	5
Plantações	5

não interferiu no momento em que os especialistas as avaliaram. Isso permite descartar a hipótese de que houve dificuldade de interpretação pelos especialistas, ao julgarem as regras deste conjunto como inconsistentes.

Entretanto, a média para cobertura do Conjunto B indica que há maior probabilidade de haver regras consideradas como senso comum. Apesar de se tratar de um conjunto com regras que confirmam o conhecimento dos especialistas, a taxa de acerto não superou a do Conjunto A.

A média de cobertura do Conjunto B foi muito superior em relação ao Conjunto A. Esta relação sugere que as regras contidas no Conjunto B tendem a representar o

senso comum, o que foi confirmado pela opinião dos especialistas ao indicarem-nas como a confirmação do seu conhecimento.

Assim, as regras do Conjunto B (Tabela 3), apesar de terem menor desempenho computacional pela avaliação dos especialistas, são as que melhor classificam uma microárea em relação ao risco contido no ambiente físico.

Esta inesperada divergência entre a opinião dos especialistas e as medidas estatísticas indica que este conjunto pode conter algumas regras interessantes. Entretanto, ao avaliar as regras, os especialistas podem ter sido resistentes a aceitar novos padrões ou para compreender modelos que contrapõem o conhecimento prévio.

A diversidade presente, oriunda das diferentes percepções sobre o território, ao ser explicitada e discutida no coletivo dos envolvidos no processo de territorialização, colabora para uma aproximação entre os problemas identificados e suas possíveis resoluções, que deverão ser priorizados coletivamente.⁵

A mineração de dados ofereceu um conjunto de regras úteis e compreensíveis, capazes de caracterizar microáreas, classificando-as quanto ao grau do risco, quando consideradas as características do ambiente físico. No entanto, o ambiente físico não se constitui no único fator para classificar uma microárea, pois, para tornar a classificação efetiva, devem ser acrescentadas informações epidemiológicas da região, organizacionais da comunidade e administrativas.

A utilização das regras propostas permite que a classificação de uma microárea possa ser realizada de forma mais rápida, menos subjetiva e mantendo um padrão entre as equipes de saúde, superando a influência da percepção particular de cada componente da equipe.

Tal influência subjetiva pode ser justificada pelo fato de que diferentes atores sociais participantes do processo de avaliação possuem um conjunto de valores construídos com base na sua experiência e na sua inserção em diferentes contextos cultural, econômico e social. Isso influencia, sobremaneira, a importância de determinados atributos em detrimento de outros.¹¹

Considerando que a classificação das microáreas de risco é uma importante ferramenta gerencial e assistencial, por envolver a distribuição dos recursos e cuidados para a população de um determinado território, realizá-la de forma a convergir a subjetividade inerente deste processo com métodos mais objetivos de análise, permitirá a otimização de ações e recursos.

REFERÊNCIAS

1. Chiesa AM, Westphal MF, Kashiwagi NM. Geoprocessamento e a promoção da saúde: desigualdades sociais e ambientais em São Paulo. *Rev Saude Publica*. 2002;36(5):559-67. DOI:10.1590/S0034-89102002000600004
2. Fayyad U, Piatetsky-Shapiro G; Smyth P. From data mining to knowledge discovery in databases. *AI Magazine*. 1996;17(3):37-54.
3. Han J, Kamber M. Data mining: concepts and techniques. San Francisco: Morgan Kaufmann; 2001.
4. Rezende SO, Plugliesi JB, Melanda EA, de Paula MF. Mineração de dados. In: Rezende SO, editor. *Sistemas inteligentes: fundamentos e aplicações*. Barueri: Manole; 2005. p.307-35.
5. Ribeiro PT. Direito à saúde: integridade, diversidade e territorialidade. *Cienc Saude Coletiva*. 2007;12(6):1525-32. DOI:10.1590/S1413-81232007000600014
6. Silva AMR, Oliveira MSM, Nunes EFPA, Torres ZF. A unidade básica de saúde e seu território. In: Andrade SM, Soares DA, Cordoni Junior L, organizadores. *Bases da saúde coletiva*. Londrina: UEL; 2001. p.145-60.
7. Silva LMV, Paim JS, Costa MCN. Desigualdades na mortalidade, espaço e estratos sociais. *Rev Saude Publica*. 1999;33(2):187-97. DOI:10.1590/S0034-89101999000200011
8. Souza CMN, Moraes LRS, Bernardes RS. Doenças relacionadas à precariedade dos sistemas de drenagem de águas pluviais: proposta de classificação ambiental e modelos causais. *Cad Saude Coletiva (Rio J)*. 2005;13(1):157-68.
9. Takeda S. A organização de serviços de atenção primária à saúde. In: Duncan BB, Schmidt MI, Giuciani WRJ, organizadores. *Medicina ambulatorial: condutas clínicas em atenção primária baseadas em evidências*. 3.ed. Porto Alegre: Artmed; 2004. p.76-87.
10. Teixeira CF. Promoção e vigilância da saúde no contexto da regionalização da assistência à saúde no SUS. *Cad Saude Publica*. 2002;18(Supl):153-62. DOI:10.1590/S0102-311X2002000700015
11. Uchimura KY, Bosi MLM. Qualidade e subjetividade na avaliação de programas e serviços em saúde. *Cad Saude Publica*. 2002;18(6):1561-9. DOI:10.1590/S0102-311X2002000600009

Artigo baseado em dissertação de mestrado de Von Stein Júnior A, apresentada ao Programa de Pós-Graduação em Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná, em 2008.
Os autores declaram não haver conflito de interesses.