

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
TRABALHO DE CONCLUSÃO DE CURSO– TCC

Thiago Deschamps Reis Felix

**Implementação de uma Arquitetura
para Minerar Regras de Associação
Espaciais a partir de Dados Espaciais
*Fuzzy***

Thiago Deschamps Reis Felix

**Implementação de uma Arquitetura
para Minerar Regras de Associação
Espaciais a partir de Dados Espaciais
*Fuzzy***

Monografia apresentada ao Trabalho de Conclusão de Curso do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Área de concentração: Banco de Dados

Orientador: Prof. Dr. Anderson Chaves Carniel

São Carlos

2022

UNIVERSIDADE FEDERAL DE SÃO CARLOS

THIAGO DESCHAMPS REIS FELIX

Esta Monografia foi julgada adequada para a obtenção do título de Bacharel em Engenharia de Computação, sendo aprovada em sua forma final pela banca examinadora:

Orientador: Prof. Dr. Anderson Chaves
Carniel
Universidade Federal de São Carlos
UFSCar

Prof. Dr. Renato Bueno
Universidade Federal de São Carlos
UFSCar

Profa. Dra. Marcela Xavier Ribeiro
Universidade Federal de São Carlos
UFSCar

São Carlos, Setembro de 2022

Dedico este trabalho à todas as pessoas que encontrei nessa jornada, que me ajudaram, apoiaram e foram crucial no meu desenvolvimento e no ponto o qual o cheguei. Em especialmente aos meus pais, Maxwell e Ana Paula, por sempre me apoiar, minha namorada Laura, que nunca soltou a minha mão mesmo nos momentos mais difíceis e à minha família, que rezou tanto para que eu pudesse estar aqui hoje.

Agradecimentos

Agradeço primeiramente a Deus, por me dar o presente da vida, com bastante saúde e habilidade de pensar e raciocinar para que eu possa ter concluído este trabalho.

Aos meus pais, Maxwell Bezerra Felix da Silva e Ana Paula Deschamps Reis Felix, os quais estiveram comigo desde o momento que eu vim para este mundo, dando sempre muito amor e carinho desde o dia que nasci, me incentivando a correr atrás dos meus sonhos e nunca desistir.

A minha companheira de vida, minha parceira para todos os momentos, meu amor, Laura Cristina Rios Ferreira, que sempre me acalma nos momentos difíceis e me faz rir nos momentos tristes, sempre esteve comigo independente da situação, me apoiando, incentivando e dizendo que tudo vai dar certo. Rezando dia e noite para que este trabalho que eu escrevo seja concluído.

A minha avó Edacy Deschamps do Nascimento Reis, que também esteve cuidando de mim desde o momento do meu nascimento, e hoje, ainda longe sempre se preocupa e reza por mim.

Aos amigos que fiz durante a graduação, que sempre me ajudaram em matérias acadêmicas ou pessoais, sempre estavam lá quando eu precisava.

Aos meus professores da Universidade Federal de São Carlos, em especial ao meu orientador, Professor Dr. Anderson Chaves Acarniel, que me deu a oportunidade de participar deste projeto incrível que uma parte acabou virando meu TCC, sempre dando todo suporte necessário, e a também ao Pedro Vinícius Almeida Borges de Venâncio e Henrique Pigozzo da Silva, porque sem eles, este trabalho não seria possível.

“ Você vai ficar bem. Você acabou de tropeçar em uma pedra na estrada. Não significa nada. Seu objetivo está muito além disso, não é? Tenho certeza que você vai superar isso. Você vai andar de novo. ...em breve.”

(Miura, Kentaro)

Resumo

O uso e a importância de dados espaciais vem crescendo cada vez mais. Com a assimilação do mundo real, muitos fenômenos espaciais são modelados e representados por objetos espaciais *fuzzy*, os quais possuem interiores vagos, limites incertos e/ou localizações inexatas. Inferir regras e encontrar padrões é algo que o ser humano sempre fez, com o avanço da inteligência artificial e da representação de um objeto espacial, este trabalho propõe uma implementação de uma arquitetura capaz de minerar regras de associação espacial a partir de dados espaciais *fuzzy*. Com entradas do usuário como parâmetro, a arquitetura proposta descobre padrões de associação espacial entre objetos espaciais de interesse dos usuários (por exemplo, atrações turísticas) e tais características espaciais *fuzzy* (por exemplo, condições sanitárias de restaurantes, número de avaliações e preço das acomodações). Além disso, este trabalho apresenta um estudo de caso baseado em conjuntos de dados reais que mostra a aplicabilidade do método desenvolvido.

Palavras-chave: Ciência de Dados Espaciais, Regras de Associações Espaciais, Vagueza Espacial, Dados Espaciais Fuzzy.

Abstract

The use and importance of spatial data have been growing more and more. With the assimilation of the real world, many spatial phenomena are modeled and represented by fuzzy spatial objects, which have blurred interiors, uncertain boundaries and/or inexact locations. Inferring rules and finding patterns is something that human beings have always done, with the advancement of Artificial Intelligence and the representation of a spatial object, this work proposes an implementation of an architecture capable of mining spatial association rules from fuzzy spatial data. With user input as a parameter, the proposed architecture discovers patterns of spatial association between spatial objects of interest to users (e.g. tourist attractions) and such fuzzy spatial features (e.g. sanitary conditions of restaurants, number of reviews and price of accommodations). Furthermore, this work presents a case study based on real datasets that shows the applicability of the developed method.

Keywords: Spatial Data Science, Spatial Association Rules, Spatial Fuzziness, *Fuzzy* Spatial Data.

Lista de ilustrações

Figura 1 – Exemplos de objetos espaciais: (a) um ponto simples, (b) uma linha simples, (c) uma região simples, (d) um ponto complexo, (e) uma linha complexa, (f) uma região complexa.	26
Figura 2 – Função de Pertinência.	27
Figura 3 – Demonstração da Função de Pertinência a partir de objetos de região <i>plateau</i> para cada valor linguístico da notabilidade de acomodação da característica espacial fuzzy, incluindo os limites da cidade de Nova York.	29
Figura 4 – Alguns exemplos ilustrados de relacionamentos topológicos	31
Figura 5 – Amostra de dados extraídos do <i>DOHMH</i>	38
Figura 6 – Amostra de dados extraídos do <i>Airbnb</i>	39
Figura 7 – Amostra de dados extraídos do <i>OpenStreetMaps</i>	39
Figura 8 – Esquematização da Arquitetura.	41
Figura 9 – Hierarquia de fenômenos espaciais do exemplo executado na arquitetura.	42
Figura 10 – Conjunto de dados usados para a construção de objetos espaciais <i>fuzzy</i>	43
Figura 11 – Amostra de código da arquitetura escrito em R que representa uma parte da construção de objetos espaciais <i>fuzzy</i>	45
Figura 12 – Hierarquia de fenômenos espaciais do exemplo executado na arquitetura.	45
Figura 13 – Exemplos de regras de Categoria extraídas pela arquitetura.	46
Figura 14 – Exemplos de regras de Tipo extraídas pela arquitetura.	47
Figura 15 – Amostra de código da arquitetura escrito em R que representa uma parte da mineração de regras de associação espaciais.	47

Lista de tabelas

Tabela 1 – Comparação de abordagens existentes com a implementação do método realizado neste TCC (última coluna). 35

Sumário

1	INTRODUÇÃO	21
1.1	Contextualização e Motivação	21
1.2	Justificativa	22
1.3	Objetivos	23
1.4	Contribuições	23
1.5	Organização	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Banco de dados espaciais	25
2.1.1	Definição de banco de dados espaciais	25
2.1.2	Tipos de dados espaciais	26
2.2	Sistemas <i>fuzzy</i>	27
2.2.1	Valores Linguísticos	27
2.2.2	Objetos Espaciais <i>fuzzy</i>	28
2.3	Manipulação de dados espaciais <i>fuzzy</i>	29
2.4	Mineração de regras de associação	29
2.5	Mineração de regras de associação espaciais	30
2.6	Considerações finais	31
3	TRABALHOS RELACIONADOS	33
3.1	Trabalhos que manipulam dados espaciais <i>crisp</i>	33
3.2	Trabalhos que manipulam dados espaciais <i>fuzzy</i>	34
3.3	Considerações Finais	35
4	IMPLEMENTAÇÃO DA ARQUITETURA PARA A DESCOBERTA DE REGRAS DE ASSOCIAÇÃO ESPACIAL A PARTIR DE OBJETOS ESPACIAIS <i>FUZZY</i>	37

4.1	Exemplo de execução	37
4.2	Visão geral da arquitetura	40
4.3	Parâmetros do Usuário	41
4.4	Camada de Dados Espaciais	43
4.5	Camada de Tratamento de Dados Espaciais	44
4.6	Camada de Tratamento de Conjuntos de Itens	45
4.7	Camada de Recuperação de Regras de Associação Espacial	47
5	CONCLUSÃO	49
5.1	Conclusões	49
5.2	Trabalhos Futuros	49
	Referências	51
APÊNDICE A	DISCOVERY OF SPATIAL ASSOCIATION RULES FROM FUZZY SPATIAL DATA	53

Capítulo 1

Introdução

Este capítulo é dividido nas seguintes seções. A seção 1.1 apresenta o contexto e a motivação deste trabalho. As seções 1.2, 1.3, 1.4 e 1.5 apresentam, respectivamente, a justificativa, os objetivos, contribuições e a forma como este trabalho está organizado.

1.1 Contextualização e Motivação

Com o avanço da tecnologia, meios mais sofisticados e precisos de análises são necessários para explorar as características geométricas e topológicas especiais de fenômenos espaciais, como localização e relações espaciais. A Ciência de Dados Espacial surge como uma importante área que fornece tais métodos [3]. A suposição comum é que os fenômenos espaciais são representados por instâncias de tipos de dados baseados em vetores chamados *tipos de dados espaciais* [16], como *pontos*, *linhas* e *regiões* (detalhados na Seção 2.1.2). Os quais podem ser abstraídos como localizações, formas geométricas e limites estritamente definidos no espaço, e são definidos como objetos espaciais *crisp*. Isso também significa que as informações extraídas desses objetos são exatas. Por exemplo, relacionamentos topológicos (por exemplo, sobreposição) em objetos espaciais nítidos produzem resultados exatos.

No entanto, quando se trata do mundo real, nem todo fenômeno é exato, por exemplo, a temperatura de uma cidade não pode ser descrita estritamente com uma palavra ou número, ou por exemplo uma previsão de precipitação de um estudo meteorológico, já que todo experimento empírico carrega uma incerteza por definição. Para resolver o problema de representar tais fenômenos, alguns objetos espaciais são caracterizados por *spatial fuzziness* [3]. Objetos espaciais com deste tipo têm interiores 'borrados', limites

incertos, localizações inexatas e/ou múltiplas características. Tais objetos não podem ser adequadamente representados por objetos espaciais *crisp*.

Eles são adequadamente representados por instâncias de *tipos de dados espaciais fuzzy*, como *pontos fuzzy*, *linhas fuzzy* e *regiões fuzzy*. A teoria dos conjuntos fuzzy [18] é usada para modelar tais *objetos espaciais fuzzy*. A ideia principal da teoria de conjuntos *fuzzy* é definir um *grau de adesão* entre 0 e 1 para cada ponto de um objeto espacial *fuzzy*. Este grau indica até quanto um ponto pertence ao objeto. Um exemplo deste trabalho é, um objeto de região *fuzzy* que representa a área de cobertura de *cheap restaurant*. Neste caso, a característica de um local é representada pelo valor linguístico *cheap* denotado pela variável linguística *restaurant*. Para este objeto, os pontos com valores bem próximos ao 1 representam muito certamente locais *cheap* (baratos). Pontos próximos ao valor 0 denotam locais que são mais caros. Portanto, os outros valores acabam flutuando entre $]0, 1[$ acabam ficando entre mais caros ou mais baratos (ou seja, diferentes níveis de valores linguísticos). As operações espaciais que lidam com objetos espaciais *fuzzy* também são *fuzzy*, pois precisam lidar com os graus de pertinência dos objetos. Por exemplo, uma *relação topológica fuzzy* (por exemplo, sobreposição *fuzzy*) [2] em objetos espaciais *fuzzy* produz um valor em $[0, 1]$ que indica o grau de verdade da relação.

1.2 Justificativa

As regras de associação espacial [12] têm um grande papel na ciência de dados espacial. Para isso, a extração de padrões espaciais úteis e significativos a partir de informações geométricas é uma tarefa essencial para esta área. A possibilidade de aplicações expressarem a frequência dois ou mais conjuntos de dados espaciais estão relacionados usando regras *se-então*, que possuem valores associados para medir sua força e significância é uma tarefa muito relevante nos dias atuais. Normalmente, o interesse está em minerar essas regras com um alto nível de confiança, que são denominadas como regras *fortes*. Por exemplo, é possível explorar regras fortes de associação espacial que mostram como as localizações das atrações turísticas estão relacionadas às áreas de cobertura de acomodações e restaurantes.

As abordagens anteriores disponíveis que extraem regras de associação espacial [6, 7, 9, 11, 12, 13, 14, 15] enfrentam pelo menos um dos seguintes problemas. Primeiro, seu foco está em lidar apenas com objetos espaciais *crisp*. Em segundo lugar, eles são incapazes de processar relações topológicas *fuzzy*. Esses problemas limitam seriamente a representação da imprecisão espacial. O último problema refere-se à falta de uma estratégia para lidar com variáveis linguísticas e valores que representem os níveis de imprecisão dos objetos.

1.3 Objetivos

Este Trabalho de Conclusão de Curso (TCC) é uma implementação de uma arquitetura desenvolvida conceitualmente dentro do contexto do artigo [17]. A cópia deste artigo pode ser visualizada no apêndice A. Este artigo foi aceito para publicação na *ER 2022 : 41st International Conference on Conceptual Modeling*. Dessa forma, seus objetivos específicos são listados abaixo:

- fornecer uma implementação de que consiga manipular os dados espaciais *fuzzy* para o processo de extração de regras.
- Utilizar o pacote *fsr* [5] para manipulação de dados espaciais *fuzzy*, principalmente o que se refere ao processamento de relacionamento topológicos *fuzzy*.
- Aplicar o método proposto em um caso de estudo real.
- Implementar a arquitetura como um fluxo contínuo de dados.

Para implementar o estudo de caso deste TCC, utilizou-se dados do *OpenStreetMap* para extrair objetos de regiões *crisp* dentro da cidade de Nova York que representam atrações turísticas, como cultura, lazer e áreas históricas. Como resultado, esse conjunto de dados contém 12.328 objetos de região *crisp* rotulados com a categoria (por exemplo, *historical*) e o tipo (por exemplo, *city*) da atração turística. E para que essa arquitetura seja como um fluxo contínuo de dados, será implementada sequência lógica de etapas que começa desde os tratamento dos dados mais brutos até as regras geradas. Portanto, resolver os problemas mencionados e implementar uma arquitetura para a extração regras de associação espacial a partir de dados espaciais *fuzzy* é o papel deste trabalho.

1.4 Contribuições

As principais contribuições deste trabalho são:

- Implementação de um método que resolve o problema de descobrir a associação espacial de fenômenos espaciais caracterizados por imprecisões espaciais.
- Mostra a aplicabilidade do método proposto por meio de um estudo de caso com base em conjuntos de dados espaciais reais.
- Utilização do pacote *fsr* [5] para a extração de regras de associação espaciais *fuzzy*.
- Construir uma arquitetura para tratar os dados espaciais e minerar as regras de forma otimizada.

1.5 Organização

Este trabalho possui 5 capítulos, iniciando com o capítulo atual de Introdução e outros 4 capítulos, organizados da seguinte forma:

- Capítulo 2 - Neste capítulo de fundamentação teórica, são apresentados os conceitos necessários para compreensão dos assuntos abordados ao longo dos demais capítulos, incluindo: (i) Banco de Dados Espaciais; (ii) Sistemas *fuzzy*; (iii) Tratamento de dados espaciais *fuzzy*; (iv) Mineração de regras de associação; e (v) Mineração de regras de associação espaciais.
- Capítulo 3 - É realizada uma revisão bibliográfica de trabalhos relacionados, nos quais abordam mineração de regras de associação espacial. Neste sentido, é apresentado um resumo acerca de cada um destes trabalhos, seus pontos positivos e negativos, além de apresentar o diferencial deste TCC em relação aos trabalhos apresentados e na conclusão uma tabela como comparativo de todos os trabalhos.
- Capítulo 4 - A arquitetura para descoberta de regras de associação espacial de objetos espaciais fuzzy é apresentada, descrevendo os algoritmos desenvolvidos para as etapas de (i) Parâmetros do Usuário; (ii) Camada de Dados Espaciais; (iii) Camada de tratamento de dados espaciais; (iv) Camada de Tratamento de Conjuntos de Itens; e (v) Camada de Recuperação de Regras de Associação Espacial. Como resultado, este capítulo demonstra como este TCC atinge aos objetivos 1, 2 e 3.
- Capítulo 5 - Apresenta as conclusões e trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste capítulo são apresentados termos técnicos e conceitos utilizados ao longo deste trabalho. Na seção 2.1 descreve os tipos de dados espaciais. A seção 2.2 aborda o conceito de sistemas *fuzzy*. Na seção 2.3 aborda conceitos de tratamento de dados espaciais *fuzzy*. Nas seções 2.4 e 2.5 são abordados mineração de regras de associação e mineração de regras de associação espaciais, respectivamente. Por fim, a seção 2.6 contém as considerações finais sobre este capítulo.

2.1 Banco de dados espaciais

Um dado espacial é qualquer tipo de dado que referencia direta ou indiretamente uma área ou local específico de forma alfanumérica, como por exemplo, um espaço como o mundo físico (geografia ou astronomia), partes de organismos vivos (anatomia do corpo humano), projetos de engenharia (circuitos integrados de grande escala, o projeto de um automóvel ou uma estrutura molecular de um projeto de um medicamento farmacêutico) e um espaço de informação conceitual (fluxo de fluido ou um campo eletromagnético).

2.1.1 Definição de banco de dados espaciais

Com fins de consulta, armazenamento e outras aplicações, foi desenvolvido o sistema de Banco de Dados Espaciais, que, segundo Güting [10], um banco de dados espaciais é um banco de dados, pois está associado a uma visão de um banco de dados como contendo conjuntos de objetos no espaço ao invés de imagens. Portanto, um banco de dados espacial ainda está conectado com tipos de dados não-espaciais, devido ainda ser preciso ter uma

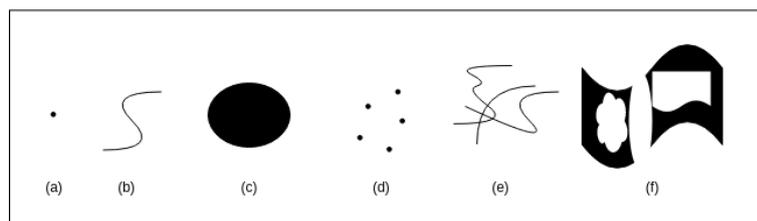
representação alfanumérica, oferecendo suporte aos tipos de dados espaciais, sua própria linguagem de consulta e modelação com algoritmos para as operações otimizadas.

2.1.2 Tipos de dados espaciais

Tipos de dados espaciais fornecem uma abstração fundamental para modelar e representar fenômenos reais em objetos espaciais, bem como seus relacionamentos, propriedades e operações. Exemplos de tipos de dados fundamentais são: pontos, os quais podem ser definidos como um objeto no espaço, sem extensão ou área (por exemplo, representando a localização de uma igreja, castelo ou cidade), linhas, as quais podem ser definidas como pontos conectados, com extensão mas sem área (por exemplo, representando uma estrada, rio, cabo ou rota) e regiões, que pode ser definida como uma área com extensão (por exemplo, cidade, lago ou floresta). A partir deles, é possível formar tipos de dados espaciais mais complexos, como pontos, linhas e regiões complexas (consiste de vários pontos, linhas ou regiões, que podem estar disjuntas ou adjacentes) e tipos de dados tridimensionais, usado para modelar paisagens ou volume, por exemplo.

No plano bidimensional, os tipos de dados espaciais simples fornecem apenas estruturas de objetos simples, como pontos únicos, linhas contínuas e regiões simples (Figura 1(a)-(c)). No entanto, de uma perspectiva de aplicação, os tipos de dados espaciais simples revelam-se abstrações inadequadas para aplicações espaciais, uma vez que são insuficientes para lidar com a variedade e complexidade da realidade geográfica. De uma perspectiva formal, elas não são fechadas sob as operações do conjunto geométrico interseção, união e diferença. Isso significa que essas operações aplicadas a dois objetos espaciais simples podem produzir um objeto espacial que não é simples. Tipos de dados espaciais complexos resolvem esses problemas. Eles fornecem objetos espaciais universais e versáteis e são fechados sob operações de conjunto geométrico. Eles permitem objetos com múltiplos componentes, que podem ter furos e componentes de linha que podem modelar redes geométricas ramificadas e conectadas (Figura 1(d)-(f)).

Figura 1 – Exemplos de objetos espaciais: (a) um ponto simples, (b) uma linha simples, (c) uma região simples, (d) um ponto complexo, (e) uma linha complexa, (f) uma região complexa.



Fonte: O autor

2.2 Sistemas *fuzzy*

Com a necessidade de lidar e fazer processamentos baseados em informações imprecisas ou aproximadas, utiliza-se a teoria dos conjuntos *fuzzy*, que é uma extensão e generalização da teoria dos conjuntos booleana.

Dado um subconjunto A em um universo X, na teoria clássica dos conjuntos, os elementos deste universo estritamente pertencem ou não pertencem àquele subconjunto, ou seja, X só pode assumir o valor 0 ou 1, os quais são chamados de crisp. O que pode ser expresso pela seguinte equação:

$$f(x_A) = X \rightarrow \{0, 1\} \quad (1)$$

Na Teoria de Conjuntos *fuzzy*, X pode assumir infinitos valores de 0 e 1, que é definido por uma função de pertinência:

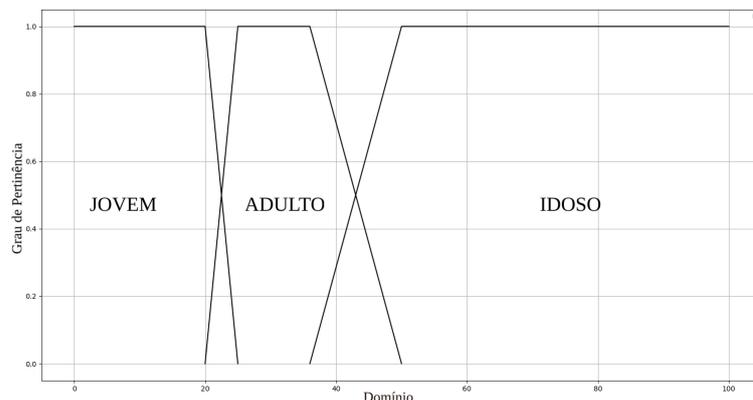
$$\mu_A(x) = X \rightarrow [0, 1] \quad (2)$$

Valores maiores ou menores para x, significam graus mais altos ou mais baixos de compatibilidade ao conjunto, respectivamente. Assim, a teoria dos conjuntos *fuzzy* permite que um determinado elemento possa pertencer a mais de um conjunto *fuzzy*, com vários e diferentes graus de pertinência em diferentes conjuntos *fuzzy*.

2.2.1 Valores Linguísticos

Uma variável linguística (VarL) representa a imprecisão de um conceito utilizando valores linguísticos (ValL), cujos valores são nomes de conjuntos *fuzzy*. Por exemplo, a idade pode ser caracterizada por uma ValL assumindo valores como jovem, adulto ou idoso. Estes valores são descritos por intermédio de conjuntos *fuzzy*, representados por funções de pertinência, conforme mostrado na Figura a seguir:

Figura 2 – Função de Pertinência.



Fonte: O autor

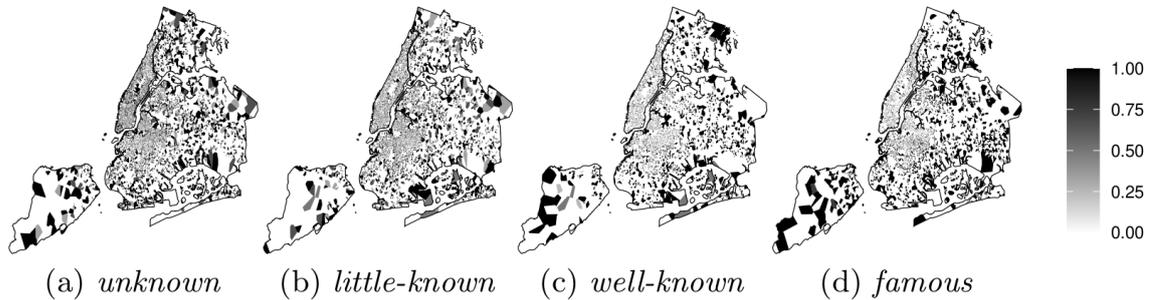
Os valores de uma variável linguística podem ser construídas a partir de adjetivos como alto, baixo, pequeno, médio, grande, por exemplo, de conectivos lógicos como de negação ou conectivos e de modificadores como muito, pouco, levemente e extremamente. O exemplo utilizado na arquitetura proposta deste trabalho tem três características espaciais (*fuzzy*) denotadas pelas seguintes variáveis linguísticas e seus valores linguísticos correspondentes (entre parênteses): preço de acomodação (*cut-rate, cheap, affordable, expensive, premium*), notabilidade de acomodação (*unknown, little-known, well-known, famous*) e segurança alimentar (*very low, low, medium, high, very high*).

2.2.2 Objetos Espaciais *fuzzy*

Sistemas de banco de dados espaciais representam principalmente fenômenos do mundo real como objetos espaciais *crisp* que usam tipos de dados espaciais, como pontos, linhas e regiões *crisp*. Eles representam fenômenos do mundo real com localização exata e uma extensão, forma e limite no espaço precisamente definidos. Os geocientistas estão cada vez mais interessados em modelar fenômenos espaciais que não possuem localizações exatas, limites estritos e interiores uniformes e que, portanto, não podem ser representados adequadamente por objetos espaciais *crisp* [4]. Esse recurso é chamado de imprecisão espacial. Um exemplo é um mapa de Nova York com diferentes tonalidades a partir de um Função de Pertinência utilizando objetos de região plateau para cada valor linguístico da notabilidade de acomodação da característica espacial *fuzzy*, como pode ser visto na figura 3, onde quanto mais escura for a área do mapa, mais forte é o valor de pertinência associada à variável linguística, ou seja, se tomar a Figura 3d como exemplo, áreas mais escuras significam uma maior certeza de que a área da figura é famosa. Outros exemplos são na área de análise financeira, onde os valores de diferentes ativos flutuam rápida e/ou gradualmente ao longo do tempo e na biologia, os quais são usados para representar visualmente grandes e pequenos conjuntos de dados, com foco em analisar padrões e semelhanças em DNA e RNA.

Tipos de dados espaciais *fuzzy* para pontos *fuzzy*, linhas *fuzzy* e regiões *fuzzy* foram formalmente definidos para representá-los. A ideia central é relaxar a estrita decisão de pertencimento (valor 1) ou não pertencimento (valor 0) de um ponto do plano a um objeto espacial. Em vez disso, o pertencimento parcial é permitido e expresso por um valor de associação no intervalo $[0, 1]$. Regiões *fuzzy* são formalmente definidas em [2].

Figura 3 – Demonstração da Função de Pertinência a partir de objetos de região *plateau* para cada valor linguístico da notabilidade de acomodação da característica espacial fuzzy, incluindo os limites da cidade de Nova York.



Fonte: adaptado de [5]

2.3 Manipulação de dados espaciais *fuzzy*

Da mesma forma que conjuntos *crisp* são estendidos para conjuntos *fuzzy*, objetos espaciais *crisp* são generalizados para objetos espaciais *fuzzy*, logo, operações com objetos espaciais *crisp* foram definidas para processar objetos espaciais *fuzzy* analogamente. Alguns exemplos são operações de conjuntos geométricos *fuzzy*, por exemplo, interseção ou sobreposição geométrica *fuzzy*. Assim como as relações topológicas *crisp*, as relações topológicas *fuzzy* foram definidas para avaliar e denominar a posição relativa de dois objetos espaciais *fuzzy*. Esse relacionamento calcula um grau de associação entre 0 e 1 que indica até que ponto o relacionamento é válido. Tal grau pode ser mapeado para um valor linguístico de alto nível e transformado em um predicado booleano.

Um objeto de espacial *fuzzy* é rotulado com um valor linguístico para caracterizar uma instância específica de uma característica espacial, que é representada por uma variável linguística. É necessário aplicar relações topológicas *fuzzy* [2] em regiões *fuzzy* para descobrir regras de associação espacial. Diferentemente de um relacionamento topológico clássico que produz um valor booleano $\{0, 1\}$, um relacionamento topológico *fuzzy* produz um grau de pertinência em $[0, 1]$ que expressa até que ponto uma posição relativa entre dois objetos de região *fuzzy* é mantida. Na arquitetura proposta por este trabalho, foi empregada a relação de sobreposição *fuzzy* para calcular o grau de sobreposição de duas regiões *fuzzy*. Neste caso, será calculada a 'intensidade' desses relacionamentos, a qual pode ser mapeada para um valor linguístico de alto nível. Por exemplo, dois objetos espaciais *fuzzy* se sobrepõem *muito*.

2.4 Mineração de regras de associação

Mineração de dados [8] é o processo de extração de informações ou padrões interessantes (não triviais, implícitos, previamente desconhecidos e potencialmente úteis) de grandes

repositórios de informações como: banco de dados relacional, *data warehouses*, arquivos XML, entre outros, sendo os processos centrais de Descoberta de Conhecimento em Banco de Dados [19].

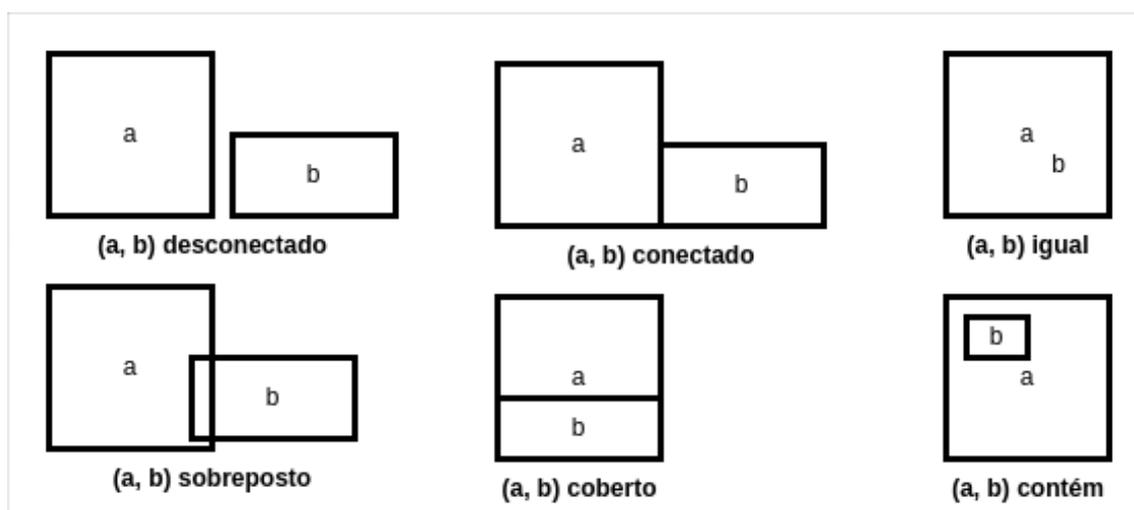
Uma regra de associação é expressa como $A \rightarrow B$ (s%, c%) onde A e B são conjuntos de itens, ou seja, são conjuntos de itens ou elementos que aparecem juntos em uma determinada transação de um banco de dados [1]. Tal regra significa que se o antecedente A ocorre, então o conseqüente B também ocorre com um suporte de s% e confiança de c%. O suporte e a confiança medem a força e o significado das regras. O suporte indica com que frequência os conjuntos de itens A e B aparecem nas transações. Confiança denota a porcentagem de transações contendo A que também contém B. Normalmente, os usuários definem valores mínimos para suporte e confiança para obter regras relevantes.

A mineração de regras de associação, uma das técnicas de mineração de dados mais importantes e bem pesquisadas, foi introduzida pela primeira vez em [1]. Ela visa extrair correlações interessantes, padrões frequentes, associações ou estruturas casuais entre conjuntos de itens nos bancos de dados de transações ou outros repositórios de dados.

2.5 Mineração de regras de associação espaciais

Uma regra de associação espacial é uma regra que indica certas relações de associação entre um conjunto de predicados espaciais e possivelmente alguns predicados não espaciais. Algoritmos para mineração de regras de associação espacial são semelhantes à mineração de regras de associação, exceto a consideração de dados espaciais. Uma regra de associação espacial estende o significado de uma regra de associação clássica, incluindo relacionamentos espaciais nos conjuntos de itens A ou B [12]. A ideia principal é construir transações que armazenem relacionamentos espaciais entre um conjunto de dados espaciais de interesse do usuário e outros conjuntos de dados espaciais que representem diferentes características do aplicativo. Comumente, as relações espaciais empregadas são relações topológicas (como por exemplo, as ilustradas na Figura 4) [16], que expressam como dois ou mais objetos espaciais estão relacionados em relação à sua posição relativa. Por exemplo, é possível ter uma regra que expresse a força e a significância da situação de sobreposição entre a área de cobertura dos alojamentos de tarifa reduzida e as regiões representativas de atrações turísticas. Neste TCC, é considerado que os conjuntos de dados espaciais dos aplicativos armazenam objetos espaciais *fuzzy* e, portanto, é empregado relacionamentos topológicos *fuzzy* na mineração de regras de associação espacial.

Figura 4 – Alguns exemplos ilustrados de relacionamentos topológicos



Fonte: O autor

2.6 Considerações finais

Os conceitos vistos nesse capítulo pertencem a duas áreas principais: banco de dados espaciais e sistemas *fuzzy*. As duas áreas são então combinadas para representar fenômenos geográficos do mundo real com alguma imprecisão. Além dessas áreas, houve a discussão de mineração de dados usadas regras de associação, que são usadas para descobrir elementos que ocorrem em comum dentro de um determinado conjunto de dados. Todos esses conceitos serão aplicados e mesclados na arquitetura implementada proposta, onde a compreensão de cada um deles será fundamental para o entendimento geral da arquitetura.

Capítulo 3

Trabalhos relacionados

Este capítulo resume as características principais de oito trabalhos da literatura relacionados ao tema deste TCC, uma vez que descrevem implementações para mineração de regras espaciais. Este capítulo será dividido em 3 seções, a seção 3.1 descreve os trabalhos que tratam de dados espaciais *crisp*, a seção 3.2 descreve os trabalhos que tratam de dados espaciais *fuzzy* e a seção 3.3 apresenta uma conclusão geral sobre os trabalhos discutidos e uma tabela de comparação dos trabalhos.

3.1 Trabalhos que manipulam dados espaciais *crisp*

Para o primeiro trabalho analisado, há o método proposto em [12], até onde é sabido, é o primeiro método que extrai regras de associação espacial de conjuntos de dados espaciais *crisp*. Ele define formalmente o conceito de regras de associação espacial e emprega ideias-chave de [1], como o uso de limites mínimos para suporte e confiança na extração de regras. Além disso, apresenta hierarquias de conceitos para conjuntos de dados espaciais e relacionamentos topológicos. Uma hierarquia para um conjunto de dados espaciais descreve como os objetos espaciais são organizados com base em seus dados alfanuméricos associados, enquanto uma hierarquia de relacionamentos topológicos agrupa relações semelhantes em uma relação mais grosseira. Tais hierarquias permitem que os usuários explorem as regras extraídas em diferentes níveis de agregações (ou seja, regras de associação de vários níveis).

Na segunda abordagem analisada, os autores do trabalho [15] aplicam conceitos semelhantes para extrair transações espaciais, que são usadas para identificar e gerar as regras de transações espaciais a partir de um conjunto de dados espaciais e seu uso na computação de regras de associação espacial. Propondo explorar a organização taxonômica

do conjunto de dados SIG utilizando uma das camadas como seletor: cada objeto desta camada seleciona um conjunto de objetos das demais camadas juntamente com a relação espacial que relaciona as duas entidades.

Já no trabalho [14], os autores mineram as regras de associação por uma abordagem diferente, utilizando a programação lógica indutiva. Donato Malerba e Francesca A. Lisi implementaram este método no sistema SPADA (Spatial Pattern Discovery Algorithm) que opera em um banco de dados relacional dedutivo configurado por uma etapa inicial de extração de características de um banco de dados espacial.

Os autores do trabalho [6] também desenvolveram uma abordagem diferente, eles mineram as regras baseadas em padrões de células. Mas também não se concentra na expressão do conceito espacial. No entanto, ele se esforça mais na formação da associação da camada de conceito espacial, que pode ser usada para minerar as regras de associação espacial, e então a eficiência do algoritmo de mineração é levantada.

E no último trabalho analisado que manipula dados espaciais *crisp*, o [7] apresentou um algoritmo usado para mineração de regras de associação espacial extraindo conjuntos de itens máximos frequentes com base em uma matriz *booleana*. A ideia principal do algoritmo é criar uma matriz *booleana* com conjuntos de itens de comprimento 1 frequentes como cabeçalhos de linha e IDs de registros de transação como cabeçalhos de coluna. Na matriz, existem apenas dois tipos de valores, '1' e '0', o que significa que o registro da transação contém ou não o conjunto de itens de comprimento 1 frequente correspondente. Então é necessário calcular o número de valor 1 em cada coluna e a contagem das colunas com o mesmo número de valor 1. Se a contagem dessas colunas for maior que o suporte mínimo, conforme, o número de valor 1 na coluna pode ser o tamanho do conjunto de itens frequente máximo, vice-versa.

3.2 Trabalhos que manipulam dados espaciais *fuzzy*

Para este tipo de trabalho, os autores em [9] que especificaram um método para lidar com dados espaciais *fuzzy*. Eles representam a imprecisão espacial usando objetos de região com limites amplos. Isso significa que um objeto real possui três partes adjacentes disjuntas: (i) exterior, que compreende todos os pontos que certamente não pertencem ao objeto, (ii) núcleo, que contém todos os pontos que certamente pertencem ao objeto, e (iii) limite amplo, que consiste em todos os pontos possivelmente pertencentes ao objeto. As regras são baseadas no cálculo de relacionamentos topológicos para objetos de região com limites amplos. Conforme mostrado na tabela 1, este método é limitado pois só trata da expressividade de um conjunto de dados *fuzzy* baseado nessas três regras.

Já os autores em [11] permitem que os usuários especifiquem hierarquias *fuzzy* no sentido de que os graus de pertinência representam em que medida um item de uma hierarquia pertence ao seu pai, permitindo que um item tenha vários e diferentes graus

de pertinência.

E por último trabalho analisado que trata de dados espaciais *fuzzy*, o [13], os autores consideram os graus de pertinência resultantes das relações espaciais (por exemplo, relações de distância) ao calcular o suporte e a confiança das regras. Para isso, definem uma função de pertinência que determina em que medida um valor pertence a um determinado valor linguístico que representa uma relação espacial, como próximo. Conforme mostrado na Tabela 1, essa abordagem lida com a imprecisão espacial de forma limitada. A razão é que ele mapeia atributos alfanuméricos associados a objetos espaciais em funções de associação em vez de usar objetos espaciais difusos.

3.3 Considerações Finais

Ao analisar os trabalhos relacionados, os quais foram oito no total, pode-se concluir que as abordagens mencionadas enfrentam problemas que afetam negativamente a tarefa de descobrir regras de associação espacial a partir de dados espaciais *fuzzy*.

Na Tabela 1, é possível ver uma comparação deste trabalho com os outros trabalhos mencionados anteriormente. Neste TCC é implementado uma arquitetura que consegue manipular os dados espaciais *fuzzy* para o processo de extração de regras. Ao calcular relações topológicas *fuzzy* entre esses recursos e um conjunto de dados espaciais de interesse, a arquitetura permite que os usuários explorem regras de vários níveis. Em particular, o benefício de usar variáveis e valores linguísticos é que eles permitem que os usuários interpretem intuitivamente o que um objeto espacial representa e o significado das relações topológicas nas regras extraídas.

Tabela 1 – Comparação de abordagens existentes com a implementação do método realizado neste TCC (última coluna).

Critério de Comparação	[12]	[15]	[14]	[6]	[7]	[9]	[11]	[13]	Este TCC
Conjuntos espaciais <i>crisp</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓
Conjuntos espaciais <i>fuzzy</i>						✓*		✓*	✓
Relações topológicas <i>crisp</i>	✓	✓	✓				✓		✓
Relações topológicas <i>fuzzy</i>						✓*			✓
Hierarquias de conceitos	✓	✓	✓			✓	✓	✓	✓
Variáveis e valores linguísticos								✓	✓
Regras de associação de múltiplos níveis	✓	✓	✓			✓	✓	✓	✓

*representação ou tratamento limitado de dados espaciais *fuzzy*

Como pode ser visto na tabela 1, lendo-a linha por linha, tem-se que para o critério de Conjuntos espaciais *crisp*, todos os trabalhos satisfazem, por ser um critério mais básico, isso já era o esperado. Já no critério de conjuntos espaciais *fuzzy*, com exceção deste TCC, apenas o [9] e [13], mas ainda com uma manipulação limitada de dados *fuzzy* como foi comentado anteriormente no texto. No terceiro critério, relações topológicas *crisp*, os

trabalhos [6], [7] e [13] não abordaram sobre este tema. No quarto critério, sobre relações topológicas *fuzzy*, exceto este trabalho, apenas o [9] tratou deste tema, mas novamente com uma visão limitada pelas três regras que foram utilizadas no trabalho comentadas anteriormente. Para o critério de hierarquia de conceitos, apenas os trabalhos [6] e [7] não utilizaram deste tema, do restante foi comum para os trabalhos por se tratarem de mineração de regras de associação. Para o critério de Variáveis e valores linguísticos, apenas este TCC e o trabalho [13] utilizaram isso na abordagem. E para o último critério analisado de Regras de associação de múltiplos níveis teve-se os trabalhos [6] e [7] que não trabalharam com este aspecto.

Capítulo 4

Implementação da arquitetura para a descoberta de regras de associação espacial a partir de objetos espaciais *fuzzy*

Este capítulo apresenta a implementação da arquitetura estudada neste TCC, sendo que a seção 4.1 apresenta o exemplo de execução utilizado neste trabalho, a Seção 4.2 apresenta uma visão geral da arquitetura implementada, a Seção 4.3 detalha parâmetros do usuário utilizados nesta arquitetura, a Seção 4.4 mostra em detalhes a camada de dados espaciais da arquitetura, a Seção 4.5 comenta sobre o tratamento para os dados espaciais, a Seção 4.6 detalha o tratamento e manipulação de conjunto de itens, e por fim, a Seção 4.7 exemplifica a Recuperação de Regras de Associação Espacial.

4.1 Exemplo de execução

Para a demonstração da arquitetura, foi feito um exemplo de execução baseado em dados da região da cidade de Nova York e informações de restaurantes extraídos da base de dados da *DOHMH (Department of Health and Mental Hygiene)*, informações de acomodações extraídas do site de hospedagem *Airbnb* e de atrações turísticas extraídos da base de dados do *OpenStreetMaps*, uma amostra desses dados podem ser vistos nas Figuras 5, 6 e 7, respectivamente. O objetivo desse exemplo de execução é extrair regras de associações espaciais que relacionem atrações turísticas e características de acomodações e

restaurantes localizados na cidade de Nova York. O código completo, escrito em R, pode ser acessado no repositório do Github através deste link.

Figura 5 – Amostra de dados extraídos do *DOHMH*

DBA	BORO	STREET	CUISINE DESCRIPTION	INSPECTION DATE	SCORE
SILVER RICE	Brooklyn	FLATBUSH AVENUE	Japanese	01/25/2019	7
EL QUINTO PINO	Manhattan	WEST 24 STREET	Latin American	10/26/2018	13
EARNEST TEA CAFE	Brooklyn	8 AVENUE	Coffee/Tea	12/09/2019	32
WINGSTOP	Manhattan	WEST 125 STREET	Chicken	12/09/2019	40
POPEYES	Brooklyn	KINGS HIGHWAY	Chicken	12/04/2017	
TROQUET	Manhattan	GRAND STREET	French	05/29/2019	9
SHASHLICHNAYA	Queens	QUEENS BLVD	Jewish/Kosher	11/30/2018	12
PROSPECT	Brooklyn	5 AVENUE	American	09/05/2019	4
MUSCLE MAKER GRILL	Manhattan	2 AVENUE		01/01/1900	
FAMOUS CALABRIA PIZZERIA	Manhattan	SAINT JAMES PLACE	Pizza	08/29/2018	32
JUICE PRESS @EQUINOX	Manhattan	AMSTERDAM AVENUE	Juice, Smoothies, Fruit Sa	10/16/2019	34
IHOP	Bronx	ALLERTON AVENUE	Pancakes/Waffles	04/29/2019	12
MAC JUICE	Manhattan	PARK AVENUE	Juice, Smoothies, Fruit Sa	02/13/2018	7
HOOTERS	Queens	190 STREET	American	08/13/2018	17
PERFECTO PIZZERIA	Manhattan	BROADWAY	Pizza	02/28/2018	12
DELICIOUS HUANG BAKERY	Manhattan	HESTER STREET	Chinese	09/07/2021	18
NEW JW GOLDEN BAKERY	Brooklyn	4 AVENUE	Bakery Products/Desserts	12/13/2018	21
TULCINGO RESTAURANT	Queens	ROOSEVELT AVE	Mexican	02/15/2018	25
THE VIG BAR	Manhattan	SPRING STREET	American	01/25/2018	9
THE VIG BAR	Manhattan	SPRING STREET	American	01/25/2018	9
GIUSEPPINA'S BRICK OVEN	Brooklyn	6 AVENUE	Pizza	01/10/2020	7
FRANK AND JOE'S DELI	Bronx	EAST TREMONT AVENUE	Sandwiches	02/01/2017	11

Fonte: O autor

Neste conjunto de dados extraídos do *DOHMH*, visto na Figura 5, pode-se destacar as colunas de *SCORE*, *INSPECTION DATE* e características do estabelecimento como *CUISINE DESCRIPTION* e *BORO*. Para construir esse conjunto de dados, foi executado um script R fornecido pelo próprio *DOHMH*, excluindo as linhas com pontuações negativas e coordenadas de latitude e longitude ausentes/inválidas. Para garantir que os dados sejam compatíveis com os dados do *Airbnb*, foi certificado que o último resultado da inspeção ocorreu antes da última data de extração das acomodações do *Airbnb*, totalizando 22.642 restaurantes.

Figura 6 – Amostra de dados extraídos do Airbnb

name	host_name	latitude	longitude	price	number_of_reviews
Skylit Midtown Castle	Jennifer	4075356	-7398559	150	48
Whole flr w/private bdrm, bath & kitchen(pls read)	LisaRoxanne	4068494	-7395765	75	409
BlissArtsSpace!	Garon	4068535	-7395512	60	50
Spacious Brooklyn Duplex, Patio + Garden	Rebecca	4066265	-7399454	275	2
Large Furnished Room Near B'way	Shunichi	4076457	-7398317	68	507
Cozy Clean Guest Room - Family Apt	MaryEllen	408038	-7396751	75	118
Lovely Room 1, Garden, Best Area, Legal rental	Laurie	4066801	-7398784	98	204
Only 2 stops to Manhattan studio	Allen & Irina	4070935	-7395342	89	181
Uptown Sanctuary w/ Private Bath (Month to Month)	Kae	4080107	-7394255	65	0
UES Beautiful Blue Room	Cyn	4078778	-7394759	62	234
Amazing location! Wburg. Large, bright & tranquil	Joelle	4071248	-7395881	90	9
Perfect for Your Parents: Privacy + Garden	Jane	4069121	-7397277	199	280
2 Furnished bedrooms next to Central Park	Sing	4079607	-7394949	96	0
Sweet and Spacious Brooklyn Loft	Chaya	4071778	-739568	299	9
Maison des Sirenes1,bohemian, luminous apartment	Nathalie	4068462	-7393839	140	139
Midtown Pied-a-terre	Tommi	4076724	-7398664	175	58
Modern 1 BR / NYC / East Village	Dana	4072712	-7398598	175	43
Spacious 1 bedroom in luxe building	Teri	4081317	-7395465	55	98
Large B&B Style rooms	Angela	4068294	-7395682	82	77
Lovely Room 2; Garden; Best area, Legal	Laurie	4066692	-7398981	108	195

Fonte: O autor

Na Figura 6 é exemplificado uma amostra de dados extraído da base do *Airbnb*. Nela pode-se destacar as seguintes colunas: *name*, *price*, *number_of_reviews*, *latitude* e *longitude*, que significam o título do anúncio da acomodação no site do *Airbnb*, preço, quantidade de avaliações, e coordenadas geográficas da acomodação, respectivamente. No conjunto de dados total obteve-se 38.277 acomodações.

Figura 7 – Amostra de dados extraídos do *OpenStreetMaps*

osm_id	name	geom	category	type
1	-9462533	Fort Tilden Historic District	c(-8224723.90215783, -8224767.18317584, -8225252.68087...	historic district
2	-9460479	Floyd Bennett Field Historic District	c(-8226519.7861067, -8226932.88160503, -8224746.533410...	historic district
3	-3971052	BMT 12	c(-8233635.20550577, -8233637.46529144, -8233639.46904...	historic yes
4	-3971052	BMT 12	c(-8233627.73596794, -8233630.0848092, -8233632.322330...	historic yes
5	-3971051	BMT 10	c(-8225313.39452124, -8225324.7713732, -8225333.131466...	historic yes
6	-3971051	BMT 10	c(-8225324.78250515, -8225336.14822516, -8225344.57511...	historic yes
7	-3959638	MJ	c(-8225313.39452124, -8225324.7713732, -8225333.131466...	historic yes
8	-3959638	MJ	c(-8225324.78250515, -8225336.14822516, -8225344.57511...	historic yes
9	-3959637	BMT 11	c(-8225313.39452124, -8225324.7713732, -8225333.131466...	historic yes
10	-3959637	BMT 11	c(-8225324.78250515, -8225336.14822516, -8225344.57511...	historic yes
11	-1376500	Main Line (New York Central Railroad)	c(-8229084.09736846, -8229090.47597528, -8229093.01405...	historic yes
12	1007920043	Horse Brook	c(-8219822.14875651, -8220228.88791191, -8220353.81064...	historic stream
13	1007920044	Kissena Creek	c(-8219609.71777226, -8218905.5552014, -8218695.439662...	historic stream
14	1013146461	Vaults	c(-8245929.73081744, -8245931.43400565, -8245931.40060...	historic maritime
15	-13562750	Lemon Creek Tidal Wetlands Area	c(-8260459.23980935, -8260446.51599155, -8260450.80179...	leisure nature_reserve
16	-13562750	Lemon Creek Tidal Wetlands Area	c(-8260444.34526148, -8260373.12305128, -8260360.69979...	leisure nature_reserve
17	-6463871	Clay Pit Ponds State Park Preserve	c(-8262944.99290646, -8262903.11451403, -8262882.29776...	leisure nature_reserve
18	-6463871	Clay Pit Ponds State Park Preserve	c(-8262718.32415934, -8262725.21483582, -8262731.97192...	leisure nature_reserve

Fonte: O autor

Como pode ser visto na Figura 7, tem-se a coluna *osm_id* (um identificador espacial geográfico), *name* para o nome do local, *geom* contendo informações do objeto geométrico

(ponto, linha ou região) e *category* e *type* para a categoria e tipo do local, respectivamente. Como resultado, obteve-se 12.328 objetos espaciais de regiões *crisp*, com informações do tipo e categoria das atrações turísticas.

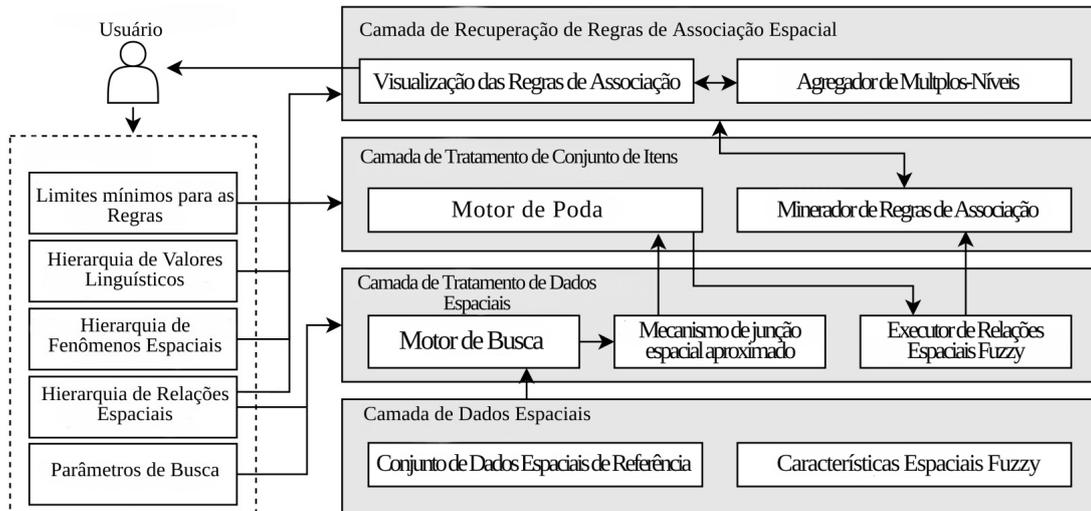
Utilizando-se os conjuntos de dados apresentados, no exemplo de execução tem-se como objetivo entender como as áreas associadas com turismo se relacionam com tais áreas como acomodações e restaurantes.

4.2 Visão geral da arquitetura

A Figura 8 mostra a arquitetura do método desenvolvido, que extrai regras de associação espacial de conjuntos de dados espaciais *fuzzy*. A descoberta é feita de forma dinâmica, realizada por uma entrada do usuário, indicado por um conjunto de parâmetros (Seção 4.3), os quais são usados no processo de mineração das regras. Eles especificam um conjunto de hierarquias para descrever como os relacionamentos espaciais e os próprios dados são organizados semanticamente e definem os limites mínimos que as regras devem satisfazer, determinando quais subconjuntos dos conjuntos de dados disponíveis serão usados no processo de extração, afim de deixar o processo mais eficiente e flexível.

Na figura, a arquitetura foi dividida em 4 camadas, e são elas: (i) Camada de Dados Espaciais, (ii) Camada de Tratamento de Dados Espaciais, (iii) Camada de Tratamento de Conjuntos de Itens e (iv) Camada de Recuperação de Regras de Associação Espacial. Conforme indicado na Figura 8, cada camada aplica alguns parâmetros específicos do usuário para realizar seu processamento. Em resumo, a Camada de Dados Espaciais fornece os conjuntos de dados a serem usados no processo de extração. A Camada de Tratamento de Dados Espaciais busca os dados relevantes da Camada de Dados Espaciais de acordo com os parâmetros do usuário. Além disso, esta camada computa relações espaciais *fuzzy* em objetos espaciais *fuzzy* representados como objetos espaciais *plateau*. O objetivos na Camada de Manipulação de Conjuntos de Itens: (i) organizar os relacionamentos e os conjuntos de itens processados e filtrá-los em relação ao suporte mínimo indicado pelo usuário, e (ii) descobrir as regras de associação espacial usando um algoritmo de mineração, como o Apriori [1]. A extração de regras é baseada em atributos alfanuméricos de Conjuntos Espaciais de Referência e valores linguísticos que representam os resultados de relações topológicas *fuzzy* e as caracterizações de Características Espaciais *Fuzzy*. Em seguida, a Camada de Recuperação de Regras de Associação Espacial é responsável por apresentar de forma interativa as regras mineradas, pois elas podem ser organizadas em vários níveis de acordo com a definição de hierarquias. As camadas da arquitetura implementada são detalhadas nas Seções 4.3 a 4.7.

Figura 8 – Esquematização da Arquitetura.



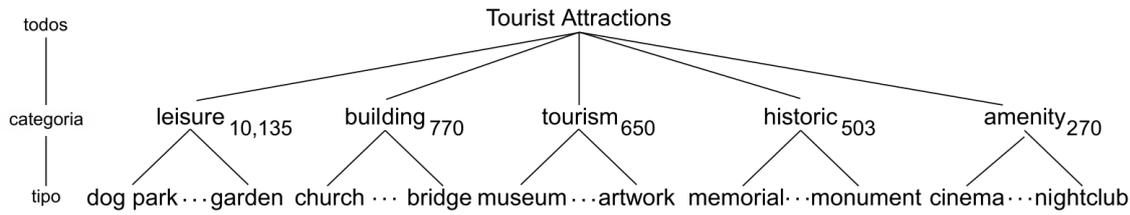
Fonte: o próprio a autor

4.3 Parâmetros do Usuário

Os parâmetros fornecidos pelo usuário desempenham um papel importante na arquitetura. Conforme mostrado na Figura 8, eles consistem em (i) Parâmetros de Consulta, (ii) Hierarquia de Relações Espaciais, (iii) Hierarquias de Fenômenos Espaciais, (iv) Hierarquias de Valores Linguísticos, e (v) Limites Mínimos para as Regras. O objetivo dos Parâmetros de Consulta é indicar (i) o Conjunto de Dados Espaciais de Referência relevante da aplicação, (ii) as Características Espaciais *fuzzy* que serão consideradas no processo de mineração, (iii) a relação topológica *fuzzy* que deve ser empregada ao associar os dados espaciais, e (iv) condições aplicadas aos atributos alfanuméricos que estão associados ao Conjunto de Dados Espaciais de Referência escolhido. O objetivo desta arquitetura é encontrar regras de associação espacial baseadas no grau de sobreposição entre atrações turísticas e características espaciais *fuzzy* que representam áreas de acomodações e restaurantes com características distintas. Para isso, os Parâmetros de Consulta são: (i) atrações turísticas como Conjunto de Dados Espaciais de Referência, (ii) três Características Espaciais *fuzzy* contendo regiões *fuzzy* implementadas como objetos de região de *plateau*, e (iii) sobreposição *fuzzy* como a relação topológica *fuzzy*.

O objetivo de definir hierarquias (parâmetros (ii)-(iv)) é organizar semanticamente conceitos relacionados a relações espaciais e características de dados espaciais. Cada hierarquia permite pelo menos um dos seguintes processos: (i) descoberta e visualização de mineração de regras de associação espacial em níveis distintos, (ii) poda de características espaciais *fuzzy* e (iii) cálculo de relacionamentos topológicos aproximados. Uma hierarquia permite que os usuários definam como os conceitos de nível inferior estão relacionados aos

Figura 9 – Hierarquia de fenômenos espaciais do exemplo executado na arquitetura.



Fonte: O próprio autor

conceitos de nível superior. Dados dois atributos A e B onde cada atributo tem seu próprio domínio de valores (ou conceitos), A e B estão relacionados em uma hierarquia caso seja possível determinar um valor de B agregando um subconjunto de valores em A. Neste caso, têm-se $A \rightarrow B$, expressando que esses atributos são níveis de uma hierarquia.

As Hierarquias de Fenômenos Espaciais são definidas usando atributos alfanuméricos que rotulam objetos espaciais *crisp* armazenados nos Conjuntos de Dados Espaciais de Referência. Ela permite que os usuários explorem e visualizem regras de associação espacial em diferentes níveis (como ilustrado na Seção 4.5). Por exemplo, os usuários podem primeiro identificar regras contendo conjuntos de itens de um nível mais geral em uma hierarquia. Então, eles podem minerar regras em um nível mais detalhado a partir de um valor específico do nível anterior. A Figura 9 mostra a Hierarquia de Fenômenos Espaciais construída sobre os atributos alfanuméricos do Conjunto de Dados Espaciais de Referência. Ela é definido por $\text{tipo} \rightarrow \text{categoria}$, com o nível superior dessa hierarquia, denominado *todos*, o qual agrega todos os valores da categoria e, como resultado, consiste em todas as atrações turísticas.

A Hierarquia das Relações Espaciais define como as relações espaciais são organizadas, tal organização serve como referência para relacionamentos de computação na Camada de Tratamento de Dados Espaciais (Seção 4.5). Ele segue princípios semelhantes ao método em [12] no sentido de que níveis mais altos da hierarquia representam relações mais grosseiras (fracas ou menos confiáveis). A principal diferença é que neste trabalho é lidado com relações topológicas *fuzzy* no nível inferior. Por exemplo, a sobreposição *fuzzy* e o encontro *fuzzy* podem ser agrupados em um relacionamento mais grosseiro chamado interseção, que pode ser calculado usando aproximações espaciais.

A Hierarquia de Valores Linguísticos descreve como os valores linguísticos de uma Característica Espacial *Fuzzy* são estruturados. Esse tipo de hierarquia permite que os usuários explorem os padrões descobertos em níveis distintos de agrupamento de valores linguísticos. Por exemplo, é possível especificar uma hierarquia de valores linguísticos para o traço espacial *fuzzy* representando a notabilidade da acomodação como $\text{Vall} \rightarrow \text{classe}$ onde $\text{Vall} = \{unknown, little-known, well-known, famous\}$ e $\text{classe} = \{unpopular, popular\}$. Ao agrupar os valores *unknown* e *little-known*, obtem-se a classe *unpopular*, enquanto a agregação de *well-known* e *famous* leva à classe *popular*.

Finalmente, os Limites Mínimos para Regras determinam os níveis desejados de força e significância que as regras mineradas devem satisfazer. Por exemplo, valores mínimos para suporte e confiança. O usuário pode fornecer diferentes limites mínimos para cada nível da hierarquia, pois a frequência dos conjuntos de itens diminui de acordo que é descido os níveis da hierarquia. Durante a execução da arquitetura, foi extraída regras de associação variando os níveis apenas na Hierarquia de Fenômenos Espaciais. Para a categoria de nível, foi utilizado o suporte de 10% e confiança mínima de 30%. No nível seguinte, foi empregado 5% e 20%, respectivamente.

4.4 Camada de Dados Espaciais

A Camada de Dados Espaciais é composta por dois componentes: (i) Conjuntos de Dados Espaciais de Referência e (ii) Características Espaciais Fuzzy. O objetivo do processo de mineração é descobrir associações entre objetos de um Conjunto de Dados de Referência Espacial e objetos de Características Espaciais Fuzzy. Assim, o primeiro componente consiste em uma coleção de objetos espaciais extraídos *OpenStreetMap* (um site mantido pela comunidade onde tem dados de localizações geográficas do mundo todo), os quais foram tratados e extraídos apenas locais de atrações turísticas, conforme mostrado na Seção 4.1.

Figura 10 – Conjunto de dados usados para a construção de objetos espaciais *fuzzy*

	Longitude	Latitude	SCORE
1	-73.96056	40.65962	7
2	-74.00127	40.74733	13
3	-73.99072	40.66048	7
4	-73.96367	40.72952	12
5	-73.86350	40.87645	12
6	-73.87978	40.88179	9
7	-73.95628	40.71718	12
8	-73.84919	40.73216	9
9	-73.95721	40.71214	12
10	-73.97690	40.64904	10
11	-73.99010	40.71836	13
12	-73.95695	40.77764	12
13	-74.00538	40.73814	9
14	-73.96875	40.72104	10
15	-73.83110	40.84744	4
16	-74.00455	40.74718	9
17	-73.99786	40.59689	12
18	-73.90823	40.70311	5

	longitude	latitude	price	number_of_reviews
1	-73.98559	40.75356	150	48
2	-73.95765	40.68494	75	409
3	-73.95512	40.68535	60	50
4	-73.99454	40.66265	275	2
5	-73.96317	40.76457	68	507
6	-73.96751	40.80380	75	118
7	-73.98784	40.66801	98	204
8	-73.95342	40.70935	89	181
9	-73.94255	40.80107	65	0
10	-73.94759	40.78778	62	234
11	-73.95881	40.71248	90	9
12	-73.97277	40.69121	199	280
13	-73.94949	40.79607	96	0
14	-73.95680	40.71778	299	9
15	-73.93839	40.68462	140	139
16	-73.98664	40.76724	175	58
17	-73.98598	40.72712	175	43
18	-73.95465	40.81317	55	98

- (a) *Dataframe* contendo informações sobre restaurantes da cidade de Nova York
- (b) *Dataframe* contendo informações sobre acomodações da cidade de Nova York

Fonte: O autor

O segundo componente refere-se a Características Espaciais *Fuzzy*, onde cada recurso é um conjunto de dados espaciais *fuzzy* e representa uma característica particular da apli-

cação. Foi usado a função *spa_creator* do pacote *fsr* do R para gerar esse conjunto de dados espaciais *fuzzy*. Foram definidos três tipos de dados espaciais *fuzzy*, eles são para a popularidade e preço de uma acomodação e segurança de um restaurante. Uma amostra desses dados pode ser visto nas Figuras 10a e 10b. Na Figura 10a tem-se o *dataframe* contendo informações de restaurantes da cidade de Nova York, com as colunas Longitude e Latitude, afim de definir sua localização e a coluna SCORE contendo a nota obtida pelo restaurante nas inspeções da *DOHMH* (*Department of Health and Mental Hygiene*). Já a Figura 10b é um *dataframe* contendo informações de acomodações na cidade de Nova York, também com as colunas Longitude e Latitude para localização, e as colunas de *price* e *number_of_reviews* (preço e quantidade de avaliações das acomodações, respectivamente). Assim, um recurso espacial *fuzzy* é rotulado com uma variável linguística, como preço e popularidade para as acomodações e segurança para os restaurantes, e consiste em um ou mais objetos espaciais *fuzzy* anotados por um valor linguístico já discutido anteriormente.

No exemplo de aplicação, tem-se um conjunto de dados espaciais de referência que armazena atrações turísticas na cidade de Nova York. As Características Espaciais *Fuzzy* da aplicação de exemplo representam as variáveis linguísticas *food safety*, *accommodation notability* e *accommodation price*. Por exemplo, Figura 3 mostra a seguinte Característica Espacial *Fuzzy* para a variável linguística *Accommodation Notability*, sendo composta por 4 regiões *fuzzy* respectivamente mostradas na Figura 3a à 3d.

4.5 Camada de Tratamento de Dados Espaciais

Essa camada é responsável por tratar os dados carregados da Camada de Dados Espaciais e processar os relacionamentos topológicos. Para isso, ela emprega os seguintes componentes: (i) Motor de Busca, (ii) Mecanismo de Junção Espacial Aproximado e (iii) Executor de Relações Espaciais *Fuzzy*. O Motor de Busca captura o subconjunto relevante de objetos espaciais da Camada de Dados Espaciais de acordo com os Parâmetros de Consulta fornecidos pelo usuário (Seção 4.3).

Em seguida, o Mecanismo de Junção Espacial Aproximado calcula relações topológicas aproximadas entre os objetos espaciais capturados do Conjunto de Dados Espaciais de Referência e as Características Espaciais *Fuzzy*. O cálculo de uma relação aproximada é baseado em *minimum bounding rectangles* (MBRs) e maiores detalhes são dadas em [17].

Finalmente, o Executor de Relações Espaciais *Fuzzy* calcula as relações topológicas *fuzzy* entre os objetos espaciais *crisp* do Conjunto de Dados Espaciais de Referência e os objetos espaciais *fuzzy* selecionados. A Figura 12 mostra o processo de construção dos objetos espaciais *fuzzy* (conforme discutido na seção anterior) que são usados na etapa de processamento dos relacionamentos topológicos *fuzzy*. Tais processamentos são feitos usando o *fsr* [5], sendo que foi usado a sobreposição (*overlap*) como predicado topológico.

Figura 11 – Amostra de código da arquitetura escrito em R que representa uma parte da construção de objetos espaciais *fuzzy*

```

1 lvals_accom_notability = c("unkown", "little_known", "well_known", "famous")
2
3 unkown_mf <- genmf("trapmf", c(0, 1, 3, 5))
4 little_known_mf <- genmf("trapmf", c(3, 5, 8, 9))
5 well_known_mf <- genmf("trapmf", c(8, 9, 28, 32))
6 famous_mf <- genmf("trapmf", c(28, 32, 1009, 1009))
7
8 accom_notability <- accom_nyc[, c("longitude", "latitude", "number_of_reviews")]
9 accom_notability_layer <- spa_creator(accom_notability, classes = lvals_accom_notability,
10                                     mfs = c(unkown_mf, little_known_mf, well_known_mf, famous_mf),
11                                     base_poly = nyc)

```

Fonte: o próprio autor.

Figura 12 – Hierarquia de fenômenos espaciais do exemplo executado na arquitetura.

	tourist_att_type	tourist_att_category	topological_relationship	food_safety.very low	food_safety.low	food_safety.medium	food_safety.high
147	garden	leisure	overlap	a little bit	a little bit	slightly	mostly
148	park	leisure	overlap	a little bit	a little bit	averagely	mostly
149	pitch	leisure	overlap	a little bit	a little bit	a little bit	quite
150	cinema	amenity	overlap	a little bit	a little bit	a little bit	averagely
151	pitch	leisure	overlap	a little bit	a little bit	a little bit	averagely
152	pitch	leisure	overlap	a little bit	a little bit	a little bit	quite
153	pitch	leisure	overlap	a little bit	a little bit	a little bit	quite
154	garden	leisure	overlap	a little bit	a little bit	mostly	slightly
155	nature_reserve	leisure	overlap	a little bit	a little bit	mostly	mostly
156	nature_reserve	leisure	overlap	a little bit	a little bit	averagely	mostly
157	pitch	leisure	overlap	a little bit	a little bit	mostly	averagely
158	garden	leisure	overlap	a little bit	a little bit	a little bit	a little bit
159	garden	leisure	overlap	a little bit	a little bit	a little bit	a little bit

Fonte: o próprio autor.

A Figura 12 representa uma amostra do conjunto de dados obtido depois de computar os relacionamentos topológicos *fuzzy*. Nela pode-se ver como um exemplo na linha 150, uma atracção turística do tipo *cinema*, da categoria *amenity*, apresenta uma sobreposição com a área que representa alta segurança na comida (*food_safety.high*) e dos outros valores linguísticos bem pouca relação (apenas o *a little bit*).

4.6 Camada de Tratamento de Conjuntos de Itens

Essa camada é responsável por duas ações principais: (i) realizar operações de poda e (ii) minerar as regras de associação espacial. A operação de poda ocorre com base no conjunto de tuplas fornecido pelo Mecanismo de Junção Espacial Aproximado e no suporte

mínimo para o nível mais alto da Hierarquia de Fenômenos Espaciais e mais detalhes são fornecidos em [17].

O Minerador de Regras de Associação extrai regras usando o conjunto de tuplas retornado pelo Executor de Relações Espaciais Fuzzy. Este conjunto é reformulado como um conjunto de dados transacional, que é dado como entrada para o algoritmo Apriori [1]. Para cada nível de cada hierarquia (por exemplo, a Hierarquia de Fenômenos Espaciais), é encontrado conjuntos de itens com uma frequência de aparecimento superior ao suporte mínimo naquele nível específico. Esses conjuntos de itens são conjuntos de itens frequentes que serão usados para formar regras de associação. As regras fortes (por exemplo, com confiança maior que o limite mínimo) são enviadas para a Camada de Recuperação de Regras de Associação Espacial.

Na arquitetura, foi utilizado o pacote R *arules* para identificar os conjuntos de itens e regras de associação frequentes sobre o exemplo de aplicação usado neste TCC, com suporte e confiança com valores de 10% e 30%, respectivamente. Um exemplo de uma regra de associação espacial é: $\{tourist_att_category=leisure\} \rightarrow \{accom_price.affordable=quite\}$ com suporte de 31% e confiança de 82%, a qual apareceu 3808 vezes. A regra significa que a chance de ter uma acomodação com um preço aceitável perto de uma atração turística da categoria *leisure* é bem provável. Este conjunto de itens se refere ao número de transações em que o atributo *category* é igual a *leisure* no Conjunto de Dados Espaciais de Referência. A imagem 13 contém alguns exemplos de regras extraídas, as colunas representam a regra, suporte e a quantidade de vezes em que ela aparece, respectivamente.

Figura 13 – Exemplos de regras de Categoria extraídas pela arquitetura.

Regras de Categoria	Suporte	Confiança	Quantidade
$\{tourist_att_category=leisure\} \Rightarrow \{accom_price.affordable=quite\}$	31%	82%	3808
$\{food_safety.very\ low=a\ little\ bit,accom_price.affordable=quite\} \Rightarrow \{tourist_att_category=leisure\}$	30%	82%	3763
$\{food_safety.very\ low=a\ little\ bit,food_safety.high=quite\} \Rightarrow \{tourist_att_category=leisure\}$	30%	81%	3713

Fonte: o próprio autor.

Para as regras extraídas da camada de hierarquia de tipo, não obteve-se um suporte alto com a confiança também alta, o que indica que o suporte alto indica a obtenção de conjuntos de itens frequentes, e não em uma implicação propriamente dita. A partir de suportes mais baixos foi obtido uma confiança maior. Um exemplo é a primeira regra da figura 14, onde a regra $\{tourist_att_type=pitch, food_safety.medium=a\ little\ bit\} \rightarrow \{food_safety.high=quite\}$ que pode ser lida como: atrações turísticas do tipo *pitch* com um pouco de segurança alimentar média, implicam numa alta probabilidade de ter restaurantes de alto nível de segurança alimentar aos arredores.

Figura 14 – Exemplos de regras de Tipo extraídas pela arquitetura.

Regras de Tipo	Suporte	Confiança	Quantidade
$\{tourist_att_type=pitch, food_safety.medium=a\ little\ bit\} \Rightarrow \{food_safety.high=quite\}$	10%	68%	1258
$\{food_safety.high=quite\} \Rightarrow \{tourist_att_type=pitch\}$	12%	33%	1537
$\{tourist_att_type=garden\} \Rightarrow \{food_safety.high=quite\}$	6%	40%	795

Fonte: o próprio autor.

Figura 15 – Amostra de código da arquitetura escrito em R que representa uma parte da mineração de regras de associação espaciais.

```

1 # Generating rules with the APRIORI algorithm - PARAMETER FOR THE HIGHEST LEVEL OF THE HIERARCHY
2 support = 0.1
3 confidence = 0.3
4 maxlength = 11
5
6 # depicting the most frequent itemsets in the selected hierarchy
7 itemsets_category <- apriori(transaction_data_filtered, supp = support, conf = confidence,
8                             maxlen = maxlength, target = "frequent itemsets")
9

```

Fonte: o próprio autor.

Na Figura 15 mostra uma parte do código que minera as regras de associação espaciais utilizando a função do R *apriori*, que é utilizada para encontrar conjunto de itens frequentes. Nas variáveis como *support*, *confidence* e *maxlength*, que significam o suporte e confiança das regras, e a quantidade máxima das regras geradas, respectivamente.

4.7 Camada de Recuperação de Regras de Associação Espacial

O objetivo desta camada é apresentar interativamente o conhecimento descoberto ao usuário. Ele consiste em dois componentes: (i) Visualização de Regras de Associação e (ii) Agregador Multinível. O primeiro componente é uma interface de usuário que permite

a visualização geral e exploração das regras de associação espacial fortes extraídas pelo Minerador de Regras de Associação. As regras podem ser visualizadas em formatos distintos, como tabelas, gráficos, gráficos de dispersão e coordenadas paralelas. A Figura 13 mostra três regras do exemplo de execução usando uma planilha para representá-las.

O Agregador de Múltiplos Níveis permite aos usuários identificar regras em diferentes níveis de uma hierarquia de conceitos. Isso é realizado principalmente selecionando as regras que contêm conjuntos de itens com os membros desejados de um determinado nível. Este componente foi integrado para minerar regras em cada nível na Hierarquia de Fenômenos Espaciais (Figura 9) usando os parâmetros fornecidos na Seção 4.3. As três primeiras regras na Figura 13 estão relacionadas ao nível mais alto dessa hierarquia (ou seja, *category*). Esse componente interage diretamente com o componente de Minerador de Regras de Associação, conforme descrito na seção anterior.

Capítulo 5

Conclusão

Este capítulo encerra o trabalho atual, comentando as conclusões sobre o desenvolvimento e os resultados obtidos. A Seção 5.1 apresenta as considerações finais sobre o trabalho desenvolvido, a seção Seção 5.2 descreve orientações para trabalhos futuros.

5.1 Conclusões

Com o crescimento da ciência de dados espaciais, faz-se necessário buscar novas formas para obter meios para avançar em conceitos da área. Há várias abordagens descritas na literatura, porém muitas acabam não satisfazendo a utilização delas no mundo real.

Dentro deste contexto, neste trabalho foi apresentado uma arquitetura para extrair regras de associação espacial de dados espaciais *fuzzy*. Foi lidado adequadamente com a *vagueza espacial*, uma vez que foi extraída regras de associação espacial baseadas em valores linguísticos expressando relações entre objetos espaciais *fuzzy* e objetos espaciais *crisp*. Este TCC ainda mostrou um exemplo de aplicação usando um estudo de caso. Foi ainda utilizado o pacote *fsr* na arquitetura para a computação de valores topológicos e outras relações *fuzzy* e também foi tudo implementando usando uma sequência passos.

5.2 Trabalhos Futuros

Os trabalhos futuros serão focados em dois tópicos principais:

- Integração da arquitetura com o pacote *fsr*. O objetivo é implementar a presente arquitetura na forma de uma função e melhorando sua usabilidade em aplicações.

- Adaptar a arquitetura e deixá-la mais auto-parametrizada. Isto é, planeja-se diminuir os parâmetros requeridos ou implementar métodos que atribuam valores automáticos a eles.

Referências

- [1] Rakesh Agrawal, Tomasz Imieliński e Arun Swami. “Mining Association Rules between Sets of Items in Large Databases”. Em: *ACM SIGMOD Int. Conf. on Management of Data*. 1993, pp. 207–216.
- [2] A. C. Carniel e M. Schneider. “A Conceptual Model of Fuzzy Topological Relationships for Fuzzy Regions”. Em: *IEEE Int. Conf. on Fuzzy Systems*. 2016, pp. 2271–2278.
- [3] A. C. Carniel e M. Schneider. “A Survey of Fuzzy Approaches in Spatial Data Science”. Em: *IEEE Int. Conf. on Fuzzy Systems*. 2021, pp. 1–6.
- [4] Anderson Chaves Carniel e Markus Schneider. “A conceptual model of fuzzy topological relationships for fuzzy regions”. Em: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2016, pp. 2271–2278. DOI: <10.1109/FUZZ-IEEE.2016.7737976>.
- [5] Anderson Chaves Carniel et al. “Handling Fuzzy Spatial Data in R Using the *fsr* Package”. Em: *ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems*. 2021, pp. 526–535.
- [6] Jiangping Chen et al. “An algorithm about spatial association rule mining based on cell pattern”. Em: *Geoinformatics: Geospatial Information Science*. 2006, pp. 662–671.
- [7] Junming Chen, Guangfa Lin e Zhihai Yang. “Extracting spatial association rules from the maximum frequent itemsets based on Boolean matrix”. Em: *Int. Conf. on Geoinformatics*. 2011, pp. 1–5.
- [8] Ming-Syan Chen, Jiawei Han e P.S. Yu. “Data mining: an overview from a database perspective”. Em: *IEEE Transactions on Knowledge and Data Engineering* 8.6 (1996), pp. 866–883. DOI: <10.1109/69.553155>.

-
- [9] Eliseo Clementini, Paolino Di Felice e Krzysztof Koperski. “Mining multiple-level spatial association rules for objects with a broad boundary”. Em: *Data & Knowledge Engineering* 34.3 (2000), pp. 251–270.
- [10] Ralf Hartmut Güting Dr.rer.nat. “An introduction to spatial database systems”. Em: *The VLDB Journal* 3 (2005), pp. 357–399.
- [11] Esen Kacar e Nihan K. Cicekli. “Discovering fuzzy spatial association rules”. Em: *Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV*. 2002, pp. 94–102.
- [12] Krzysztof Koperski e Jiawei Han. “Discovery of spatial association rules in geographic information databases”. Em: *Int. Symp. on Spatial Databases*. 1995, pp. 47–66.
- [13] R. Ladner, F. E. Petry e M. A. Cobb. “Fuzzy Set Approaches to Spatial Data Mining of Association Rules”. Em: *Transactions in GIS* 7.1 (2003), pp. 123–138.
- [14] Donato Malerba e Francesca A. Lisi. “An ILP method for spatial association rule mining”. Em: *First Workshop on Multi-Relational Data Mining*. 2001, pp. 18–29.
- [15] Salvatore Rinzivillo e Franco Turini. “Extracting spatial association rules from spatial transactions”. Em: *ACM Int. Workshop on Geographic Information Systems*. 2005, pp. 79–86.
- [16] M. Schneider e T. Behr. “Topological Relationships between Complex Spatial Objects”. Em: *ACM Trans. on Database Systems* 31.1 (2006), pp. 39–81.
- [17] Henrique P. da Silva et al. “Discovery of Spatial Association Rules from Fuzzy Spatial Data”. Em: *Int. Conf. on Conceptual Modeling*. 2022, pp. 1–15.
- [18] L. A. Zadeh. “Fuzzy Sets”. Em: *Information and Control* 8.3 (1965), pp. 338–353.
- [19] Qiankun Zhao e Sourav S Bhowmick. “Association rule mining: A survey”. Em: *Nanyang Technological University, Singapore* 135 (2003).

APÊNDICE A

Discovery of Spatial Association Rules from Fuzzy Spatial Data

This preprint has not undergone peer review (when applicable) or any post-submission improvements or corrections. The Version of Record of this contribution is published in Conceptual Modeling, 41st International Conference, ER 2022, and is available online at <https://doi.org/10.1007/978-3-031-17995-2_13>.

Discovery of Spatial Association Rules from Fuzzy Spatial Data

Henrique P. da Silva¹, Thiago D. R. Felix², Pedro V. A. B. de Venâncio³[0000-0003-4665-562X], and Anderson C. Carniel²[0000-0002-8297-9894]

¹ Federal University of Technology - Paraná, Brazil

henriquepigozzo@gmail.com

² Dept. of Computer Science, Federal University of São Carlos, Brazil

thiagofelix@estudante.ufscar.br, accarniel@ufscar.br

³ Grad. Program in Electrical Engineering, Federal University of Minas Gerais, Brazil

pedrovincius@ufmg.br

Abstract. The discovery of *spatial association rules* is a core task in spatial data science projects and focuses on extracting useful and meaningful spatial patterns and relationships from spatial and geometric information. Many spatial phenomena have been modeled and represented by *fuzzy spatial objects*, which have blurred interiors, uncertain boundaries, and inexact locations. In this paper, we introduce a novel method for mining spatial association rules from fuzzy spatial data. By allowing users to represent spatial features of their applications as fuzzy spatial objects and by employing fuzzy topological relationships, our method discovers spatial association patterns between spatial objects of users' interest (e.g., tourist attractions) and such fuzzy spatial features (e.g., sanitary conditions of restaurants, number of reviews and price of accommodations). Further, this paper presents a case study based on real datasets that shows the applicability of our method.

Keywords: Spatial data science · Spatial association rule · Spatial fuzziness · Fuzzy spatial data · Fuzzy topological relationship.

1 Introduction

Increasingly, applications have required specialized and sophisticated methods for exploring the special geometric and topological characteristics of spatial phenomena, such as location and spatial relationships. *Spatial data science* emerges as an important area that provides such methods [3]. The common assumption is that spatial phenomena are represented by instances of vector-based data types called *spatial data types* [14], such as *points*, *lines*, and *regions*. The locations, geometric shapes, and boundaries of such instances are precisely defined in space. Hence, these instances are denominated as *crisp spatial objects*. This also means that the information extracted from these objects is exact. For instance, topological relationships (e.g., overlap) on crisp spatial objects yield exact results.

However, many spatial phenomena are characterized by *spatial fuzziness* [3]. Spatial objects with this feature have blurred interiors, uncertain boundaries,

and/or inexact locations. Such objects cannot be adequately represented by crisp spatial objects. They are properly represented by instances of *fuzzy spatial data types*, such as *fuzzy points*, *fuzzy lines*, and *fuzzy regions*. Fuzzy set theory [15] is used to model such *fuzzy spatial objects*. The key idea is to assign a *membership degree* between 0 and 1 to each point of a fuzzy spatial object. This degree indicates to which extent a point belongs to the object. An example is a fuzzy region object that represents the coverage area of *expensive accommodations*. In this case, we represent areas with a particular characterization expressed by a *linguistic value* (i.e., *expensive*) in a given context denoted by a *linguistic variable* (i.e., *accommodations*). In this object, points with membership degree 1 represent certainly expensive locations. Points with degree 0 denote locations that are definitely not expensive. The remaining points (i.e., with degrees in $]0, 1[$) characterize locations that are partially expensive (i.e., different degrees of truth). Spatial operations that handle fuzzy spatial objects are also fuzzy since they have to deal with the membership degrees of the objects. For instance, a *fuzzy topological relationship* (e.g., fuzzy overlap) [2] on fuzzy spatial objects yields a value in $[0, 1]$ that indicates the degree of truth of the relationship.

The discovery of *spatial association rules* [10] is a core aspect of spatial data science applications. The focus is on extracting useful and meaningful spatial patterns from geometric information. That is, applications can express how frequently two or more spatial datasets are related by using *if-then* rules, which have associated values to measure their strength and significance. Usually, the interest is in mining those rules with large associated values (i.e., greater than minimum thresholds), which are termed as *strong* rules. For instance, we can mine strong spatial association rules that show how locations of tourist attractions are related to coverage areas of accommodations and restaurants.

Unfortunately, the available approaches that extract spatial association rules [6–13] face at least one of the following problems. First, their focus is on dealing with crisp spatial objects only. Second, they are incapable of processing fuzzy topological relationships. These problems seriously limit the representation of spatial fuzziness. The last problem refers to the lack of a strategy to handle linguistic variables and values representing the fuzziness levels of objects.

In this paper, our goal is to solve the aforementioned problems by proposing a novel method for extracting spatial association rules from fuzzy spatial data. The central idea is to adequately deal with spatial fuzziness by representing spatial features of users' applications as fuzzy spatial objects and to compute fuzzy topological relationships between spatial information of users' interest (called *reference spatial dataset*) and such fuzzy spatial objects. Our method transforms the degrees returned by fuzzy topological relationships into linguistic values to intuitively express the meaning of the resulting relationships. Further, we represent concepts related to topological relationships, linguistic values, and reference spatial datasets in *hierarchies* to mine spatial association rules on multiple levels.

The main contributions of this paper are:

- It introduces a novel method that solves the problem of discovering spatial association from spatial phenomena afflicted by spatial fuzziness.

Table 1. Comparison of existing approaches with our work (last column).

Comparison Criteria	[10]	[13]	[12]	[6]	[7]	[8]	[9]	[11]	Our work
Crisp spatial datasets	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fuzzy spatial datasets						✓*		✓*	✓
Crisp topological relationships	✓	✓	✓				✓		✓
Fuzzy topological relationships						✓*			✓
Hierarchies of concepts	✓	✓	✓			✓	✓	✓	✓
Linguistic variables and values								✓	✓
Multiple level association rules	✓	✓	✓			✓	✓	✓	✓

*limited support

- It describes how to discover such rules by computing fuzzy topological relationships between a reference spatial dataset and fuzzy spatial features.
- It provides the possibility of mining multiple-level spatial association rules based on hierarchies of concepts.
- It shows the applicability of the proposed method by using a case study based on real spatial datasets.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces our case of study. Sections 4 and 5 summarize basic concepts needed to understand the proposed method. Section 6 details the architecture of our method and shows its applicability by using our case of study. Finally, Section 7 draws some conclusions and presents future work.

2 Related Work

Table 1 compares existing approaches that extract spatial association rules with our proposed approach. We check whether a given approach deals with different types of datasets and provides solutions to implementing specific concepts. We can group these existing approaches as follows: (i) approaches that deal with crisp spatial data, and (ii) approaches that apply concepts from fuzzy set theory.

With respect to the first group, we highlight the method proposed in [10]. To the best of our knowledge, it is the first method that extracts spatial association rules from crisp spatial datasets. It formally defines the concept of a spatial association rule and employs key ideas from [1], such as the use of minimum thresholds for support and confidence when extracting rules. In addition, it presents *hierarchies of concepts* for spatial datasets and topological relationships. While the hierarchy for a spatial dataset describes how spatial objects are organized based on their associated alphanumeric data, the hierarchy of topological relationships group similar relations into a coarser relation. Such hierarchies allow users to explore the extracted rules in different levels of aggregations (i.e., multiple level association rules). The approach in [13] applies similar concepts to extract spatial transactions, which are used to identify the rules. Other approaches mine such rules by using different strategies, such as

inductive logic programming [12], cell patterns [6], and Boolean matrix [7]. As depicted in Table 1, the main limitation of the aforementioned approaches is that they do not consider concepts related to spatial fuzziness.

The second group consists of approaches that incorporate fuzzy concepts to the discovery of spatial association rules. The authors in [8] specify a method to deal with fuzzy spatial data. They represent spatial fuzziness by using region objects with *broad boundaries*. This means that an areal object has three disjoint or adjacent parts: (i) exterior, which comprises all points that certainly do not belong to the object, (ii) core, which contains all points that certainly belong to the object, and (iii) broad boundary, which consists of all points possibly belong to the object. The rules are based on the computation of topological relationships for region objects with broad boundaries. As discussed in [2], this model limits the expressiveness of spatial fuzziness since it is based on the three-valued logic only. The authors in [9] allow users to specify fuzzy hierarchies in the sense that membership degrees represent to which extent an item of a hierarchy belongs to its parent, allowing that an item has multiple and different membership degrees. In [11], the authors consider the resulting membership degrees of spatial relationships (e.g., distance relations) when calculating the support and confidence of the rules. For this, they define a membership function that determines to which extent a value belong to a given linguistic value that represents a spatial relationship, such as *near*. Further, this approach deals with spatial fuzziness by mapping alphanumeric attributes associated to spatial objects into membership functions instead of using fuzzy spatial data types.

As previously discussed and shown in Table 1, we conclude that the aforementioned approaches face problems that negatively affect the task of discovering spatial association rules from spatial phenomena characterized by spatial fuzziness. On the other hand, we introduce a novel method that adequately handles and manages spatial fuzziness when extracting such rules. This method stores spatial features of an application (labeled with linguistic variables) by using fuzzy spatial objects (labeled with linguistic values). By computing fuzzy topological relationships between these features and a spatial dataset of users' interest, our method allows users to mining multiple level rules.

3 Running Example

Our goal is to discover associations between tourist attractions and characteristics of accommodations and restaurants located in New York City. Hence, our application is based on three real spatial datasets. The first one stores the tourist attractions represented by crisp region objects. To build this dataset, we have used OpenStreetMap data to extract crisp region objects inside New York City that represent tourist attractions, such as culture, leisure, and historic areas. As a result, this dataset contains 12,328 crisp region objects labeled with the category (e.g., *historic*) and type (e.g., *monument*) of the tourist attraction.

The other datasets are based on the application in [4]. The second dataset refers to the locations of Airbnb accommodations in New York City compiled

on December 4th, 2021. The characteristics of interest are the price and number of reviews of accommodations. We have excluded the lines with missing data in these attributes from the dataset. The last dataset comprises the most recent graded inspection results of restaurants in New York City provided by the Department of Health and Mental Hygiene (DOHMH). To build this dataset, we have executed the R script supplied by the DOHMH, excluded the lines with negative score and missing latitude and longitude coordinates, and guaranteed that the last inspection result occurred before the last extraction date of the Airbnb accommodations. We aim at representing the coverage area of each characteristic of accommodations and restaurants as fuzzy region objects (Section 5) to understand how such areas are associated to tourist attractions.

4 Basic Concepts of Spatial Association Rules Mining

An association rule is expressed as $A \rightarrow B (s\%, c\%)$ where A and B are *itemsets*, i.e., they are sets of items or elements that appear together in a given transaction of a database [1]. Such a rule means that if the antecedent A occurs, then the consequent B also occurs with a *support* of $s\%$ and *confidence* of $c\%$. Support and confidence measure the strength and significance of rules. Support indicates how frequently the itemsets A and B appear in the transactions. Confidence denotes the percentage of transactions containing A that also contain B . Commonly, users define minimum values for support and confidence to get relevant rules.

A spatial association rule extends the meaning of a classical association rule by including spatial relationships in the itemsets A or B [10]. The key idea is to build transactions that store spatial relationships between a spatial dataset of the user interest and other spatial datasets that represent different characteristics of the application. Commonly, the employed spatial relationships are *topological relationships* (e.g., overlap, inside) [14], which express how two or more spatial objects are related with respect to their relative position. For instance, we can have a rule that expresses the strength and significance of the overlapping situation between the coverage area of cut-rate accommodations and regions representing tourist attractions. In this paper, we consider that the spatial datasets of the applications store fuzzy spatial objects and thus, we employ fuzzy topological relationships when mining spatial association rules.

5 Fuzzy Spatial Data Handling

5.1 Fuzzy Regions and Fuzzy Topological Relationships

In this paper, we deal with fuzzy regions, which are formally defined by using concepts from fuzzy set theory [15]. Fuzzy set theory extends and generalizes Boolean set theory by allowing that an element can have partial membership in the set. Let X be the *universe*. A fuzzy set \tilde{A} uses a *membership function* $\mu_{\tilde{A}} : X \rightarrow [0, 1]$ to determine the *membership degree* of an element to \tilde{A} .

In the same way as crisp sets are extended to fuzzy sets, crisp spatial objects are generalized to fuzzy spatial objects. For a fuzzy region object \tilde{R} , this means that its geometric structure is the same as a *crisp* region object. \tilde{R} consists of a finite set of disjoint fuzzy faces, with special properties. A fuzzy face \tilde{F} is a connected, bounded, and regular closed fuzzy set in \mathbb{R}^2 with a membership function $\mu_{\tilde{F}} : \mathbb{R}^2 \rightarrow]0, 1]$ that assigns a membership degree to each point in \tilde{F} . A crisp region object can be represented by a fuzzy region object that contains points with degree 1 only. Fuzzy regions are formally defined in [2].

A fuzzy region object is labeled with a *linguistic value* to characterize a specific instance of a spatial feature, which is represented by a *linguistic variable*. Commonly, linguistic variables are denoted by substantives, while linguistic values are denoted by adjectives. Our running example has three (fuzzy) spatial features denoted by the following linguistic variables and their corresponding linguistic values (in parentheses): *accommodation price* (*cut-rate, cheap, affordable, expensive, premium*), *accommodation notability* (*unknown, little-known, well-known, famous*), and *food safety* (*very low, low, medium, high, very high*).

Different types of fuzzy spatial operations have been defined to handle fuzzy region objects [3]. We are interested in applying *fuzzy topological relationships* [2] on fuzzy regions to discover spatial association rules. Differently from a classical topological relationship that yields a Boolean value, a fuzzy topological relationship yields a membership degree in $[0, 1]$ that expresses to which extent a relative position between two fuzzy region objects holds. For instance, we can compute the *overlapping degree* of two fuzzy region objects. The membership degree returned by a fuzzy topological relationship can be mapped to a *high-level linguistic value*, which provides the semantics of the relationship to the user. For instance, two fuzzy region objects can *quite* overlap.

5.2 Spatial Plateau Algebra and its Implementation

The *Spatial Plateau Algebra* (SPA) [5] is a *executable type system* since it provides data structures for fuzzy spatial data types and specifications for fuzzy spatial operations. The SPA represents fuzzy spatial data types as *spatial plateau data types* where a spatial plateau object can be a *plateau point*, *plateau line*, or *plateau region*. A plateau region object consists of a list of pairs $\langle (r_1, m_1), \dots, (r_n, m_n) \rangle$ where r_i is a crisp region object annotated with the membership degree $m_i \in]0, 1]$ with $i \leq n$ for some $n \in \mathbb{N}$. The crisp region objects of all pairs must have different membership degrees and be disjoint or adjacent to each other. Further, the SPA specifies fuzzy spatial operations and fuzzy topological relationships as *spatial plateau operations* and *spatial plateau topological relationships*. The SPA's operations are specified by using well-defined concepts from crisp spatial algebras that are implemented by existing spatial libraries (e.g., GEOS).

The implementation of the SPA is given by the R package *fsr* [4]. This package implements all data types and operations specified by the SPA, including a two-stage method for building plateau regions from point datasets. We employ this method to create the plateau region objects that characterize the linguistic values of each linguistic variable of our running example. Figure 1 shows the

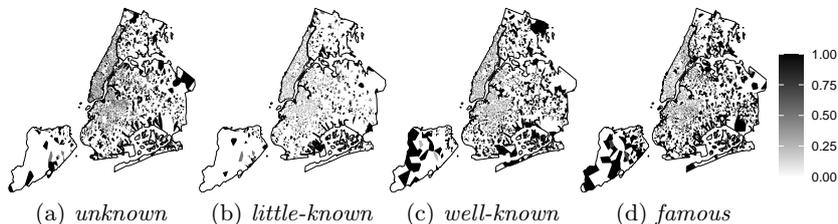


Fig. 1. The plateau region objects for each linguistic value of the fuzzy spatial feature *accommodation notability* including New York city boundaries.

built plateau region objects for the linguistic variable *accommodation notability*. In this figure, each point has a membership degree that indicates to which extent the point belongs to a specific linguistic value (as indicated by the side bar).

6 Discovery of Spatial Association Rules from Fuzzy Spatial Objects

6.1 Architectural Overview

Figure 2 shows the architecture of our method, which extracts spatial association rules from fuzzy spatial datasets. The discovery is guided by the user’s interest, as indicated by a set of parameters (Section 6.2). They determine which subsets of the available datasets will be used in the extraction process, specify a set of hierarchies to describe how spatial relationships and the data itself are semantically organized, and set the minimum thresholds that rules must satisfy.

The layers of our architecture are: (i) Spatial Data Layer, (ii) Spatial Data Handling Layer, (iii) Itemsets Handling Layer, and (iv) Spatial Association Rules Retrieval Layer. As indicated in Figure 2, each layer applies some specific user parameters to perform its processing. In summary, the Spatial Data Layer provides the datasets to be used in the extraction process. The Spatial Data Handling Layer fetches the relevant data from the Spatial Data Handling Layer, according to the user parameters, and computes fuzzy spatial relationships. The goal of the Itemsets Handling Layer is twofold: (i) to organize the processed relationships into itemsets and filter them with respect to the minimum support indicated by the user, and (ii) to discover the spatial association rules by using a mining algorithm, such as the apriori [1]. Then, the Spatial Association Rules Retrieval Layer is responsible for interactively presenting the mined rules since they can be organized in multiple levels according to the definition of hierarchies. The layers of our architecture are detailed in Sections 6.3 to 6.6, including examples in the context of our running example.

6.2 User Parameters

The parameters provided by the user play an important role in our method. As shown in Figure 2, they consist of (i) Query Parameters, (ii) Hierarchy of

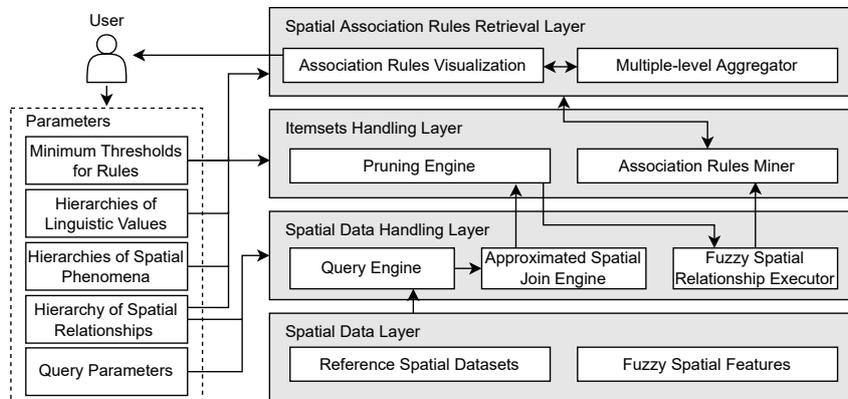


Fig. 2. The architecture of our method.

Spatial Relationships, (iii) Hierarchies of Spatial Phenomena, (iv) Hierarchies of Linguistic Values, and (v) Minimum Thresholds for Rules. The purpose of Query Parameters is to indicate (i) the relevant Reference Spatial Dataset of the application, (ii) the Fuzzy Spatial Features that will be considered in the mining process, (iii) the fuzzy topological relationship that should be employed when associating spatial data, and (iv) conditions applied to alphanumeric attributes that are associated to the chosen Reference Spatial Dataset. In our running example (Section 3), we are interested in finding spatial association rules based on the overlapping degree between tourist attractions and the fuzzy spatial features that represent coverage areas of distinct characteristics of accommodations and restaurants. For this, our Query Parameters are: (i) tourist attractions as the Reference Spatial Dataset, (ii) three Fuzzy Spatial Features containing fuzzy regions implemented as plateau region objects (Section 5), and (iii) fuzzy overlap as the fuzzy topological relationship.

The goal of defining hierarchies (parameters (ii)-(iv)) is to semantically organize concepts related to spatial relationships and characteristics of spatial data. Each hierarchy enables at least one of the following processes: (i) discovery and visualization of spatial association rules mining at distinct levels, (ii) pruning of fuzzy spatial features, and (iii) computation of approximated topological relationships. A hierarchy allows users to define how lower-level concepts are related to higher-level concepts. Given two attributes A, B where each attribute has its own domain of values (or concepts). A and B are related in a hierarchy if we can determine a value of B by aggregating a subset of values in A . In this case, we have $A \rightarrow B$, expressing that these attributes are levels of a hierarchy.

The Hierarchies of Spatial Phenomena are defined by using alphanumeric attributes that label crisp spatial objects stored in the Reference Spatial Datasets. It allow users to mine and visualize spatial association rules in different levels (as illustrated in Section 6.6). For instance, users can first identify rules containing itemsets of a more general level in a hierarchy. Then, they can

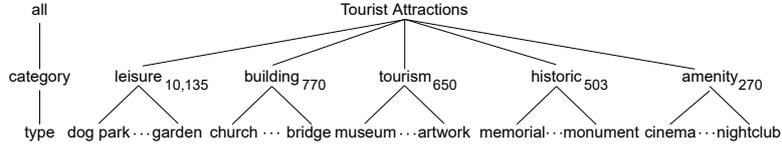


Fig. 3. The Hierarchy of Spatial Phenomena for our running example.

mine rules on a more detailed level from a particular value of the previous level. Figure 3 depicts the Hierarchy of Spatial Phenomena built on the alphanumeric attributes of our Reference Spatial Dataset. It is defined by $type \rightarrow category$. Note that the upper level of this hierarchy (named *all*) aggregates all values from *category* and as a result, consists of all touristic attractions. Further, this figure shows the number of members for each distinct value of *category*.

The Hierarchy of Spatial Relationships defines how spatial relationships are organized into levels of detail. Such an organization serves as reference for computing relationships in the Spatial Data Handling Layer (Section 6.4). It follows similar principles as the method in [10] in the sense that higher levels of the hierarchy represent coarser relations. The main difference is that we are dealing with fuzzy topological relationships in the bottom level. For instance, *fuzzy overlap* and *fuzzy meet* can be grouped to a coarser relationship named *intersect*, which can be computed by using spatial approximations.

A Hierarchy of Linguistic Values describes how linguistic values of a Fuzzy Spatial Features are structured. This kind of hierarchy allows users to explore the discovered patterns at distinct grouping levels of linguistic values. For instance, we can specify a Hierarchy of Linguistic Values for the Fuzzy Spatial Feature representing the *accommodation notability* as $lvalues \rightarrow class$ where $lvalues = \{unknown, little-known, well-known, famous\}$ and $class = \{unpopular, popular\}$. By grouping the values *unknown* and *little-known*, we obtain the class *unpopular*, whereas the aggregation of *well-known* and *famous* lead to the class *popular*.

Finally, the Minimum Thresholds for Rules determine the desired levels of strength and significance that the mined rules must satisfy. For instance, minimum values for support and confidence. The user can provide different minimum thresholds for each level of the hierarchy since the frequency of itemsets intrinsically decreases as we descend the levels of the hierarchy. In our running example and without loss of generality, we extract association rules by varying the levels in the Hierarchy of Spatial Phenomena only. For the level *category*, we use the support of 10% and a minimum confidence of 30%. In the next level, we employ 5% and 20%, respectively.

6.3 Spatial Data Layer

The Spatial Data Layer is composed of two components: (i) Reference Spatial Datasets, and (ii) Fuzzy Spatial Features. The goal of the mining process is to discover associations between objects of a Spatial Reference Dataset and objects

of Fuzzy Spatial Features. Hence, the first component consists of a collection of crisp spatial objects associated with alphanumerical attribute values that represent the users' interest. Let $n, m \in \mathbb{N}$. Let further $C = \{c_1, \dots, c_n\}$ be a set of alphanumerical attributes where each $c \in C$ has a domain of values cv_1, \dots, cv_m , i.e., $dom(c) = \{cv_1, \dots, cv_m\}$. We define a Reference Spatial Dataset as a pair $\langle N, S \rangle$ where N provides its name and S is a set of tuples (s, v_1, \dots, v_n) such that $s \in \{point, line, region\}$ and $v_i \in dom(c_i)$ with $i \leq n$. By using the Query Parameters, the user picks one Reference Spatial Dataset that will be employed in the extraction of spatial association rules.

The second component refers to Fuzzy Spatial Features, where each feature is a fuzzy spatial dataset and represents a particular characteristic of the application. We define a Fuzzy Spatial Feature as a pair $\langle L, F \rangle$ such that (i) L is a linguistic variable with a domain of linguistic values l_1, \dots, l_k for some $k \in \mathbb{N}$, and (ii) F is a set of tuples (f, l) where $f \in \{fpoint, fline, fregion\}$ and $l \in dom(L)$. Hence, a Fuzzy Spatial Feature is labeled with a linguistic variable and consists of one or more fuzzy spatial objects annotated by a linguistic value.

Recall that, as discussed in Section 6.2, the attributes in a Reference Spatial Dataset and linguistic values of the Fuzzy Spatial Features can be organized in hierarchical structures to represent different aggregation levels of concepts. Further, the user can also apply filters to the spatial datasets aiming to get a subset of tuples that will be used in the extraction process.

In our running example, we have one Reference Spatial Dataset that stores tourist attractions in New York City. In this case, $C = \{category, type\}$ and our Reference Spatial Dataset is given by $\langle Tourist Attractions, \{(s_1, leisure, dog\ park), \dots, (s_{12,328}, tourism, museum)\} \rangle$ with $s_1, \dots, s_{12,328} \in region$. Our Fuzzy Spatial Features represent the linguistic variables *food safety*, *accommodation notability*, and *accommodation price*. For instance, Figure 1 shows the following Fuzzy Spatial Feature: $\langle Accommodation Notability, \{(f_1, unknown), (f_2, little-known), (f_3, well-known), (f_4, famous)\} \rangle$ where f_1, \dots, f_4 are fuzzy regions implemented as plateau region objects respectively shown in Figure 1a to d.

6.4 Spatial Data Handling Layer

This layer is responsible for handling the data loaded from the Spatial Data Layer and computing topological relationships. For this, it employs the following interacting components: (i) Query Engine, (ii) Approximated Spatial Join Engine, and (iii) Fuzzy Spatial Relationship Executor. The Query Engine captures the relevant subset of spatial objects from the Spatial Data Layer according to the Query Parameters provided by the user (Section 6.2).

Next, the Approximated Spatial Join Executor computes *approximated topological relationships* between the captured spatial objects from the Reference Spatial Dataset and the Fuzzy Spatial Features. The computation of an approximated relationship is based on *minimum bounding rectangles* (MBRs). For this, we employ a function named *approx.rel* that yields a Boolean value to indicate whether the MBR of a crisp spatial object satisfies a given relationship (e.g., *intersect*) with respect to the crisp MBR of a fuzzy spatial object. Such a result

can be a false positive since an MBR of a spatial object may include points that do not belong to the object. However, the use of MBRs allows us to quickly discover cases where two spatial objects are certainly not related and thus, identify those cases where an association does not exist. The approximated relationship is chosen based on the Hierarchy of Spatial Relationship provided by the user. For instance, the parent level of the fuzzy topological relationship *overlap* is the relationship *intersect*, which is a coarser relationship to be computed by using MBRs. Let RSD be a set of tuples containing the spatial objects of interest of a Reference Spatial Dataset and the pair $\langle L, F \rangle$ be a Fuzzy Spatial Feature with n tuples in F . Let also ext be a function that extracts the i -th fuzzy spatial object of F . The Approximated Spatial Join Executor computes the function $approx_rel$ for each crisp spatial object in RSD with respect to each fuzzy spatial object in FSF . Hence, it yields a set of tuples (s, a_1, \dots, a_n) where s is a crisp spatial object belonging to RSD and $a_1, \dots, a_n \in bool$ such that each element a_i is given by $approx_rel(s, ext(F, i))$ with $1 \leq i \leq n$. The attribute name of a_i is a concatenation of the linguistic value of the $ext(F, i)$ and L . We compute this procedure for each Fuzzy Spatial Feature. In the end, we have a collection of sets where each set contains the results of how RSD is roughly related to each Fuzzy Spatial Feature. Since these sets share a common attribute (i.e., elements of the Reference Spatial Dataset), we join them to build a unique set of tuples. This set is then sent to the Pruning Engine of the Itemsets Handling Layer (Section 6.5).

Finally, the Fuzzy Spatial Relationship Executor computes fuzzy topological relationships between the crisp spatial objects of the Reference Spatial Dataset and the fuzzy spatial objects selected by the Pruning Engine. The main advantage of this is that number of required computations to process the costly relationships can be decreased since the Pruning Engine might identify some fuzzy spatial object that do not have strong associations with the Reference Spatial Dataset. As discussed in Section 5.1, a fuzzy topological relationship yields a membership degree in $[0, 1]$ that indicates to which extent the given relationship occurs. Such a degree is then transformed into a linguistic value. Let ftr be a fuzzy topological relationship that returns a linguistic value in LT (e.g., as given in [2]). Let further $n' \in \mathbb{N}$ be the number of fuzzy spatial objects in a Fuzzy Spatial Feature $\langle L, F \rangle$ after processing the Pruning Engine. We obtain a set of tuples $(s, v_1, \dots, v_m, tl_1, \dots, tl_{n'})$ where s is a crisp spatial object annotated with a set of alphanumeric attributes v_1, \dots, v_m , which are members of RSD , and $tl_1, \dots, tl_{n'}$ are elements of LT such that tl_i is given by $ftr(s, ext(F, i))$ with $1 \leq i \leq n'$. Similarly to the set of tuples returned by the Approximated Spatial Join Executor, the attribute name of tl_i is a combination of the linguistic value of the $ext(F, i)$ with L . As a result, we have a collection of sets of tuples that are joined and sent to the Itemset Handling Layer.

Due to space constraints, we show only one example of the result obtained by the Fuzzy Spatial Relationship Executor for our application. To compute the fuzzy overlap on spatial objects, we employ the fsr [4] with $LT = \{a\ little\ bit, somewhat, slightly, averagely, mostly, quite\}$. The set of tuples with the overlapping results between our Reference Spatial Dataset and the Fuzzy Spatial Feature

representing the linguistic variable *accommodation notability* (Section 6.3) is defined as $\{(s_1, \textit{leisure}, \textit{dog park}, \textit{quite}, \textit{a little bit}, \textit{a little bit}), \dots, (s_{12,328}, \textit{tourism}, \textit{museum}, \textit{a little bit}, \textit{mostly}, \textit{slightly})\}$. The attribute names respectively consists of *region_obj*, *category*, *type*, *unknown accomm. notability*, *well-known accomm. notability*, and *famous accomm. notability* (*accom.* stands for *accommodation*). Note that the linguistic value *little-known* is not included, which means that the fuzzy spatial object representing this particular situation was not selected by the Pruning Engine. This set is then joined to the results for other Fuzzy Spatial Features. The resulting set of tuples is effectively used to extract the spatial association rules in our running example.

6.5 Itemsets Handling Layer

This layer is responsible for two key actions: (i) to perform pruning operations, and (ii) to mine the spatial association rules. The pruning operation takes place based on the set of tuples given by the Approximated Spatial Join Engine, named *AT*, and the minimum support for the highest level of the Spatial Phenomena Hierarchy. The goal is to identify the fuzzy spatial objects that are not frequently spatially related to the crisp spatial objects of the Reference Spatial Dataset. Then, these fuzzy spatial objects are not used in the next step (i.e., the Fuzzy Spatial Relationship Executor). To accomplish this goal, the Pruning Engine first calculates the ratio of the number of times that the value *true* appears in each Boolean attribute of *AT* and the total number of tuples in *AT*. Then, it excludes all fuzzy spatial objects whose corresponding ratio values are lesser than a given minimum support. The non-excluded fuzzy spatial objects are sent to the Spatial Data Handling Layer.

The Association Rules Miner extracts rules by using the set of tuples returned by the Fuzzy Spatial Relationship Executor. This set is reshaped as a transactional dataset, which is given as input to the Apriori algorithm [1]. For each level of each hierarchy (e.g., the Hierarchy of Spatial Phenomena), we find itemsets with a frequency of appearance higher than the minimum support at that particular level. These itemset are then frequent itemsets that will be used to form association rules. The strong rules (e.g., with confidence greater than the minimum threshold) are sent to the Spatial Association Rules Retrieval Layer.

In our running example, we employ the R package *arules*⁴ to identify the frequent itemsets and association rules from the execution of the Apriori algorithm. An example of frequent itemset is $\{\textit{tourist attraction}_{\textit{category}=\textit{leisure}}\}$ with support of 82.21%. This itemset refers to the number of transactions where the attribute *category* is equal to *leisure* in the Reference Spatial Dataset. Table 2 depicts examples of spatial association rules for our running example.

6.6 Spatial Association Rules Retrieval Layer

The purpose of this layer is to interactively present the knowledge discovered to the user. It consists of two components: (i) Association Rules Visualization,

⁴ <https://cran.r-project.org/package=arules>

Table 2. Some examples of rules for our running example.

Antecedent	Consequent	Support Confidence	
		(%)	(%)
$\{tour. att. category=leisure\}$	$\{overlap_{quite}, accom. notability_{unknown}\}$	27.12	32.98
$\{tour. att. category=leisure\} \wedge \{overlap_{quite}, food\ safety_{high}\}$	$\{overlap_{quite}, accom. price_{affordable}\}$	11.86	39.28
$\{tour. att. category=leisure\} \wedge \{overlap_{quite}, accom. price_{cheap}\}$	$\{overlap_{quite}, accom. notability_{famous}\}$	10.46	48.92
$\{tour. att. type=pitch\}$	$\{overlap_{quite}, accom. price_{affordable}\}$	12.89	37.62
$\{tour. att. type=pitch\}$	$\{overlap_{quite}, food\ safety_{high}\}$	12.34	36.02
$\{overlap_{quite}, accom. notability_{well-known}\}$	$\{tour. att. type=pitch\}$	10.12	30.61

tour. att. and *accom.* stand for *tourist attraction* and *accommodation*, respectively.

and (ii) Multi-level Aggregator. The first one is a user interface that enables the general visualization and exploration of the strong spatial association rules extracted by the Association Rules Miner. The rules can be visualized in distinct formats, such as tables, graphs, scatter plots, and parallel coordinates. Table 2 shows six rules for our running example by using a formal tabular format inspired by the notation given in [13]. We also employed this component to get only rules that include attributes related to tourist attractions either on the antecedent or on the consequent.

The Multi-level Aggregator allows users to identify rules in different levels of a hierarchy of concepts. This is mainly performed by selecting those rules that contain itemsets with the desired members of a given level. We employed this component in our running example to mine rules in each level of our Hierarchy of Spatial Phenomena (Figure 3) by using the parameters given in Section 6.2. The first three rules in Table 2 are related to the highest level of this hierarchy (i.e., *category*), while the other rules refer to the lowest level (i.e., *type*). For instance, the first rule states that spatial objects representing *leisure* tourist attractions *quite* overlap coverage areas of accommodations characterized by *unknown notability*. This rule has support of 27.12% and confidence of 32.98%. By descending the hierarchy, the fourth rule shows that tourist attractions of type *pitch* quite overlap *affordable accommodations* with support of 12.89% and confidence of 37.62%. These results show that our method can effectively and adequately correlate distinct crisp and fuzzy spatial datasets by mining association rules according to minimum thresholds of strength and significance.

7 Conclusions and Future Work

In this paper, we have presented a novel method for extracting spatial association rules from fuzzy spatial data. We adequately deal with spatial fuzziness since we

represent spatial features of an application as fuzzy spatial objects. The extracted spatial association rules are based on linguistic values that express the meaning of fuzzy topological relationships between such fuzzy spatial objects and crisp spatial objects representing the users' interest. We have shown the applicability of our method by using a case study based on real spatial datasets.

Future work will deal with two main topics. First, we aim to study automatic strategies to define hierarchies of concepts. Hence, we would not need such types of parameters from the user. Second, we plan to conduct experimental evaluations to characterize the runtime of our method.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: ACM SIGMOD Int. Conf. on Management of Data. pp. 207–216 (1993)
2. Carniel, A.C., Schneider, M.: A conceptual model of fuzzy topological relationships for fuzzy regions. In: IEEE Int. Conf. on Fuzzy Systems. pp. 2271–2278 (2016)
3. Carniel, A.C., Schneider, M.: A survey of fuzzy approaches in spatial data science. In: IEEE Int. Conf. on Fuzzy Systems. pp. 1–6 (2021)
4. Carniel, A.C., Galdino, F., Philippsen, J.S., Schneider, M.: Handling fuzzy spatial data in R using the *fsr* package. In: ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems. pp. 526–535 (2021)
5. Carniel, A.C., Schneider, M.: Spatial Plateau Algebra: An executable type system for fuzzy spatial data types. In: IEEE Int. Conf. on Fuzzy Systems. pp. 1–8 (2018)
6. Chen, J., Li, P., Fei, H., Wang, R.: An algorithm about spatial association rule mining based on cell pattern. In: Geoinformatics: Geospatial Information Science. pp. 662–671 (2006)
7. Chen, J., Lin, G., Yang, Z.: Extracting spatial association rules from the maximum frequent itemsets based on boolean matrix. In: Int. Conf. on Geoinformatics. pp. 1–5 (2011)
8. Clementini, E., Felice, P.D., Koperski, K.: Mining multiple-level spatial association rules for objects with a broad boundary. *Data & Knowledge Engineering* **34**(3), 251–270 (2000)
9. Kacar, E., Cicekli, N.K.: Discovering fuzzy spatial association rules. In: *Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV*. pp. 94–102 (2002)
10. Koperski, K., Han, J.: Discovery of spatial association rules in geographic information databases. In: *Int. Symp. on Spatial Databases*. pp. 47–66 (1995)
11. Ladner, R., Petry, F.E., Cobb, M.A.: Fuzzy set approaches to spatial data mining of association rules. *Transactions in GIS* **7**(1), 123–138 (2003)
12. Malerba, D., Lisi, F.A.: An ILP method for spatial association rule mining. In: *First Workshop on Multi-Relational Data Mining*. pp. 18–29 (2001)
13. Rinzivillo, S., Turini, F.: Extracting spatial association rules from spatial transactions. In: *ACM Int. Workshop on Geographic Information Systems*. pp. 79–86 (2005)
14. Schneider, M., Behr, T.: Topological relationships between complex spatial objects. *ACM Trans. on Database Systems* **31**(1), 39–81 (2006)
15. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8**(3), 338–353 (1965)