UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

HENRIQUE PIGOZZO DA SILVA

DESCOBERTA DE REGRAS DE ASSOCIAÇÃO ESPACIAIS A PARTIR DE DADOS ESPACIAIS FUZZY

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

HENRIQUE PIGOZZO DA SILVA

DESCOBERTA DE REGRAS DE ASSOCIAÇÃO ESPACIAIS A PARTIR DE DADOS ESPACIAIS FUZZY

DISCOVERY OF SPATIAL ASSOATION RULES FROM FUZZY SPATIAL DATA

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Anderson Chaves Carniel

DOIS VIZINHOS
2022

HENRIQUE PIGOZZO DA SILVA

DESCOBERTA DE REGRAS DE ASSOCIAÇÃO ESPACIAIS A PARTIR DE DADOS ESPACIAIS FUZZY

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 04/agosto/2022

Anderson Chaves Carniel
Doutorado
Universidade Federal de São Carlos

Ives Renê Venturini Pola
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

Jefferson Tales Oliva
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

DOIS VIZINHOS
2022

Ao meu pai, Carlos.

AGRADECIMENTOS

Agradeço ao Prof. Anderson Carniel pela paciência, dedicação, confiança e parceria.

Aos colegas Thiago Felix e Pedro Venâncio pela fundamental contribuição para o trabalho.

Aos professores, tutores, à coordenação e aos demais envolvidos no curso da Especialização pela excelência do trabalho desenvolvido.

Por fim, à minha família, sem eles este trabalho não seria possível. Em particular à Gabriela, que esteve ao meu lado em todos os momentos.

RESUMO

A descoberta de regras de associação espacial é uma tarefa central em projetos de ciência de dados espaciais e se concentra na extração de padrões e relacionamentos espaciais úteis e significativos a partir de informações espaciais e geométricas. Muitos fenômenos espaciais foram modelados e representados por objetos espaciais difusos, que possuem interiores desfocados, limites incertos e/ou localizações inexatas. Neste artigo, apresentamos um novo método para extrair regras de associação espacial a partir de dados espaciais difusos. Ao permitir que os usuários representem características espaciais de suas aplicações como objetos espaciais difusos e ao empregar relações topológicas difusas, nosso método descobre padrões de associação espacial entre objetos espaciais de interesse dos usuários (por exemplo, atrações turísticas) e tais características espaciais difusas (por exemplo, condições sanitárias). de restaurantes, número de avaliações e preço das acomodações). Além disso, este artigo apresenta um estudo de caso baseado em conjuntos de dados reais que mostra a aplicabilidade do nosso método.

Palavras-chave: Ciência de dados espaciais. Regra de associação espacial. Vagueza espacial. Dados espaciais difusos. Relacionamento topológico difuso

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/or 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

Discovery of Spatial Association Rules from Fuzzy Spatial Data

Henrique P. da Silva[1], Thiago D. R. Felix[2], Pedro V. A. B. de Venâncio[3][0000−0003−4665−562X], and Anderson C. Carniel[2][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
pedrovinicius@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 alphanumerical 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 alphanumerical 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 $\mu_{\tilde{F}} : \mathbb{R}^2 \to]0, 1]$ 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 $\langle (r_1, m_1), \ldots, (r_n, m_n) \rangle$ 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

(a) *unknown* (b) *little-known* (c) *well-known* (d) *famous*

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

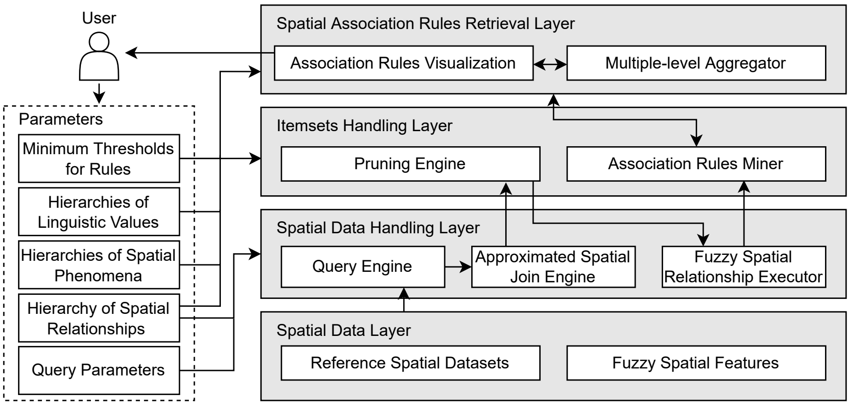

Fig. 2. The architecture of our method.

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 alphanumerical 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 $A \to B$, expressing that these attributes are levels of a hierarchy.

The Hierarchies of Spatial Phenomena are defined by using alphanumerical 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

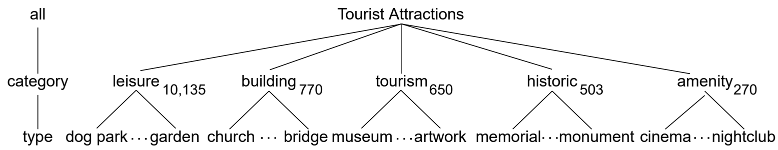

Fig. 3. The Hierarchy of Spatial Phenomena for our running example.

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 alphanumerical 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 $C = \{c_1, ..., c_n\}$ be a set of alphanumerical attributes where each $c \in C$ has a domain of values $cv_1, ..., cv_m$, i.e., $dom(c) = \{cv_1, ..., cv_m\}$. 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 $(s, v_1, ..., v_n)$ 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 $l_1, ..., l_k$ 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*, $\{(s_1, leisure, dog\ park)$, ..., $(s_{12,328}, tourism, museum)\}\rangle$ with $s_1, ..., s_{12,328} \in region$. 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*, $\{(f_1, unknown), (f_2, little\text{-}known), (f_3, well\text{-}known), (f_4, famous)\}\rangle$ where $f_1, ..., f_4$ 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 $(s, a_1, ..., a_n)$ where s is a crisp spatial object belonging to RSD and $a_1, ..., a_n \in bool$ 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 $(s, v_1, ..., v_m, tl_1, ..., tl_{n'})$ where s is a crisp spatial object annotated with a set of alphanumerical attributes $v_1, ..., v_m$, which are members of RSD, and $tl_1, ..., tl_{n'}$ 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 $\{(s_1$, *leisure*, *dog park*, *quite*, *a little bit*, *a little bit*), ..., $(s_{12,328}$, *tourism*, *museum*, *a little bit*, *mostly*, *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 minim 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*$_{well\text{-}known}\}$	$\{$*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)