

Um Modelo para Derivação de Relacionamentos Espaciais em Equivalentes Semânticos Relacionais

WESLEY VAZ E GEOVANE CAYRES MAGALHÃES

Instituto de Computação - UNICAMP - CP. 6176 - 13081-970 - Campinas, SP, Brasil

wvaz,geovane@ic.unicamp.br

Abstract. Some kinds of applications that deals with spatial data requires information about relationships of spatial objects. Spatial data mining methods, for instance, can involve spatial analysis and statistical methods directly on spatial objects, increasing the complexity to obtain solutions. This paper proposes a model to transform relationships between spatial objects in relational datasets semantically equivalent. The main purpose of this model is to obtain topological and distance-based relationships on instances of objects, transforming them in a set of predicates that references a specific context. Applications that works only with these relationships can use this repository of spatial predicates instead to work directly on spatial objects, increasing modularity and decreasing complexity on solution's obtention.

1 Introdução

A quantidade de dados geográficos armazenados e coletados têm crescido nos dias atuais. Isso se deve, dentre outros, a demanda de aplicações que utilizam dados espaciais, como as de controle ambiental e telecomunicações, por exemplo. A utilização da variável espaço em um vasto número de aplicações motiva o desenvolvimento de técnicas para armazenamento, manipulação e análise de dados georeferenciados.

Algumas das aplicações em destaque são as que envolvem ferramentas de *geomarketing*. Essas aplicações auxiliam o processo de tomada de decisões estratégicas através da descoberta de relacionamentos implícitos e interessantes em contextos onde a variável espacial é importante [9].

O desenvolvimento de aplicativos que implementam essas funcionalidades de *geomarketing* em geral envolve a criação de métodos estatísticos para *obtenção de conhecimento implícito, previamente desconhecido em dados espaciais* [12], denominado "*Data Mining Espacial*". Alguns métodos de mineração de dados espacial fazem a análise dos relacionamentos entre os objetos espaciais e análise estatística dos dados de forma simultânea [14]. Essa simultaneidade acarreta um aumento da complexidade na obtenção do "conhecimento" e no tempo de execução dos algoritmos.

Entretanto, um método alternativo visando transformar os relacionamentos entre dados espaciais em dados convencionais pode ser considerado. A partir de um conjunto de predicados espaciais, algoritmos tradicionais de mineração de dados podem ser aplicados diretamente sobre eles. A criação desse repositório visa aumentar a escalabilidade e a modularidade do processo de mineração dos dados espaciais. A utilização de um repositório convencional contendo apenas os relacionamentos relevantes entre os objetos espa-

ciais faz com que não seja necessário modificar algoritmos de mineração de dados tradicionais para lidar especificamente com dados espaciais.

Um exemplo de utilização deste método é a obtenção de regras de associação espacial utilizando predicados referentes aos relacionamentos entre os objetos espaciais. A obtenção destes predicados é realizada através do modelo de derivação a ser proposto. Regras como "*90% dos clientes que moram na Região X trabalham próximos à Região Y*" podem ser obtidas através da execução de algoritmos tradicionais de obtenção de regras de associação [1, 13] sobre os predicados espaciais obtidos entre as instâncias dos objetos espaciais.

A proposta deste trabalho é ilustrar um modelo de derivação de relacionamentos espaciais em seus equivalentes semânticos no formato relacional. As principais contribuições do modelo são as seguintes:

1. Modelagem de estruturas que possibilite ao usuário escolher os objetos e os relacionamentos espaciais a serem verificados;
2. Mapeamento do conhecimento do usuário do domínio da aplicação para definir semântica entre os relacionamentos espaciais;
3. Otimização na execução das operações espaciais de verificação para a obtenção de predicados espaciais.

Este trabalho propõe um modelo a obtenção de predicados espaciais em relacionamentos baseados em distância e topológicos. através da verificação dos resultados e da ordem de execução dessas operações, os predicados espaciais são gerados e associados a cada instância dos objetos espaciais. Esses predicados são obtidos mediante a verificação

dos dados originais frente à base de conhecimento do domínio (mapeados através de hierarquias de conceito) e pela escolha dos objetos espaciais e dos relacionamentos a serem verificados (fornecidos pelo usuário), de forma a restringir os predicados de acordo com o contexto da aplicação.

Esse artigo é organizado como se segue: na Seção 2 serão abordados os trabalhos da literatura relacionados à obtenção de um conjunto mínimo de relacionamentos topológicos e seus modelos formais. A Seção 3 descreve o Modelo de Derivação Espacial e sua funcionalidade de acordo com os relacionamentos (topológicos e de distância). A Seção 4 finaliza o artigo com as conclusões do trabalho e suas futuras extensões.

2 Trabalhos Relacionados

Modelos formais para objetos e operações espaciais têm sido pesquisados no intuito de, dentre outros, prover a interoperabilidade entre SIGs [4, 8, 16, 11, 2, 5]. Além dessa funcionalidade, os modelos auxiliam a obtenção dos relacionamentos entre os objetos através de operações espaciais de uma maneira eficiente. A caracterização formal desses relacionamentos possibilita que novas técnicas e modelos sejam criados para descrever e obter "predicados espaciais".

De forma intuitiva, define-se que predicados espaciais topológicos são aqueles provenientes da descrição de relacionamentos topológicos entre dois objetos espaciais. O predicado ("**Objeto X**", *toca*, "**Objeto Y**") é um predicado topológico por ter como característica a presença do relacionamento "toca". Este subconjunto de relacionamentos espaciais caracterizados pela propriedade de ser preservada mediante transformações topológicas, como translação, rotação e mudança de escala são informalmente caracterizados relacionamentos topológicos [5].

Várias descrições formais de relacionamentos topológicos entre dois objetos espaciais podem ser encontradas em [3, 7, 18]. Na descrição do método *4-intersection* [6], o autor tenta obter semântica exata dos relacionamentos topológicos binários. Esse método tenta distinguir os relacionamentos somente verificando as interseções vazias ou não-vazias entre de fronteiras e interiores de objetos geométricos. Esse método resulta em vários tipos de relacionamentos diferentes, pouco inteligíveis para o usuário. Uma extensão de *4-intersection*, denominado "Método de Dimensão Estendida", considera também a dimensão do resultado das interseções entre os objetos. No entanto, a lista de casos resultantes dessa abordagem resulta em um total de 52 casos reais, o que dificulta a interpretação humana para os relacionamentos topológicos [5]. Mark et. al em [20, 19] descrevem os resultados de experiências sobre como as pessoas "entendem" os relacionamentos entre objetos espaciais do tipo linha e polígono. Nestes trabalhos, 19 rela-

cionamentos válidos (gerados através do modelo formal *9-intersection*) são fornecidos para que os usuários os analisem. Os resultados confirmam que o método usado tem potencial para definir semanticamente predicados espaciais entendíveis ao usuário.

Em [5], os autores propõem um método cujo objetivo é manter o número de relacionamentos topológicos potenciais o menor possível, tentando ser o subconjunto completo mais entendível para os usuários finais. Essa meta foi obtida através de operações de interseção de fronteiras e interiores, além da análise da dimensão do resultado, resultando em cinco relacionamentos mais gerais: *toca*, *está-contido*, *cruza*, *sobreposição* e *disjunto*. Essa abordagem é chamada "Método Baseado em Cálculo" e verifica relacionamentos topológicos entre objetos espaciais do tipo ponto, linha e polígono.

O autor demonstra que este é o menor conjunto de relacionamentos capaz de representar todos os casos do "Método de Dimensão Estendida" sobre a condição que somente operadores de fronteira de linhas e áreas são considerados. Esse conjunto de relacionamentos é pequeno e, portanto, de utilização normal para o ser humano e poderoso o suficiente para representar uma grande variedade de casos.

Outras descrições formais de relacionamentos topológicos podem ser obtidas em [21, 11, 2]. Elas se utilizam de lógica proposicional e um modelo descritivo lógico para formação de operadores de predicados. Além disso, alguns trabalhos fazem uso dessas descrições para a tentativa de otimização de "queries" que envolvam relacionamentos topológicos [15, 21, 16].

No caso deste artigo, utilizamos a descrição formal de Clementini et. al [5] para a obtenção de predicados topológicos a partir da execução de operações entre os objetos. A modificação na ordem a qual essas operações são executadas pode otimizar o processamento a partir dos dados de entrada fornecidos pelo usuário.

Outra característica importante para a funcionalidade do modelo proposto é a utilização de hierarquias de conceito. Informalmente, essa hierarquia é um conjunto parcialmente ordenado, geralmente mapeado em forma de árvores, onde cada nó corresponde a um nível de generalização de um conceito. A definição formal de hierarquias de conceito pode ser encontrada em [17]. Essa estrutura servirá para generalização e especialização semânticas dos dados originais visando a formação de predicados espaciais baseados em distância.

3 Modelo de Derivação Espacial

Ao conjunto de objetos espaciais com dados relacionais descritivos referentes ao mesmo conceito denomina-se "tema". Portanto, um tema é um conjunto de instâncias de objetos espaciais que possuem dados que descrevem suas carac-

terísticas em relação a uma aplicação. Esses objetos podem estar dispostos em forma de pontos, linhas ou polígonos, como citado anteriormente. Por exemplo, se consideramos um tema "Poste", temos um conjunto de pontos em um mapa, onde cada ponto possui, associado à localização geográfica, as características do poste, como altura, número de linhas instaladas, etc.

Os predicados espaciais são as representações dos relacionamentos entre dois objetos. Potencialmente, todas as instâncias de todos os temas podem ser submetidos a todas as operações espaciais definidas e gerar predicados espaciais entre elas. No entanto, os gastos computacionais das operações espaciais são altos e alguns tipos de relacionamentos podem ser previamente descartados entre dois objetos caso haja um "pré-conhecimento" das características, tanto das operações quanto dos objetos.

Além disso, é possível otimizar a execução e cálculo de operações espaciais entre as instâncias dos temas. Dadas as características das operações e os tipos de objetos espaciais (ponto, linha ou polígono), consegue-se otimizar o processo de obtenção dos predicados. A ordem com que relacionamentos topológicos são verificados, por exemplo, pode resultar na economia de um grande número de operações intermediárias entre as instâncias.

Os relacionamentos que serão considerados neste modelo para a montagem dos predicados espaciais são os topológicos ("está-contido", "cruza", "sobreposição", etc.) e os baseados em distância. Deve-se considerar que um conjunto de relacionamentos pré-definido acima deverá ser escolhido para ser executado entre instâncias de dois temas diferentes, e a execução das operações de verificação será feita de modo a otimizar a obtenção dos predicados. O processo de escolha das triplas (tema, relacionamento, tema) é feita pelo especialista do domínio e a verificação da veracidade dessas triplas é funcionalidade do modelo.

Com a geração dos resultados, obtém-se um conjunto de predicados entre dois objetos espaciais, pertencentes às instâncias de vários temas. A partir daí, propõe-se a substituição do objeto espacial pelo conjunto de predicados associados àquela instância. Cada predicado espacial gerado pode conter uma ou mais instâncias para cada objeto espacial a ele associado. O conjunto desses predicados formará o repositório semântico requerido.

A figura 1 ilustra o Modelo de Derivação Espacial, composto de três módulos: *SpatialC*, *AlgTop* e *AlgDist*. *SpatialC* é responsável por retornar o resultado das verificações do relacionamento "está-contido" sobre as instâncias dos temas pré-definidos pelo usuário. *AlgTop* utiliza o resultado de *SpatialC* e deriva os relacionamentos topológicos entre as instâncias dos temas, também fornecidos pela base de conhecimento em poder do usuário. O último módulo retorna predicados referentes aos relacionamentos de distância entre os objetos espaciais, utilizando as hierar-

quias de conceito contextualizadas à aplicação.

Formalmente, o problema a ser solucionado pelo modelo pode ser descrito abaixo:

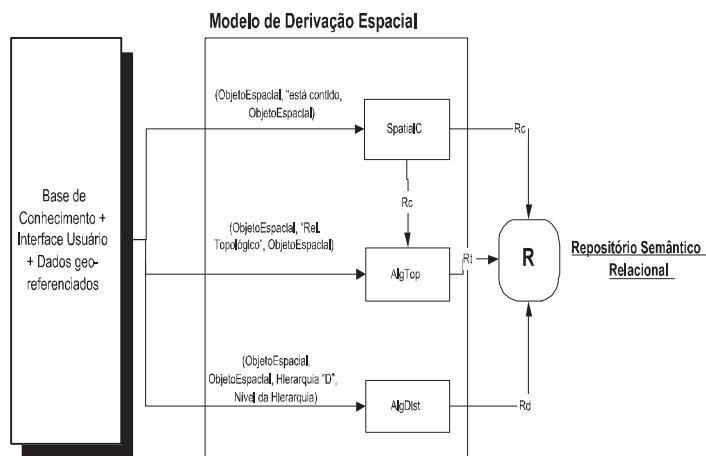


Figura 1: Modelo de Derivação Espacial

Enunciado 1 Seja o número de temas considerado. Considere o conjunto , de cardinalidade , onde cada elemento de é um conjunto de objetos espaciais. Portanto, dado o tema espacial , então é composto das instâncias de todos os objetos espaciais pertencentes ao tema . A instância é dada por . Considere o conjunto de operações topológicas disponíveis e o conjunto das operações espaciais de distância. A partir da escolha da tripla , e (ou), encontrar o conjunto composto das instâncias dos objetos espaciais considerados e dos predicados espaciais () entre eles.

Algorithm 1 SPG

Entrada: O conjunto de triplas e para . Ambos definem qual subconjunto de operações será aplicado sobre os objetos espaciais de cada tema.

Saída: O conjunto-resposta de triplas

Ordenar de acordo com as operações na ordem pré-definida e dimensão dos objetos

Retornar

O algoritmo SPG utiliza três outros algoritmos dependendo das operações pré-definidas no conjunto . Para a operação de "está-contido", o conjunto é verificado no intuito de aplicar aos temas o algoritmo *SpatialC*. De forma análoga, para os temas pré-definidos a gerarem predicados topológicos, executa-se *AlgTop* em cada instância. Para temas os quais a operação de distância foi requerida entre suas instâncias, o algoritmo *AlgDist* recebe os objetos e as hierarquias de conceito referentes à semântica das distâncias entre temas e retorna triplas com os predicados. A união dessas triplas forma o conjunto-repositório de pares de objetos e predicados espaciais.

3.1 Relacionamentos Topológicos

Clementini et. al, em [5], propõe um método baseado em cálculo para manter o conjunto de relacionamentos topológicos entre geo-objetos o menor possível. Seu modelo agrupa todos os casos possíveis em um conjunto pequeno de relações topológicas que podem ser derivados para a obtenção de novos relacionamentos. Esses relacionamentos topológicos mais genéricos são os equivalentes as operações: "toca", "está-contido", "cruza", "sobreposição" e "está-disjunto".

Duas propriedades dos relacionamentos topológicos são importantes para que possa ser criada uma estratégia de minimização de operações espaciais para a obtenção de predicados topológicos. A primeira é a de transitividade, que será explicada na próxima sub-seção. A segunda propriedade é a de simetria, que garante que o resultado de uma operação entre um objeto espacial e , dada por produz o mesmo resultado que . A propriedade de transitividade é encontrada na operação "está-contido", enquanto todos os outros quatro relacionamentos aqui considerados (*toca*, *cruza*, *sobreposição* e *está-disjunto*) possuem a propriedade de simetria entre os relacionamentos.

3.1.1 Relacionamento "Está Contido"

O cálculo do relacionamento topológico "está-contido" pode ser bastante otimizado de forma a reduzir o espaço de busca aplicando transitividade nos resultados prévios.

Devemos atentar para duas propriedades importantes da operação "está-contido" em objetos espaciais:

1. A operação "está-contido" é transitiva, ou seja, se e então .
2. A operação "está-contido" é complementar à operação "contém" (). Portanto se então .

Baseado nestas propriedades, o objetivo é propor um método para eliminar operações espaciais desnecessárias entre as instâncias dos temas considerados, ou seja, obter

da melhor forma o super-conjunto de pares de objetos que farão parte do predicado "Objeto *está-contido-em* Objeto" derivado para a forma relacional.

O método utilizará os resultados do algoritmo *PoliCPoli* além das informações de transitividade e hierarquia entre os temas. Os níveis de especialização e generalização entre objetos espaciais nesse algoritmo são considerados através da análise das dimensões dos objetos. Objetos do tipo ponto são mais específicos do que os do tipo linha, que por sua vez são mais específicos do que os poligonais.

1. Algoritmo PoliCPoli

Algorithm 2 *PoliCPoli*

Entrada: O conjunto de temas poligonais, onde é o número de temas poligonais considerados

Saída: O conjunto de respostas composto dos resultados entre os pares de instâncias de todos os temas considerado.

```

while do
    (Obj
    EstaContido(SS
    (Obj
    EstaContido(SS
if then
else
    if then
        end if
    end if
for to do
    ((Obj
    EstaContido(SS
    ((Obj
    EstaContido(SS
end for
end while

```

A idéia do algoritmo é, dadas instâncias de temas do tipo "polígono", retornar os predicados que denotem que uma instância de objeto "está contida" em outra instância. A otimização na obtenção desses predicados é dada através da característica de transitividade dessa operação.

No contexto deste artigo, "espaço de busca" (SS) é definido como sendo o conjunto de pares de instâncias os quais o predicado pode ser gerado através de uma operação espacial específica (nesse caso "está contido"). A otimização é possível quando reduz-se ao máximo o espaço de busca de uma instância através da análise de outras instâncias diferentes dela mesma. Por exemplo, se um objeto A "não está contido" em B e B "está contido" em C, é certo que A "não está contido" em C mesmo sem efetuarmos a operação e verificarmos o resultado. Essa consideração causa a redução do espaço de busca do objeto A, fazendo desnecessária a operação entre A e C.

O problema da obtenção do predicado "está contido" entre objetos do tipo polígono pode ser formalmente descrito abaixo:

Enunciado 2 *Seja S e T o conjunto de objetos espaciais do tema T e S respectivamente. Seja a uma instância do tema T e b uma instância do tema S . O espaço de busca entre a instância a e todas as instâncias de S é dada por $B(a, S)$. Esse conjunto armazena os pares de instâncias que devem ser verificados para que todos os predicados "está-contido" possam ser obtidos. A partir de $B(a, S)$ as operações entre as instâncias são verificadas. $B(a, S)$ é o conjunto resposta que contém as instâncias de S que estão contidas nas instâncias de a . A ideia é obter o conjunto $B(a, S)$ que contenha todos os pares de instâncias entre todos os objetos poligonais considerados que denotem o predicado "está contido".*

O algoritmo *PoliCPoli* é mostrado em 2.

A função "EstaContido" recebe como entrada o espaço de busca a ser considerado para que a operação espacial seja realizada. A notação $B(a, S)$ imprime que deverá ser considerado no espaço de busca todas as relações transitivas entre $(i+1)$ e j . Isso é realizado como uma operação "em cascata" entre os resultados já calculados, o que pode resultar na descoberta de novos predicados sem a necessidade de calculá-los. Portanto, a redução do espaço de busca para cada conjunto de objetos espaciais faz com que as operações espaciais propriamente ditas possam ser feitas apenas quando necessárias entre duas instâncias cujos resultados não possam inferir qualquer associação transitiva.

Portanto, a verificação da transitividade entre os pares de respostas já obtidos é essencial para que um vasto número de operações espaciais possam ser poupados quando não houver necessidade de cálculo ou já existirem informações suficientes para garantir que um objeto poligonal estará ou não contido em outro.

O resultado desse processo é importante para conseguirmos verificar as operações entre instâncias de conceitos mapeados por linhas e pontos com polígonos.

2. Algoritmo *SpatialC*

A sequência de passos abaixo ilustra o método para a obtenção do super-conjunto de respostas entre objetos espaciais do tipo ponto (), linha () e polígono () para o relacionamento "está-contido".

- (a) Execute *PoliCPoli*. A partir da resposta ordene em ordem decrescente de cardinalidade os elementos do conjunto resposta do algoritmo (). Isso fará com que o espaço de busca dos objetos pontos e linhas para as futuras operações com polígonos estejam definidos sobre esse conjunto de resposta e não sobre as instâncias dos temas poligonais. Ordenar decrescentemente a quantidade de respostas positivas para cada conceito tenta prever transitividade. Quanto mais objetos um tema contém, maiores as chances de haver ocorrências de transitividade entre as outras instâncias com esses objetos.
- (b) Verifique, para cada objeto a , se ele está contido em cada instância dos objetos S e armazene no conjunto de respostas $B(a, S)$. Nesse caso, o espaço de busca para as instâncias de pontos são todas as instâncias de objetos pertencentes a temas do tipo S .
- (c) Verifique, para cada instância do objeto a , se ele está contido em instâncias de polígono da seguinte forma:
 - i. Se a instância de a estiver contida no polígono, verifica transitividade em $B(a, S)$ já calculado por *PoliCPoli* e, se for o caso, gera
 - ii. Se a instância de a não estiver contida no polígono, verifica transitividade em $B(a, S)$ para reduzir o espaço de busca dessa instância de a . Por exemplo, se a não está contido no polígono P_1 , e b está contido no polígono P_2 , então a não pode estar contido em P_2 , portanto, essa verificação não precisa ser realizada.
- (d) Calcule, para cada instância do objeto a , se ele está contido em instâncias de linha L e armazene as ocorrências em $B(a, S)$. Isso envolve atualizar o espaço de busca para cada ponto, de acordo com a análise da transitividade dos objetos que satisfazem e não satisfazem o relacionamento.

Esse algoritmo segue o princípio de análise dos objetos mais específicos (pontos -) para o mais genéricos

espacialmente (polígono -). A idéia é usar os resultados da operação entre polígonos para facilitar a execução dos outros tipos de objetos, tentando prever relacionamentos entre eles sem precisar executar operações espaciais. O conjunto será composto da união entre as triplas contidas em , e .

A funcionalidade da redução de espaços de busca pode ser implementada através de uma lista de objetos candidatos a satisfazerem a operação para cada instância. A "poda" dos objetos não candidatos pode ser feita durante a execução e verificação considerando a hierarquia específica espacial (de pontos até polígonos) e pela propriedade de transitividade do relacionamento "está contido".

3.1.2 Relacionamentos Topológicos Simétricos

Para otimizar a obtenção de relacionamentos topológicos entre os objetos, recorre-se novamente a [5], que define uma árvore de decisão para a obtenção destes a partir da definição e execução de operações na fronteira e interiores de objetos espaciais do tipo polígono, linha ou ponto. No caso deste artigo, uma nova árvore de decisão foi criada baseada na intenção de se utilizar resultados da operação "está-contido" para tentar minimizar o número de cálculos na obtenção dos outros relacionamentos topológicos.

A partir da verificação das propriedades de transitividade e simetria dos relacionamentos topológicos e da utilização dos resultados de *SpatialC* propomos o algoritmo *AlgTop* que verifica, dado o par de objetos espaciais e , se existem um dos quatro relacionamentos topológicos ("está-disjunto", "toca", "cruza" e "sobrepõe") entre suas instâncias.

Enunciado 3 Seja e o conjunto de objetos espaciais do tema e respectivamente. Seja a instância do tema e a instância do tema . Considere o interior do objeto espacial referente à instância do tema e a função a dimensão da instância considerada¹. Considere o conjunto das quatro operações topológicas consideradas: "toca", "está-disjunto", "cruza" e "sobrepõe". A partir do conjunto composto das triplas () previamente definido pelo usuário, onde é um subconjunto total ou parcial de , determinar o conjunto de triplas () correspondente aos predicados espaciais caso exista entre as duas instâncias o relacionamento topológico , onde .

O algoritmo *AlgTop* toma como uma das entradas o conjunto de respostas da operação "está-contido". Os objetos espaciais a serem analisados serão os pares que não es-

¹ As dimensões consideradas nos objetos espaciais são zero para objetos do tipo ponto, 1 para objetos do tipo linha e 2 para polígonos

tejam presentes neste conjunto de respostas. Com essas instâncias, é aplicada a operação de interseção e o objeto resultante da operação é armazenado em . A partir da análise de , consegue-se prever o relacionamento espacial entre e . Isso é feito através da aplicação dos testes de interseção de interiores dos objetos espaciais e análise da dimensão dos resultados, ilustrados no algoritmo.

O conjunto resposta conterá todas as instâncias dos objetos considerados que possuam predicados espaciais válidos com outros objetos de temas diferentes. Os relacionamentos topológicos entre os objetos referentes aos temas escolhidos pelos usuários estão contidos totalmente neste conjunto. Portanto, temos um conjunto cuja semântica equivale aos relacionamentos topológicos entre os objetos espaciais considerados.

Algorithm 3 *AlgTop*

Entrada: O conjunto de resultados de "*SpatialC*" e o conjunto de triplas ()

Saída: O conjunto-resposta composto da união de e do conjunto de triplas ()

```

for all triplas e pertencentes a do
  for all par de instâncias e
    do
      if then
        if then
          , "está-disjunto",
        end if
        if then
          if then
            , "toca",
          end if
          if then
            , "cruza",
          end if
          else
            , "sobrepõe",
          end if
        end if
      end if
    end for
  end for

```

3.2 Operações de Distância

O segundo conjunto de operações a serem contempladas neste modelo refere-se às operações baseadas em distância entre os objetos espaciais. Os predicados gerados a partir dos relacionamentos de distância poderão descrever distâncias pontuais entre objetos ou considerações semânticas a

partir do valor da distância sobre hierarquias de conceitos.

3.2.1 Hierarquias "D" Padrão

Hierarquias "D" Padrão são aquelas que são definidas entre todos os temas de dimensões diferentes, ou seja, existe uma hierarquia para cada par de combinação das três dimensões de objetos espaciais (ponto, linha ou polígono). Por exemplo, uma hierarquia "D" será criada entre objetos do tipo ponto: , entre pontos e linhas: ; e assim por diante. Isso significa dizer que podem ser criadas, a priori, 6 hierarquias distintas de distância considerando apenas as dimensões dos temas considerados. A criação dessas hierarquias é o primeiro passo para a generalização dos relacionamentos de distância entre os conceitos espaciais do conjunto .

Entretanto, a criação dessas hierarquias desconsidera uma característica importante presente em vários relacionamentos espaciais: a simetria. O questionamento é: caso haja uma hierarquia "D" padrão entre ponto e linha (), essa mesma pode ser considerada quando a análise for simétrica, ou seja, quando o objetivo for analisar a distância entre linha e ponto? Essa resposta é facilmente respondida quando a análise do contexto é dispensada; ou seja, se um ponto está a uma distância de uma linha, é óbvio dizer que a mesma linha está a uma distância do mesmo ponto. A análise pontual de faz com que tenha-se a impressão de que a "direção" da análise (linha para ponto ou ponto para linha) não altere o significado. Como os predicados serão gerados a partir das generalizações dos valores pontuais, a ordem da análise tem sua importância, podendo alterar os predicados a serem obtidos.

Tome como exemplo o seguinte cenário: a análise entre pontos de distribuição de linhas telefônicas () e linhas de transmissão (). No contexto considerado, um ponto de distribuição está "próximo" a uma linha de transmissão se ele estiver a uma distância inferior a . Em contrapartida, no mesmo contexto, uma linha de transmissão estará "próxima" de um ponto de distribuição se ela estiver a uma distância inferior a . Portanto, dois predicados idênticos "próximo" entre dois objetos diferentes podem ser obtidos a partir de distâncias pontuais diferentes (). Esse fato sugere que sejam inseridas novas hierarquias no conjunto das hierarquias "D" padrão, considerando agora a "ordem" na qual serão analisados os objetos espaciais para que sejam derivados. A partir dessa inserção, 12 hierarquias "D" padrões são geradas considerando a dimensão dos temas considerados e a direção da análise na derivação dos predicados. A figura 2 ilustra exemplos de uma hierarquia "D"

Existe a possibilidade de que não sejam necessárias novas hierarquias dependentes da ordem a qual os conceitos são considerados na derivação. No entanto, esse fato também

é diretamente ligado ao contexto da aplicação e, no caso da descrição deste modelo, opta-se por cobrir todas as possibilidades independente da simplicidade que algumas aplicações apresentam em detrimento da complexidade do modelo e do número de estruturas a serem criadas.

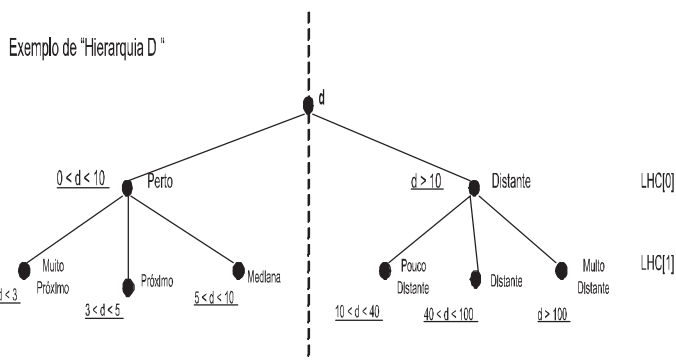


Figura 2: Hierarquia de Conceito "D"

3.2.2 Hierarquias "D" Específicas

Além de considerar a dimensão dos temas e a ordem a qual os predicados de distância serão derivados, o modelo deve possibilitar que hierarquias específicas entre dois pares de temas possam ser criadas e consideradas. Essas hierarquias denominam-se "*Hierarquias D Específicas*", e são criadas não a partir do conjunto dos temas agrupados por dimensão, e sim por temas específicos escolhidos pelo usuário especialista.

Por exemplo, considere dois temas do tipo ponto: lotes e postes. Existem hierarquias "D" padrões para a análise , tanto na direção lote/postes quanto na direção contrária. Entretanto, o especialista pode querer construir uma hierarquia específica que considere a ordem postes/lotos sobre uma semântica diferente das hierarquias padrões. Isso deve ser possível, haja visto que podem existir um grande número de temas pertencentes à mesma dimensão e cuja semântica difere dos outros que fazem parte do mesmo conjunto.

As hierarquias "D" específicas têm prioridade sob quaisquer hierarquias "D" padrões existentes, pelo fato de terem sido criada pelo usuário especialista no contexto considerado, e ser mais específica do que as hierarquias padrões.

3.2.3 Algoritmo AlgDist

Baseado na definição e modelagem das hierarquias de distância entre dois objetos, o algoritmo de derivação de relacionamentos de distância *AlgDist* pode ser criado. O problema é descrito formalmente através do enunciado abaixo:

Enunciado 4 Seja S e T o conjunto de objetos espaciais do tema s e t respectivamente. A partir disto, seja n e m o número de instâncias de objetos dos temas s e t . Seja d a distância entre a instância i do tema s e a instância j do tema t . Seja L uma lista que contém as distâncias entre duas instâncias de quaisquer objetos espaciais considerados ($L = \{d_{ij} \mid i \in S, j \in T\}$, para todos os temas s e instâncias consideradas). Seja H a hierarquia de distância escolhida pelo usuário entre os temas s e t nesta ordem e N o nível da hierarquia H a ser considerado na derivação. A partir do conjunto C composto das quádruplas ((s, t, d, N)), deseja-se obter o conjunto R de triplas da forma ((s, t, d)), onde d refere-se ao conteúdo semântico das hierarquias de distância pré-definidas pelo usuário.

Baseado nesta contextualização formal, o algoritmo *AlgDist* é ilustrado abaixo:

Algorithm 4 AlgDist

Entrada: O conjunto

de quádruplas contendo os temas s a serem derivados (na ordem), a hierarquia H e o nível da hierarquia N a ser considerada na mineração.

Saída: O conjunto R de triplas (s, t, d) , onde d é um predicado espacial de distância entre as instâncias consideradas.

```

for all quádrupla pertencente em  $C$  do
  for  $s$  to  $S$  do
    for  $t$  to  $T$  do
      if  $d_{st} \geq d_{th}$  then
         $Obj = (s, t, d_{st})$ 
        Insira  $Obj$  em  $R$ 
      else
         $d = d_{st}$ 
      end if
    end for
  end for
end for
Retornar  $R$ 

```

A função *"Calculadistância"* retorna a distância pontual (na unidade considerada pela aplicação) entre duas instâncias de objetos espaciais.

A função *"Insira"* tem como objetivo inserir o valor da distância já calculada entre duas instâncias na lista R . O principal objetivo dessa lista é tentar reduzir o número de cálculo de distâncias entre dois objetos. Por exemplo, se uma quádrupla exige que sejam calculadas distâncias entre s e t , o algoritmo já terá calculado as distâncias de todas as instâncias de objetos entre esses dois te-

mas. Entretanto, é possível que haja uma quádrupla que exigem que sejam derivados os relacionamentos de distâncias entre s e t . Como citado anteriormente, a ordem a qual os temas são considerados neste caso é relevante para a derivação semântica, por possivelmente possuir hierarquias diferentes dependendo da ordem. No entanto, a distância entre os dois objetos é a mesma, variando apenas a semântica de cada derivação, o que torna desnecessário o novo cálculo da distância pontual entre os dois objetos. A função *"Extraia"* retorna o valor da distância entre duas instâncias de objetos armazenados em R .

"PHierarquiaD" implementa o processo de derivação em si. Baseado na distância, na hierarquia e no nível da hierarquia, ela percorre a árvore e encontra o nó o qual a distância pontual calculada se insere. Essa semântica do nó é repassada para a variável d que se tornará o predicado espacial deste tipo de relacionamento entre essas instâncias. Exemplificando, considere a figura 2 que contém a hierarquia H de distância entre dois objetos espaciais. Supondo que a distância pontual entre duas instâncias seja 4, e que o nível da hierarquia considerado seja 1, então a derivação semântica se dará percorrendo todos os nós da hierarquia no nível 1 até que o valor pontual 4 seja encontrado em algum nó. Nesse caso, o predicado a ser gerado será o *"Próximo"*, que denota todos os relacionamentos de distância entre objetos espaciais entre 3 e 5. Portanto, a função percorre a hierarquia de forma que as distâncias sejam analisadas até que os valores pontuais sejam transformados em predicados espaciais de distância semanticamente equivalentes.

4 Conclusões e Extensões

Esse artigo propôs um modelo de derivação de relacionamentos entre objetos espaciais para um conjunto de predicados semanticamente equivalente no formato relacional. O objetivo é a obtenção de um repositório que contenha apenas os relacionamentos entre os objetos espaciais, ambos escolhidos pelo usuário especialista no contexto considerado e verificados pelo modelo.

A importância do usuário na execução do modelo é vital. Uma das razões é que a geração de predicados não deve ser indiscriminada porque a obtenção dos predicados entre os dados em um conjunto de dados espacial implicaria no cálculo de todos os relacionamentos possíveis entre as instâncias dos objetos, o que é inviável em termos de custos computacionais. Por essa razão, o usuário restringe o escopo escolhendo que objetos espaciais serão derivados e quais os relacionamentos entre eles serão considerados.

O desenvolvimento do modelo em questão está em andamento, e sua validação será realizada utilizando uma aplicação de mineração de dados espacial sobre o contexto específico da indústria de telecomunicações. Regras de as-

sociação Espacial serão obtidas utilizando algoritmos de extração convencionais (Apriori, FP-Growth, ECLAT [1, 13, 12]) sobre o repositório gerado pelo modelo proposto.

Extensões deste trabalho podem vislumbrar a geração de predicados baseados em relacionamentos direcionais entre os objetos espaciais (norte, sul, leste, etc.), definição de uma interface para o usuário e validação do trabalho em contextos diferentes. Outras aplicações que propõem lidar apenas com o equivalente semântico de um conjunto de objetos espaciais podem envolver estudo em otimização de consultas espaciais específicas em determinados contextos [21], interoperabilidade entre sistemas de informação geográficas [10, 21], controle de acesso a objetos espaciais, dentre outros.

5 Agradecimentos

Esse trabalho é financiado pelo CNPq e tem apoio da Fundação CPqD através do fornecimento de dados e infra-estrutura para validação e testes.

Referências

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, pages 487–499. Morgan Kaufmann, 12–15 1994.
- [2] Brandon Bennett. Spatial reasoning with propositional logics. In Jon Doyle, Erik Sandewall, and Pietro Torasso, editors, *KR'94: Principles of Knowledge Representation and Reasoning*, pages 51–62. Morgan Kaufmann, San Francisco, California, 1994.
- [3] K. Bennis, B. David, I. Morize-Quilio, J. Thevenin, and Y. Viemont. Geograph: A topological storage model for extensible gis. *Auto-Carto 10*, pages 349–367, 1991.
- [4] Gilberto Camara, Antonio Miguel Monteiro, Joao Paiva, Jonas Gomes, and Luiz Velho. Towards a unified framework for spatial data models.
- [5] Eliseo Clementini, Paolino Di Felice, and Peter van Oosterom. A small set of formal topological relationships suitable from end-user interaction. *Lecture Notes in Computer Science*, 692:277–195, 1993.
- [6] M. Egenhofer and R. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, pages 161–174, 1991.
- [7] M. Egenhofer and J. Herring. Categorizing binary topological relationships between regions, lines, and points in geographic databases, 1990.
- [8] Max J. Egenhofer and Robert D. Franzosa. On the equivalence of topological relations. *International Journal of Geographical Information Systems*, 9(2):133–152, 1995.
- [9] Donato Malerba Floriana. Mining spatial association rules in census data, 2001.
- [10] Andrew U. Frank. *Ontology: a consumer's point of view*, 2000.
- [11] V. Haarslev, C. Lutz, and R. Möller. A description logic with concrete domains and role-forming predicates. *Journal of Logic and Computation*, 9(3):351–384, 1999.
- [12] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2001.
- [13] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. pages 1–12, 2000.
- [14] Krzysztof Koperski and Jiawei Han. Discovery of spatial association rules in geographic information databases. In M. J. Egenhofer and J. R. Herring, editors, *Proc. 4th Int. Symp. Advances in Spatial Databases, SSD*, volume 951, pages 47–66. Springer-Verlag, 6–9 1995.
- [15] B. Kuijpers, J. Paredaens, and J. Van den Bussche. Lossless representation of topological spatial data. In M. J. Egenhofer and J. R. Herring, editors, *Proceedings of the 4th International Symposium on Large Spatial Databases (SSD)*, volume 951, pages 1–13, Berlin, 1995. Springer-Verlag.
- [16] Bart Kuijpers, Jan Paredaens, and Jan Van den Bussche. On topological elementary equivalence of spatial databases. In *ICDT*, pages 432–446, 1997.
- [17] Yijun Lu. Concept hierarchy in data mining: Specification, generation and implementation.
- [18] D. Mark and M. Egenhofer. An evaluation of the 9-intersection for region-line relations. *GIS/LIS*, 1992.
- [19] D. Mark and M. Egenhofer. Modeling spatial relations between lines and regions: Combining formal mathematical models and human subjects testing, 1994.
- [20] David M. Mark and Max J. Egenhofer. Calibrating the meanings of spatial predicates from natural language: Line-region relations. In *Sixth International Symposium on Spatial Data Handling*, volume 1, pages 538–553, Edinburgh, Scotland, 1994.
- [21] C. H. Papadimitriou, D. Suciu, and V. Vianu. Topological queries in spatial databases. pages 81–92, 1996.