

Received December 17, 2019, accepted January 24, 2020, date of publication February 21, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975584

Spatial Structured Prediction Models: Applications, Challenges, and Techniques

ZHE JIANG^{ID}, (Member, IEEE)

Department of Computer Science, The University of Alabama, Tuscaloosa, AL 35487, USA

e-mail: zjiang@cs.ua.edu

This material was based upon work supported by the National Science Foundation under Grants No. 1850546, the University Corporation for Atmospheric Research (UCAR) under Contract SUBAWD000837, and Camgian Microsystems.

ABSTRACT Spatial structure patterns are prevalent in many real-world data and applications. For example, in biochemistry, the geometric topology of a molecular surface indicates protein functions; in hydrology, irregular geographic terrains and topography on the Earth's surface control water flows and distributions; in civil engineering, wetland parcels in remote sensing imagery are often made up of contiguous patches. Spatial structured prediction aims to learn a prediction model whose input and output data contain a spatial structure. Modeling spatial structural information in prediction models is critical for interdisciplinary applications due to two reasons. First, explicit spatial structural information often indicates the underlying physical process, and thus enhances model interpretability. Second, spatial structural constraints also have positive side-effects of enhancing model robustness against noise and obstacles and regularizing model learning when training labels are limited. However, spatial structured prediction also poses several unique challenges, such as the existence of implicit spatial structure in continuous space, structural complexity in geometric topology, and high computational costs. Over the years, various techniques have been proposed for spatial structured prediction in different applications. This paper aims to provide an overview of the spatial structured prediction problem. We provide a taxonomy of techniques based on the underlying approaches. We also discuss several future research directions. The paper can potentially not only help interdisciplinary researchers find relevant techniques but also help machine learning researchers identify new research opportunities.

INDEX TERMS Machine learning, structured prediction, spatial structured model, geometric deep learning, interdisciplinary applications.

I. INTRODUCTION

Structured prediction (also called structured learning, or structured output learning) aims to learn a predictive model whose input or output data contain a structure between samples [1], [2]. Structural prediction is important in many real-world applications. In natural language processing, words in a sentence often follow a syntactic structure called a parse tree. In biology, a protein consists of amino acids in a network structure, and an amino acid consists of atoms such as carbon, nitrogen, and oxygen again in a network structure. In speech recognition, vocals follow a sequential structure with a current element correlated with its precedents. The existence of structure is non-trivial in supervised learning because learning samples are no longer independent and identically distributed (i.i.d.). The i.i.d. assumption is prevalent

in many supervised learning methods such as decision tree, random forest, maximum likelihood classifiers. The assumption simplifies model representation and learning algorithms since an identical model can be used to predict each sample independently. Removing the i.i.d. assumption dramatically increases the complexity of models and learning algorithms.

Spatial structured prediction, as its name suggests, aims to learn a prediction model for which input or output samples follow a spatial structure [3]–[6]. Such a prediction model is also called a spatial structured (prediction) model. For example, in biochemistry, the geometric topology of a molecular surface indicates protein functions; in hydrology, irregular geographic terrains and topography on the Earth's surface control water flows and distributions; in civil engineering, wetland parcels in remote sensing imagery are often made up of contiguous patches.

Modeling spatial structure is important because the structural information is often closely related to the

The associate editor coordinating the review of this manuscript and approving it for publication was Berdakh Abibullaev^{ID}.

underlying physical process. It provides an opportunity to bridge machine learning models with existing theories or knowledge in a scientific discipline. For example, hydrologists have long studied the distribution and flow of water on the Earth's surface based on geographic terrain and topography, and have developed various simulation models. Incorporating such geographic terrain structure into machine learning models helps enhance interpretability of the black-box models. From this perspective, structured prediction is intrinsically interdisciplinary, involving the convergence of multiple disciplines. It represents an intersection between data-driven machine learning models in computer science and physics-driven models in an application discipline.

However, learning spatial structured models is non-trivial due to several unique challenges. First, the spatial dependency structure is often implicit in continuous space. Second, the dependency structure can be complex both in Euclidean space and in geometric topology space. Third, spatial structural patterns can be further complicated by additional spatial effects such as spatial heterogeneity (local effects) and multi-scale effects. Fourth, learning spatial structured models is computationally challenging due to model structural complexity. Finally, the problem can be challenging when training labels are limited.

Over the years, various techniques have been developed for different applications. Existing surveys on relevant topics such as spatial prediction [3] often focus on the general challenges of spatial data (e.g., spatial autocorrelation, heterogeneity, multi-scale hierarchy, etc.) without a systematic review of methods from the structured prediction perspective, particularly geometric topology structures. This paper fills the gap with a systematic review of different types of spatial structures. We provide a taxonomy of techniques based on the underlying approaches (Section V), including spatial contextual feature generation, spatial structured model representation, and deep neural networks. Specifically, within structured model representation, we further group the methods based on the types of spatial structures, including spatial neighborhood graph, spatial distance kernel, and geometric topology. Finally, we discuss several future research directions, including geometric deep learning, enhancing model transparency and interpretability, and scalable inference (Section VI). The paper can potentially not only help interdisciplinary researchers find relevant techniques but also help machine learning researchers identify new research opportunities.

II. PROBLEM STATEMENT

In order to formally define the spatial structured prediction problem, we need to first introduce the concept of spatial learning sample. In traditional prediction problems, input data is often viewed as a collection of sample records. Similarly, in spatial structured prediction, input spatial data can be viewed as a collection of spatial data samples, whereby each sample corresponds to a spatial object (e.g., point, line

TABLE 1. Math symbols and descriptions.

Symbol	Domain	Description
n	\mathbb{N}	The number of sample locations
\mathbf{s}	$\mathbb{R}^{2 \times 1}$	Sample location coordinates
$\mathbf{x}(\mathbf{s})$	$\mathbb{R}^{m \times 1}$	m -dimensional feature vector of a sample at location \mathbf{s}
$y(\mathbf{s})$	\mathbb{R} or \mathcal{C}	Continuous or categorical response of a sample at location \mathbf{s}
$\hat{y}(\mathbf{s})$	\mathbb{R} or \mathcal{C}	Predicted response of sample at location \mathbf{s} , \mathcal{C} is the set of class categories
\mathbf{X}	$\mathbb{R}^{n \times m}$	Feature matrix of n samples
\mathbf{Y}	$\mathbb{R}^{n \times 1}$ or $\mathcal{C}^{n \times 1}$	Response vector of n samples
$\hat{\mathbf{Y}}$	$\mathbb{R}^{n \times 1}$ or $\mathcal{C}^{n \times 1}$	Predicted response vector
\mathbf{W}	$\mathbb{R}_{0+}^{n \times n}$	W-matrix
W_{ij}	\mathbb{R}_{0+}	An element of W-matrix
$C(\mathbf{h})$	$\mathbb{R}^{2 \times 1} \mapsto \mathbb{R}_{0+}$	Covariogram function
ρ	\mathbb{R}_{0+}	Weight of spatial effect
μ	$\mathbb{R}^{m \times 1}$	Mean feature vector
Σ	$\mathbb{R}^{m \times m}$	Covariance matrix of features
β	$\mathbb{R}^{m \times 1}$	Coefficient vector
$\epsilon(\mathbf{s}_i)$	\mathbb{R}	Noise (residual error) of a sample
ϵ	$\mathbb{R}^{n \times 1}$	Noise (residual error) of n samples

or polygon), or a cell in a raster grid. A spatial data sample has multiple non-spatial attributes, one of which is the target response variable to be predicted and the others are explanatory features. Additionally, a spatial data sample also has location information and spatial attributes (e.g., distance to another point, length of a line, area of a polygon). These additional information distinguish spatial data samples out from traditional data samples in two important ways: first, the location information and corresponding spatial attributes can provide additional contextual information in the explanatory feature list; second and more important, implicit spatial structure exists based on sample locations, violating the common i.i.d. assumption. For example, in ground sensor observations on soil properties, a sample corresponds to information from one geo-located sensor. Explanatory features can include soil texture, nutrient level, and moisture. The response variable can be whether a type of plant can grow or not at the location.

Formally, spatial data is a set of spatial data samples $\{(\mathbf{x}(\mathbf{s}_i), y(\mathbf{s}_i)) | i \in \mathbb{N}, 1 \leq i \leq n\}$, where n is the total number of samples, \mathbf{s}_i is a 2 by 1 spatial coordinate vector for the i th sample, $\mathbf{x}(\mathbf{s}_i)$ is a m by 1 explanatory feature vector (m is the feature dimension), and $y(\mathbf{s}_i)$ is a scalar response (it is categorical for classification, and continuous for regression). The set of spatial samples can also be written in the matrix format, (\mathbf{X}, \mathbf{Y}) , where $\mathbf{X} = [\mathbf{x}(\mathbf{s}_1), \mathbf{x}(\mathbf{s}_2), \dots, \mathbf{x}(\mathbf{s}_n)]^T$ is a n by m feature matrix, and $\mathbf{Y} = [y(\mathbf{s}_1), y(\mathbf{s}_2), \dots, y(\mathbf{s}_n)]^T$ is a n by 1 response vector. A list of all symbols and descriptions are summarized in Table 1.

Given a set of spatial data samples with explanatory features $\mathbf{X} = [\mathbf{x}(\mathbf{s}_1), \mathbf{x}(\mathbf{s}_2), \dots, \mathbf{x}(\mathbf{s}_n)]^T$ and responses $\mathbf{Y} = [y(\mathbf{s}_1), y(\mathbf{s}_2), \dots, y(\mathbf{s}_n)]^T$, the spatial structured prediction problem aims to learn a model (or function) f such that $\mathbf{Y} = f(\mathbf{X})$. Once the model is learned, it can be used to predict the responses at other locations based on their explanatory

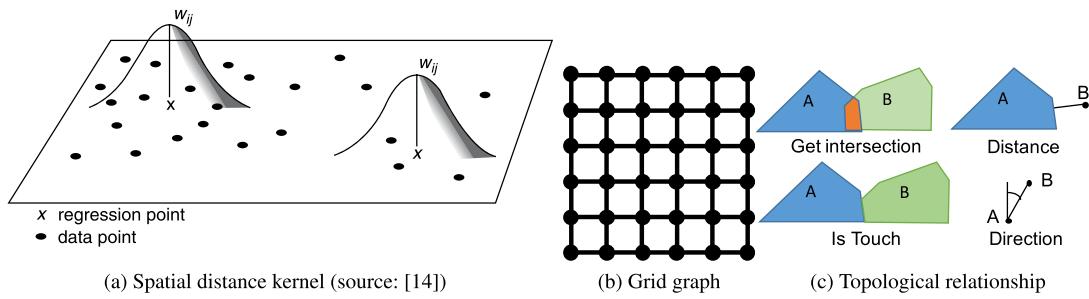


FIGURE 1. Spatial structure on 2D Euclidean plane.

features. The problem can be further categorized into spatial classification for categorical response and spatial regression for continuous response.

The problem defined above is called spatial structured prediction because there exists spatial dependency (instead of independence) between variables at different locations ($\mathbf{x}(s_1), \mathbf{x}(s_2), \dots, \mathbf{x}(s_n), y(s_1), y(s_2), \dots, y(s_n)$). It is different from other types of structured prediction (e.g., classifying sentences based on syntactic structure such as parse tree, recognizing speech signals based on the sequential structure of vocals) in that the dependency structure is related to spatial locations. Spatial structured prediction is different from traditional non-structured prediction problems. In non-structured prediction problems, samples are commonly assumed to be independent and identically distributed (i.i.d.). Thus, a same model can be used to predict every sample independently. In other words, a same model $y(\mathbf{s}) = f(\mathbf{x}(\mathbf{s}))$ can be applied to $\forall \mathbf{s}$ individually. However, the i.i.d. assumption is often violated in spatial data due to implicit spatial structure based on sample locations. Therefore, a model needs to capture the relationship between entire feature maps and response variable map, as indicated by $\mathbf{Y} = f(\mathbf{X})$.

Scope: There are other types of spatial structure prediction problems beyond the geospatial domain. Examples include image and video classification as well as retrieval in the computer vision community, where spatial structures between image parts play a critical role in prediction models. Examples include fine-grained image classification [7], [8], video captioning [9] and multi-media retrieval [10], [11]. However, we realize that the amount of literature on these topics in the computer vision community is so extensive that the topic these topics alone could be a separate survey. Thus, we do not provide exhaustive coverage in this paper.

A. TYPE OF SPATIAL STRUCTURES

This subsection introduces several common types of spatial structures based on the underlying spatial domain. A spatial domain is an underlying space in which locational coordinates are measured. Here we listed three most common spatial domains, including two-dimensional (2D) Euclidean plane, spatial network, and three dimensional (3D) geometric topological surface. Various spatial structures exist in each spatial domain. For example, in a 2D Euclidean plane,

spatial dependency structure can be established based on a distance between spatial points, adjacency between spatial raster cells, or topological relationships between geometric objects. On spatial networks, point samples often follow linear patterns along network paths (e.g., traffic crashes along road segments). On a geometric topological surface, spatial structure can be both regular geometry shapes and irregular terrains, topography, as well as contours and flow directions. We now introduce each category in detail below.

1) SPATIAL PROXIMITY ON 2D EUCLIDEAN PLANE

In a 2D Euclidean plane, spatial samples can be points, raster grid cells, or geometric objects (lines, polygons). For spatial point samples, implicit spatial structure exists between spatial points based on their distances. According to Tobler's first law of geography, "everything is related to everything else, but near things are more related than distance things" [15]. The law indicates that geographical data is essentially structured, with nearby locations mutually dependent on each other. The law is consistent with our common sense, e.g., households in the same neighborhood often have similar income levels. Thus, spatial structure can be modeled as a function of distance, with stronger dependency (or correlation) between sample attributes or classes that are closer to each other, as illustrated in Figure 1(a).

For raster grid cell samples such as earth image pixels, the most common spatial structure is the cell adjacency in a grid graph (Figure 1(b)). The most common assumption is the Markov property, i.e., a cell's attribute only depends on its neighbors. In this way, the joint distribution of all cells' attributes can be decomposed into a set of neighborhood cliques. The entire map is called a Markov random field, which has been widely used in image segmentation applications.

For spatial samples as complex geometric objects such as lines or polygons, the spatial structure between sample objects can be based on topological relationships, such as touch, overlap, disjoint, within, as shown in Figure 1(c). For example, a railroad (spatial line) may "cross" the Mississippi (spatial line) that "passes through" a national park (polygon) "within" the state of Minnesota (polygon). As another instance of example, in a construction field, the spatial relationship structure between the footprints of building,



FIGURE 2. An example of US interstate highway networks (source: [18]).

pathways, and instruments are critically important for safety maintenance.

2) LINEAR DEPENDENCY ALONG SPATIAL NETWORKS

A spatial network (also called geometric graph) is a unique type of network whose nodes and edges are geometric objects. In other words, a spatial network combines the spatial relationships both in the Euclidean space and in the network space. Common examples include road networks and river stream networks. As illustrated by Figure 2, in a road network, a node is a road intersection (spatial point), while an edge is a road segment between two intersections (spatial line segment). The spatial structure of samples in a spatial network domain has several unique characteristics. First, samples are embedded in the geometric lines of network edges. Thus, samples show a linear dependency structure along network edges [16], [17]. For example, when predicting traffic volume, samples are not spreading across an entire Euclidean plane but are constrained by the direction of a network path. Second, between different network edges, there is also a network topology structure (e.g., direction, neighbor, predecessor, and successor). For example, in a highway system, a road segment (edge) has one or two directions, and different road segments are connected based on intersections (turns).

3) GEOMETRIC TOPOLOGICAL SURFACE

A geometric topology surface is a 2D manifold embedded in a 3D Euclidean space. Complex spatial structure patterns can exist on the surface based on geometric and topological features. For example, on a geographical terrain map collected from LiDAR point clouds (Figure 4(a)), the terrain structures not only show the shape of the Earth's surface but also control water distributions (surface extents) and water flow directions. Another example is 3D point clouds in building information modeling (Figure 3). The geometric shape of objects and their topological relationships can indicate the structural properties of buildings, which is very important in construction and structural engineering.



FIGURE 3. Spatial structure on geometric topology surface based on 3D point cloud (source: [19]).

III. APPLICATIONS

This section introduce several representative applications for spatial structured prediction, with a particular focus on applications that involve complex spatial structures in geometric topology surface.

A. HYDROLOGY

One fundamental problem in hydrology is to map the distribution of water (e.g., flood, river streams) on the Earth's surface from remotely sensed imagery. The problem plays a critical role in flood disaster response [23] and national water forecasting at the NOAA National Water Center on the University of Alabama campus [24], [25]. The problem, however, is significantly more complex than image classification, since the flow and distribution of water is constrained by the complex terrain on the Earth's surface, as shown in Figure 4(a). Integrating such complex spatial structures into a data-driven model can significantly enhance the model's interpretability and allow for comparison with existing hydrological simulation models.

B. MATERIAL SCIENCE

The properties of materials arise from atomic structures. Exploring atomic structures requires a comprehensive understanding of the potential energy landscape, a function between all atomic positions and their potential energy [26]–[32]. Figure 4(b) is an illustrative example. Complex topological structures on the landscape indicate the physical process of materials transiting from a stable state (local minimum energy) to a saddle state, and then relaxing to a nearby stable state. Modeling such complex spatial patterns into data-driven models helps reveal the relationship between local atomic displacement and material properties (e.g., shear strain) towards developing new materials.

C. BIOCHEMISTRY

Analyzing the chemical properties of a molecular system plays an important role in understanding the life process. Various data have been collected on the spatial structures of molecules through X-ray crystallography, NMR, and electron microscopy [33]. Theoretical chemistry has shown that

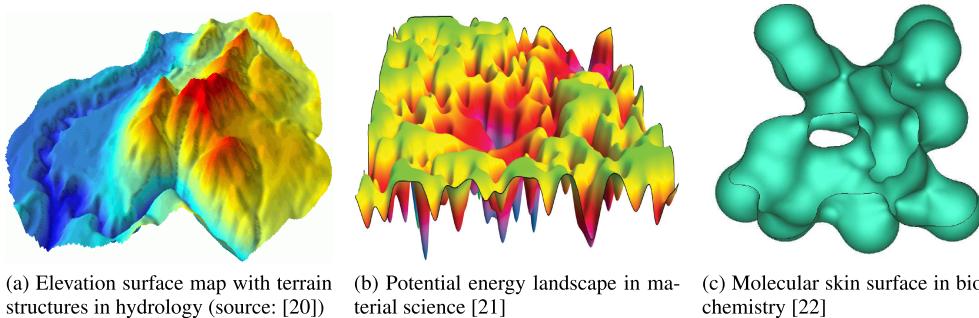


FIGURE 4. Motivation applications in hydrology, material science, and biochemistry.

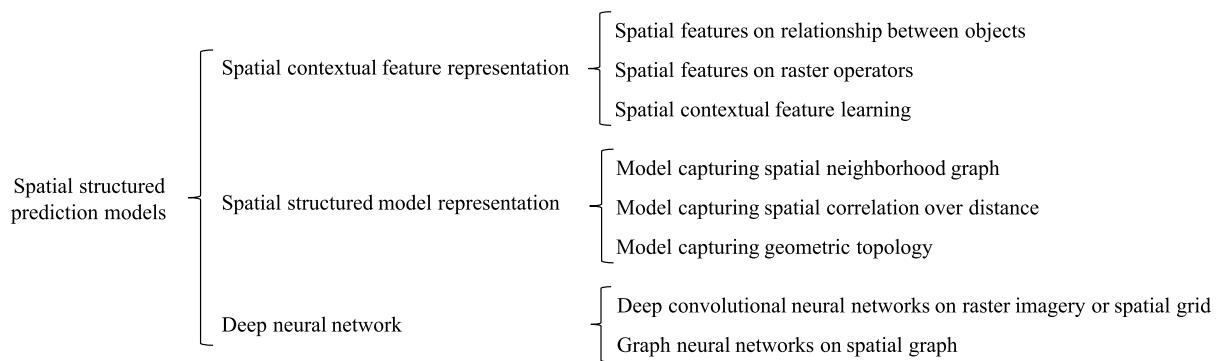


FIGURE 5. A taxonomy of spatial structured prediction models based on underlying approaches.

spatial structural information can predict protein function and interaction [34]. Building transparent data-driven models with such complex geometrical and topological structures (Figure 4(c)) can enhance model interpretability by the existing theories in the field [22], [34], [35].

IV. CHALLENGES

Learning spatial structured prediction models is challenging for several reasons. First, spatial structural dependency is often implicit in continuous space. For example, in disease risk mapping, nearby locations tend to have a similar level of disease risks due to the spatial autocorrelation effect. The effect is implicit from input data, which contains continuous maps such as environmental variables and human movements. This requires machine learning algorithms to explicitly model the spatial dependency structure. Second, spatial structural dependency can be very complex, both in an Euclidean plain and in a geometric topology surface. For example, in flood mapping, flood locations are derived not only based on spectral pixels in an earth imagery, but also constrained by the topography (e.g., contours, flow directions) of elevation map (a 3D surface). Third, spatial structure can be further complicated by additional effects such as spatial heterogeneity and the multi-scale effect. Spatial heterogeneity means that the dependency may be local instead of global. The multi-scale effect means that the spatial dependency structure may exist at different spatial scales in a spatial

hierarchy. For instance, in traffic prediction along road networks, the spatial network structured patterns can be modeled both at a coarse scale (only with major highways) and a fine scale (with county roads and local streets). Fourth, learning a spatial structured model for a large data volume can be computationally challenging, due to the structural complexity in model representation. Many learning problems related to graph structures are NP-hard. Finally, there can be limited ground truth. For example, in earth science applications, collecting ground truth training class labels often involves sending a field crew on the ground or visually interpreting earth images, which is tedious and time consuming. This is a particularly challenge for spatial structured prediction since a complex model often requires more training samples in the learning process in order to avoid overfitting.

V. A TAXONOMY OF TECHNIQUES BY APPROACHES

We now introduce a taxonomy of existing spatial structured prediction models, as shown in Figure 5. Based on the underlying approaches to capture spatial structure dependency, existing works can be categorized into three groups: spatial contextual feature representation, spatial structured model representation and deep neural network. Spatial structured feature representation approach focuses on creating or learning contextual features that captures spatial dependency structure. Examples of such features include spatial relationships between geometric objects (e.g., within, overlap),

outputs of spatial raster operators (e.g., texture), as well as learning contextual features to implicitly reflect spatial dependency (e.g., spatial network embedding). Spatial structured model representation approach aims to explicitly capture spatial dependency structure within machine learning model representation. Examples include modeling spatial neighborhood graphs in Markov random fields, modeling spatial correlation over distance in Gaussian process (Kriging) and modeling geometric topology in flow direction trees. Finally, deep neural network approach focuses on learning a blackbox model to automatically learn complex spatial structure dependency in an end-to-end manner. Specifically, it can be categorized into deep convolutional neural networks for raster imagery or regular spatial grid and graph neural networks for spatial graphs. We now introduce each approach and its specific methods, and compare different methods in their advantages and disadvantages.

A. SPATIAL CONTEXTUAL FEATURE GENERATION

One way of incorporating spatial dependency into prediction methods is to augment input data with additional spatial contextual features. The spatial context of a sample location refers to information related to the spatial structure surrounding it, such as relationships to other objects or locations, attributes of nearby samples, auxiliary semantic information from additional data sources. Once spatial contextual information is added into explanatory features, traditional non-spatial prediction methods can be used. We now introduce several approaches to generate spatial contextual features.

1) SPATIAL RELATIONSHIP FEATURES

Spatial contextual features can be generated based on spatial relationships with other locations or objects, such as distance or direction, touching, lying within or overlapping with another object. Spatial relationship features can be readily used in rule-based or decision tree-based models. Examples of techniques include spatiotemporal probability tree model to classify meteorological data on storms [36], [37], multi-relational spatial classification [38], prediction based on spatial association rules [39].

2) SPATIAL CONTEXTUAL FEATURES BASED ON RASTER OPERATORS

Spatial contextual features have long been used in classifying raster data (e.g., earth observation imagery) to reduce salt-and-pepper noise [40]. Specific methods include neighborhood window filters (e.g., median filter [41], weighted median filter [42], adaptive median filter [43], decision-based filter [41], [44], etc.), spatial contextual variables and textures [45], neighborhood spatial autocorrelation statistics [46]–[48], morphological profiling [49], object-based image analysis (e.g., mean, variance, texture of object segments) [50], and spatial zone ensembles [51], [52]. Sometimes, these methods are used in the post-processing step [53].

3) SPATIAL CONTEXTUAL FEATURE LEARNING

Spatial contextual feature learning aims to learn a low dimensional feature representation of spatial samples that reserve spatial structural information in the original feature space. Common techniques include embedding [54]–[56], matrix factorization [57], as well as recent unsupervised deep learning methods such as auto-encoder [58]. Recently, deep convolutional neural networks have been used to automatically learn complicated structured features in spatial raster data with convolutional operators [58]–[64]. One unique property of spatial data is that information from different sources can be fused into the same spatial framework, providing important spatial contexts for learning samples. For example, when predicting a fine-grained air quality map for an entire city, we can generate contextual features by fusing air quality records at ground stations, weather information, road network and traffic data, as well as POIs [65]. When predicting human behaviors from location history, auxiliary data from geosocial media can provide important semantic annotations [66], [67]. Generating contextual features through data fusion can have its own challenges (e.g., multi-modality, sparsity, noise). Various techniques have been explored, such as coupled matrix factorization, and context-aware tensor decomposition with manifold [68].

Spatial contextual feature generation is important in many practical applications (e.g., urban computing) due to two main advantages. First, generating appropriate contextual features can significantly enhance prediction accuracy due to the effectiveness of those features in explaining the response variable. Second, after spatial contextual features are generated, many traditional non-spatial predictive models can be used (e.g., random forest, support vector machine). This is sometimes convenient since there is no need to modify non-spatial prediction models or learning algorithms. At the same time, spatial contextual feature generation may require significant knowledge about the application domain.

B. STRUCTURED MODEL REPRESENTATION

Another general approach for spatial structured prediction problems is to incorporate spatial structural constraints into machine learning model representation. One of the most common framework to model dependency structure in machine learning is probabilistic graphical models [69]–[74], which use nodes and edges in a graph to model variables and dependency between variables. Graphical models can be further grouped into Markov random field for undirected graphs [75], Bayesian networks for directed graphs [76], and Markov logic networks for graphics with first order logic [77]. We now introduce several categories of techniques based on the type of spatial structural constraints being incorporated, including spatial neighborhood graph, spatial distance kernel, and geometric topology.

1) SPATIAL NEIGHBORHOOD GRAPH BASED MODELS

Spatial neighborhood graph based models assume that the spatial dependency structure follows a neighborhood

graph, i.e., dependency exists between neighboring locations. Spatial neighborhood graph can be derived based on cell adjacency on a raster framework or spatial distance in a continuous field. These models are generally considered as the Markov random field model family.

Markov random field (MRF) [71], [73], [74], [78]–[81] is a widely used model for areal data such as earth observation images, MRI medical images, and county-level disease count map. An MRF is a random field that satisfies the Markov property: the conditional probability of the observation at one cell given observations at all remaining cells only depends on observations at its neighbors. This property is consistent with the first law of geography that “nearby things are more related than distant things”. According to the Brook’s lemma [82], the joint distribution of cell observations can be uniquely determined based on conditional probability specified in the Markov property. Furthermore, according to the Hammersley-Clifford theorem [83], the corresponding joint distribution of MRF has a unique structure: it can be expressed by a set of potential functions on spatial neighbor cliques (i.e., symmetric functions that are unchanged by any permutations of input variables within a clique). Such a joint distribution is also called Gibb’s distribution. Equation 1 is an example. Its potential function is $W_{i,j}(y(\mathbf{s}_i) - y(\mathbf{s}_j))^2$ based on cliques of size two ($\mathbf{s}_i, \mathbf{s}_j$). The Markov property simplifies the modeling process: as long as the neighborhood structure is specified, the joint distribution of an MRF can be expressed by a potential function on neighbor cliques. Spatial prediction methods based on MRF include ones that explicitly capture spatial dependency such as Simultaneous Autoregressive models (SAR), ones that implicitly capture spatial dependency such as Conditional Autoregressive models, and ones integrating MRF with other models such as Bayesian classifiers and support vector machines.

$$P(y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)) \propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i,j} W_{i,j}(y(\mathbf{s}_i) - y(\mathbf{s}_j))^2\right\} \quad (1)$$

Simultaneous Autoregressive models (SAR) explicitly express spatial dependency across response variables [84], [85]. The SAR model extends traditional linear regression with an additional spatial autoregressive term, as shown in Equation 2, where \mathbf{Y} is a n by 1 column vector of all response variables, \mathbf{X} is a n by m sample covariate (feature) matrix, $\boldsymbol{\beta}$ is a m by 1 column vector of coefficients, and $\boldsymbol{\epsilon}$ is a n by 1 column vector of i.i.d. Gaussian noise (residual errors), \mathbf{W} is row-normalized \mathbf{W} -matrix, and ρ reflects the strength of spatial dependency effect. The i th row of the spatial autoregressive term \mathbf{WY} is a weighted average of response variables at all neighboring locations of \mathbf{s}_i . Parameters in SAR can be estimated based on the maximum likelihood method. SAR model can also be extended for spatial classification via logit transformation. It is worth noting that the spatial autoregressive term can also be added into other variables than the responses, such as covariates as in the spatial Durbin model or residual errors as in the spatial error

model [86].

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

Another model similar to MRF is conditional random field (CRF) [87]–[90], which directly models spatial dependency within the conditional probability function $P(\mathbf{Y}|\mathbf{X})$. Its potential function on class labels within a clique is conditioned on feature vector \mathbf{X} . Several variants of CRF have been proposed including decoupled conditional random field [88], discriminative random field [89] and support vector random field [90]. The difference between MRF and CRF is that the former is generative while the latter is discriminative.

The advantage of MRF-based models is that spatial dependency can be modeled in a very intuitive and simple way (designing potential functions). The limitations include high computational costs in parameter estimation and strong assumptions on the structure of joint probability distribution. In addition, neighborhood relationships are assumed to be given as inputs. Thus, fixed neighborhoods such as square windows are often used for simplicity. For applications where spatial data is anisotropic, determining spatial neighborhood structure is also a challenge.

There are other models for spatial model representations, including spatial statistics and spatial econometrics [84], [85] such as spatial panel model and spatial Dublin model [86], [91], as well as spatial regularization on loss function [92]–[94]. Methods based on spatial network connectivity are relatively less studied, including spatial network Kriging [95], and spatial network autoregressive models [96].

2) SPATIAL DISTANCE BASED MODELS

Spatial distance based methods assume that the strength of dependency is only determined by the distance between sample locations, regardless of directions. The most common method is the Gaussian process model (also called Kriging).

Different from MRF based models that are used for areal data, Gaussian process based models are used for spatial prediction (interpolation) on point reference data [85]. Given observations of a variable at sample locations in continuous space, the problem aims to interpolate the variable at an unobserved location. The Gaussian process assumes that observations at any set of sample locations jointly follow a multivariate Gaussian distribution. The mean term at a location is determined by its local covariates. The residual error term at a location is assumed to be weakly stationary and isotropic, so that the covariance matrix can be expressed as a function of distance (covariogram).

Specifically, Gaussian process (Kriging) assumes that any set of sample observations $\mathbf{Y} = [y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)]^T$ follows a multivariate Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is the covariance matrix $\Sigma_{ij} = \text{Cov}(y(\mathbf{s}_i), y(\mathbf{s}_j)) = C(\mathbf{s}_i - \mathbf{s}_j)$. The main difference of Gaussian process from classical regression is that the residual errors are not mutually independent (the covariance matrix is not diagonal and the non-diagonal elements can be determined based on covariogram $C(\mathbf{h})$, which is a function describing the degree of

spatial dependence over location differences [85]). It can be shown that the optimal predictor (minimizing expected square loss) of $y(\mathbf{s}_0)$ given other observations $y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)$ is the conditional expectation as shown in Equation 3, which can be estimated based on the covariance structure from covariogram [85]. This method is also called universal Kriging since it involves covariates \mathbf{X} . Special cases without covariates include simple Kriging (with known constant mean) and ordinary Kriging (with unknown constant mean) [97].

$$\hat{y}(\mathbf{s}_0) = E[y(\mathbf{s}_0)|y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)] \quad (3)$$

Gaussian process shares some similarities with MRF in that both of them incorporate spatial dependency (autocorrelation) across sample locations into model structure. There are several differences, however: first, Gaussian process is developed for point reference data while MRF is developed for areal data; second, MRF models spatial dependency through W-matrix (W) while Gaussian process models spatial dependency through covariance function on spatial distance (covariogram). Similar to MRF which assumes a given neighborhood structure and the Markov property, Gaussian process has assumptions on spatial stationarity and isotropy.

3) GEOMETRIC TOPOLOGY BASED MODELS

Geometric topology based models learn spatial structures on a topological surface (e.g., 3D terrain map) beyond the common Euclidean space (e.g., raster image or video). Such spatial structure is often related with geometric shapes or contour patterns. It has been extensively studied in computational geometry, or more specifically computational topology [33], [98]–[100].

One such spatial structured model is geographical hidden Markov tree (g-HMT) [101], [102], a probabilistic graphical model that generalizes the common hidden Markov model (HMM) from a total order sequence to a partial order tree. The tree structure is motivated by disciplinary knowledge in hydrology. As illustrated in Figure 6, due to gravity, water flows from a high elevation to nearby lower elevations. If location 5 is in the flood class, locations 2-4 and 6-7 should also be flood. Such partial order is captured in a tree structure (Figure 6(b)) in the hidden class layer. Each hidden class node has an associated observed feature node for the same pixel. Such a spatial dependency structure can potentially reduce classification errors due to noise, obstacles, and heterogeneity among spectral features of individual pixels. Compared with existing spatial structured models (e.g., Markov random field, post-processing label propagation), proposed g-HMT can reflect directed dependency following physics in hydrology. Empirical evaluations have shown that g-HMT significantly improves accuracy over existing methods due to overcoming noise and obstacles in feature maps [101]. The model has also been extended to address limited observation based on the physics-aware spatial structural constraints (e.g., water flow directions on a terrain surface) [103].

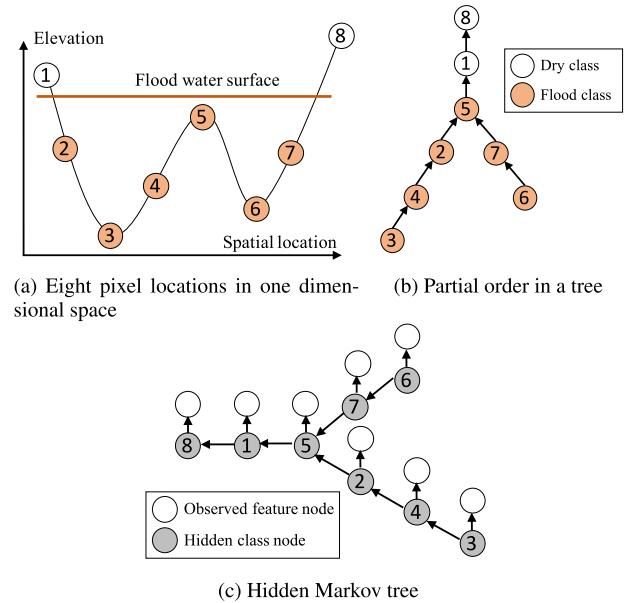


FIGURE 6. Illustration of geographical hidden Markov tree.

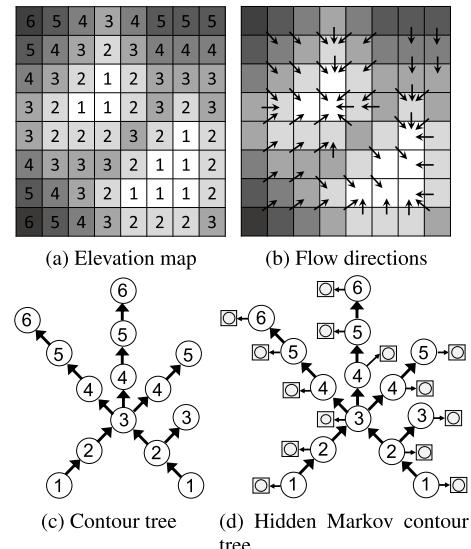


FIGURE 7. Illustration of hidden Markov contour tree.

Another model example is hidden Markov contour tree (HMCT) [104]. HMCT is a probabilistic graphical model that generalizes the common hidden Markov model from a total order sequence to a partial order contour tree. It extends geographical hidden Markov tree with 2D structures such as contours. Figure 7 illustrates an example in flood mapping in hydrology. Due to gravity, floodwater flows from one location to nearby lower locations, as illustrated by arrows in Figure 7(b). Such dependency structure is represented by a contour tree, as shown in Figure 7(c), where a node represents the class (flood or dry) of a contour segment (e.g., the three contiguous pixels with elevation 1), and an edge represents the flow direction between two adjacent contour segments. In the HMCT model, each hidden class node in the contour

tree is also associated with a set of observed feature nodes, corresponding to the spectral information of pixels in the same contour segment, as shown in Figure 7(d). In this way, class inference is based on not only local spectral information but also flow direction and contour structure. Empirical evaluations on datasets from Hurricane Mathew floods showed that HMCT outperformed existing methods with 10% to 30% gain in F-score, due to overcoming noise and obstacles in features [104]. It can be more robust to g-HMT in cases where different sides of river banks should model the dependency structure separately.

Geometric topology based methods are different from the previous two methods in that they capture topological structure on a 3D surface instead of 2D spatial plane. Such topological structures can be directed (e.g., flow directions based on surface gradient). The specific techniques discussed above can be considered as special cases of Bayesian network (they are directed trees or polytrees instead of general directed graph). Such unique topological structures on 3D surface are important for applications in hydrology, material science, and biochemistry.

C. DEEP NEURAL NETWORKS

Deep neural network is a new approach that becomes increasingly popular in recent years for structured prediction problems. Compared with the previous two approaches, i.e., spatial contextual feature generation and structured model representation, the deep learning approach could automatically learn structured features in an end-to-end manner (without manually handcrafting spatial contextual features, or explicit specifying spatial structural in model representation). Deep learning methods for spatial structured prediction can be generally categorized into deep convolutional neural networks (DCNN) and graph neural networks (GNN). DCNN is originally developed for image data in the computer vision community, and thus is naturally applicable to spatial raster data such as earth imagery [60]–[62] and a discretized spatial grid [63], [64]. Graph neural networks are generalization of deep neural networks from images to graphs. GNN can be used for spatial structured prediction because spatial data can be represented as graphs with nodes representing locations and edges representing spatial relationships between locations [105], [106]. We put deep learning as a separate approach from the previous two due to its wide popularity in recent years. Since DCNN is widely known, we introduce more details on GNN in this subsection.

Graph neural networks are a set of techniques that generalizing deep neural networks to graphs. Generalizing deep learning to graph data is non-trivial due to the lack of a regular input domain like imagery or videos, and the structural complexity of graphs. Several surveys exist on this important emerging topic [107]–[110]. Existing techniques can be categorized based on the underlying approaches to capture graph dependency between nodes, such as graph convolutional neural networks, graph recurrent neural networks, and graph attention neural networks.

Graph convolutional neural networks generalize deep convolutional neural networks into graph data. The main challenge is that there is no regular neighborhood window as in imagery data. There are two general approaches: the spectral-based approach and the non-spectral-based approach. The spectral-based approach uses the spectral representation graph (i.e., eigenvalues and eigenvectors of the graph Laplacian matrix). Techniques have been proposed to reduce the number of parameters and computational complexity, such as ChebyNet [111] and simple Graph Neural Network [112]. The non-spectral approach directly generalizes convolutional operations from a regular grid to a graph, such as using different weight matrices for nodes with different degrees [113] and graph diffusion neural networks [114]. Graph convolutional networks can generalize convolutional filters from raster imagery to graph neighborhoods. One potential issue is that the methods often assume a fix graph topology (and corresponding Laplacian matrix), and thus are not applicable to cases where the underlying graph structure changes from training data to test data.

Graph recurrent neural networks generalize recurrent neural networks from sequences to graphs. Its main advantage is the capability of capturing directed dependency (directed graph). The main idea is to use gating mechanism to control the flow of information across graph nodes along directed paths. The most common technique is graph Long Short-Term Memory (graph LSTM) [115]–[119]. An LSTM [58] is a special recurrent neural network that uses internal gates to control the memory of information from previous steps in a sequence, possibly capturing long range dependency. An LSTM can be generalized from sequences to trees and graphs by allowing for multiple predecessors or successors of a node. Such models have been applied to natural language processing [115]–[117], but the underlying directed graph is often small. Graph recurrent neural networks can potentially be applied to directed spatial graphs (e.g., river networks, road networks considering traffic directions). The potential issue is the high computational costs in learning when the graph is large.

Graph attention networks [120] uses the self-attention mechanism on graphs to perform node classification a graph structure. The main idea is to compute the hidden representations of each node in the graph by learning weights over its neighbors. Weights between a node to its neighbors are learned by a self-attention strategy. Self-attention means mapping a graph to the same graph itself while learning the weights between a node and its neighbors (the amount of attention on its neighbors required to classify the current node). The self-attention mechanism can address the issue of varying node degrees in graph convolution (this violates the assumption of a fixed window size in traditional convolutional operations) through averaging hidden representations from all neighbors based on their weights.

Research on the field of deep neural network, particularly graph neural network, is still quickly growing. The main advantages of this approach, compared with the traditional

TABLE 2. Comparison of different methods under structured model representation.

Approaches	Advantages	Disadvantages
Spatial contextual feature generation	It is often effective in improving prediction accuracy if generated features can well explain the response variable. After spatial feature generation, many traditional non-spatial predictive models can be used without the need of changing the model.	It may require significant knowledge about the application domain. Features need to be re-generated for different applications.
Structured model representation	Once a model is developed, it is generally applicable to different applications. Model structure explicitly captures spatial dependency structure.	Time costs of model learning can be high. Re-designing model representation for spatial structures requires significant machine learning expertise.
Deep neural network	End-to-end learning does not need handcrafting spatial features. It is able to model complex spatial structure patterns in neural network layers.	Blackbox model representation is hard to interpret. It often requires a large amount of training labels.

approaches, are that it does not require handcrafting or learning spatial contextual features, and it is able to learn complex spatial structure dependency patterns. The main limitations include that deep neural networks are often black-box models and thus are hard to interpret and that it requires a large amount of training labels.

D. SUMMARY AND COMPARISON

Table 2 provides a summary of comparisons between the three major approaches of spatial structured prediction models: spatial contextual feature generation, structured model representation, and deep neural networks. Incorporating spatial structures through contextual feature generation is an effective approach when the generated features are effective predictors for the response variable. It does not require the change of machine learning models, and thus traditional non-spatial machine learning methods can be used (e.g., random forests, support vector machine). Due to this reason, it is often used in practice by interdisciplinary application domain scientists. The limitation is that the approach requires sufficient domain knowledge on the application and the process needs to be repeated for different applications. In contrast, the approach of incorporating spatial structure in model representation provides general tools that can be used for different applications, but it also has the limitation of high computational costs of model learning compared with the traditional non-spatial models. Finally, the emerging deep learning approach reduces the burden of handcrafting spatial features (the deep neural networks automatically learn effective feature representations), and is able to learn complex spatial structure patterns. The main disadvantages are that the models are often hard to interpret and model training requires a large amount of training labels.

VI. FUTURE DIRECTIONS

This section identifies research gaps and discusses some potential future research directions in the field of spatial structured prediction.

A. GEOMETRIC DEEP LEARNING

Recently, the topic of geometric deep learning [121] has attracted growing interests from the deep learning research community. Geometric deep learning generalizes traditional deep learning from regular raster

frameworks (e.g., images and videos) in Euclidean to irregular spatial structures on a geometric surface (e.g., 3D point cloud, protein surface). The problem is very challenging due to several reasons: First, there is no fix input spatial domain that we can easily learn a local invariant convolutional operator; Second, explicit topological relationship can exist on a geometric surface (e.g., water flow direction along a terrain map); Third, there may be limited training labels. According to a recent overview article [121], existing methods are largely based on graph convolution neural networks [107], [108], which rely on a fix graph topology structure, and thus cannot generalize to cases where the graph structure changes from one data to another.

Addressing the challenges require the integration of deep learning with computational geometry (or computational topology, computer graphics). There is already extensive research on modeling complex spatial structures in the field of computational geometry (e.g., topological data analysis, contour trees, shape analysis). These existing works provide unique opportunities to extract spatial structural constraints from observed geometric surface data. Such spatial structural constraints can then be integrated with graph neural network models. The integration of data-driven approaches (graph neural networks) with physics-aware structural constraints (computation topology) into the backbone of model representation has several advantages. First, explicit spatial structural information often indicates the underlying physical process, and thus enhances model interpretability. Second, spatial structural constraints also have a positive side-effect of providing regularization for model learning when training labels are limited. For example, we can potentially integrate a contour tree from an elevation surface with graph recurrent neural networks to capture both non-linear relationships between variables and the flow direction and contour patterns on an elevation surface.

B. MODEL TRANSPARENCY AND INTERPRETABILITY

Another potential future research direction is to improve model transparency and interpretability. Existing spatial structured prediction models based on deep learning techniques are often based on a black-box model representation, making it hard to interpretable by the existing knowledge and theories in an application domain. This is a particular

handicap for scientific applications such as earth science, biology, and chemistry. Recently, there is a growing interest on interpretability in the machine learning research community. Interpretable machine learning [122]–[126] (also called explainable machine learning or AI [127], [128]) refers to techniques that help human users better understand the behavior of machine learning models. According to a recent survey [124], existing approaches can be categorized into intrinsic interpretation that focuses on constructing self-explanatory models, and post-hoc interpretation that creates a second model to provide explanations for an existing black-box model. Techniques for intrinsic interpretation include transparent model structures [129] such as decision trees, rule-based models [130], linear models with sparse features; adding globally interpretable constraints [131] such as regularizing loss to learn disentangled representations in CNN [132] or constructing a capsule network [122]; and locally interpretable structure such as attention mechanism [133]. Techniques for post-hoc interpretation [134], [135] includes learning interpretable models to mimic a black-box model [136]; testing model sensitivity to input features [137]–[139]; and explaining representations in black-box models [140]–[143]. Spatial structured prediction seems more suitable for intrinsic interpretation techniques given the existence of structural information. However, existing methods that build transparent model representation often assume simple structures such as trees or rules. Techniques for transparent models with complex spatial structural constraints are largely underexplored.

There are two potential strategies to address the challenge. One strategy is to incorporate domain knowledge into the design of model structures or model components [144], [145] so that the model is aware of the underlying physics. The other strategy is to explicitly model the spatial structural constraints (e.g., from computational topology) and build it into the backbone of model representation.

C. SCALABLE INFERENCE

Inference of structured prediction models are computationally very challenging due to structural complexity, which violates a common assumption that data samples are independent and identically distributed. Indeed, inference on a general graph is NP-hard [72]. Existing approximate inference algorithms include loopy belief propagation [146], variational inference [147], and sampling methods [148]. However, these methods are largely based on statistical properties instead of considering unique spatial structural properties. For example, spatial data can show a local effect (i.e., dependency is stronger within local areas) and a multi-scale effect (i.e., spatial patterns can be analyzed at different spatial scales and resolutions). Based on the local effect, we can potentially develop techniques of divide-and-conquer to conduct inference within individual zones and synchronize results together towards a global solution. Based on the multi-scale effect, we can potentially investigate multi-resolution filtering that first infers unknown classes on a coarse spatial scale

(much smaller data size) and refine the inferred class boundaries on a finer spatial scale. Another potential direction is to develop parallel algorithms for scalable inference. The problem is further complicated by the lack of a regular grid structure in input data like images or videos, making it hard to utilize existing parallel computational frameworks such as GPU. Several works have been done on parallelizing graph neural networks in GPU [149], [150], but the problem is largely underexplored.

VII. CONCLUSION

This paper provides an overview of the spatial structured prediction problem. We define the problem with different types of spatial structures and their applications. We also provide a taxonomy of common techniques categorized by the underlying approaches. In the end, we discuss several potential research directions.

REFERENCES

- [1] L.-C. Chen, A. Schwing, A. Yuille, and R. Urtasun, “Learning deep structured models,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1785–1794.
- [2] D. Weiss and B. Taskar, “Structured prediction cascades,” in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 916–923.
- [3] Z. Jiang, “A survey on spatial prediction methods,” *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 9, pp. 1645–1664, Sep. 2019.
- [4] S. Shekhar, Z. Jiang, R. Ali, E. Eftelioglu, X. Tang, V. Gunturi, and X. Zhou, “Spatiotemporal data mining: A computational perspective,” *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 4, pp. 2306–2338, Oct. 2015.
- [5] S. Shekhar, Z. Jiang, J. Kang, and V. Gandhi, “Spatial data mining,” in *Encyclopedia of Database Systems*, 2nd ed. L. Liu and M. T. Özsu, Eds. New York, NY, USA: Springer-Verlag, 2017.
- [6] Z. Jiang and S. Shekhar, *Spatial Big Data Science*. Cham, Switzerland: Springer, 2017.
- [7] X. He and Y. Peng, “Weakly supervised learning of part selection model with spatial constraints for fine-grained image classification,” in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4075–4081.
- [8] Y. Peng, X. He, and J. Zhao, “Object-part attention model for fine-grained image classification,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018.
- [9] J. Zhang and Y. Peng, “Object-aware aggregation with bidirectional temporal graph for video captioning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8327–8336.
- [10] Y. Peng and C.-W. Ngo, “Clip-based similarity measure for query-dependent clip retrieval and video summarization,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 612–627, May 2006.
- [11] Y. Peng, X. Huang, and Y. Zhao, “An overview of cross-media retrieval: Concepts, methodologies, benchmarks, and challenges,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2372–2385, Sep. 2018.
- [12] X. Zhu, Z. Ni, M. Cheng, F. Jin, J. Li, and G. Weckman, “Selective ensemble based on extreme learning machine and improved discrete artificial fish swarm algorithm for haze forecast,” *Int. J. Speech Technol.*, vol. 48, no. 7, pp. 1757–1775, Sep. 2017.
- [13] A. Onan, S. Korukoğlu, and H. Bulut, “A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification,” *Inf. Process. Manage.*, vol. 53, no. 4, pp. 814–833, Jul. 2017.
- [14] A. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Hoboken, NJ, USA: Wiley, 2002.
- [15] W. R. Tobler, “A computer movie simulating urban growth in the Detroit region,” *Econ. Geogr.*, vol. 46, pp. 234–240, Jun. 1970.
- [16] Z. Jiang, M. Evans, D. Oliver, and S. Shekhar, “Identifying k primary corridors from urban bicycle GPS trajectories on a road network,” *Inf. Systems*, vol. 57, pp. 142–159, Apr. 2016.
- [17] A. Sainju and Z. Jiang, “Mapping road safety features from streetview imagery: A deep learning approach,” 2019, *arXiv:1907.12647*. [Online]. Available: <https://arxiv.org/abs/1907.12647>

[18] Wikipedia. *Interstate Highway System*. Accessed: Nov. 1, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Interstate_Highway_System

[19] M. Ball. *NASA and Autodesk Collaborate on Sustainable Building Performance Monitoring*. Accessed: Nov. 1, 2019. [Online]. Available: <https://informedinfrastructure.com/3154/nasa-and-autodesk-collaborate-on-sustainable-building-performance-monitoring/>

[20] University of Washington. *ArcGIS Modules*. Accessed: Nov. 1, 2019. [Online]. Available: https://courses.washington.edu/gis250/lessons/introduction_arcpy

[21] D. J. Wales, "Exploring energy landscapes," *Annu. Rev. Phys. Chem.*, vol. 69, no. 1, pp. 401–425, Apr. 2018, doi: [10.1146/annurev-physchem-050317-021219](https://doi.org/10.1146/annurev-physchem-050317-021219).

[22] H. Edelsbrunner and P. Koehl, "The geometry of biomolecular solvation," *Combinat. Comput. Geometry*, vol. 52, pp. 243–275, Aug. 2005.

[23] NOAA National Weather Service. *Hydrologic Information Center—Flood Loss Data*. Accessed: Nov. 1, 2019. [Online]. Available: <http://www.nws.noaa.gov/hic/>

[24] D. Cline, "Integrated water resources science and services: An integrated and adaptive roadmap for operational implementation," *Nat. Oceanic Atmos. Admin.*, Silver Spring, MD, USA, Tech. Rep. IWRSS-2009-03-02, 2009.

[25] National Oceanic and Atmospheric Administration. *National Water Model: Improving NOAA's Water Prediction Services*. Accessed: Nov. 1, 2019. [Online]. Available: <http://water.noaa.gov/documents/wnr-national-water-model.pdf>

[26] P. G. Debenedetti and F. H. Stillinger, "Supercooled liquids and the glass transition," *Nature*, vol. 410, no. 6825, p. 259, 2001.

[27] L. Tian, L. Li, J. Ding, and N. Mousseau, "ART_data_analyzer: Automating parallelized computations to study the evolution of materials," *SoftwareX*, vol. 9, pp. 238–243, Jan. 2019.

[28] D. J. Wales, M. A. Miller, and T. R. Walsh, "Archetypal energy landscapes," *Nature*, vol. 394, no. 6695, p. 758, 1998.

[29] S. Sastry, "The relationship between fragility, configurational entropy and the potential energy landscape of glass-forming liquids," *Nature*, vol. 409, no. 6817, pp. 164–167, Jan. 2001.

[30] S. Sastry, P. G. Debenedetti, F. H. Stillinger, T. B. Schröder, J. C. Dyre, and S. C. Glotzer, "Potential energy landscape signatures of slow dynamics in glass forming liquids," *Phys. A, Stat. Mech. Appl.*, vol. 270, nos. 1–2, pp. 301–308, Aug. 1999.

[31] S. Sastry, P. G. Debenedetti, and F. H. Stillinger, "Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid," *Nature*, vol. 393, no. 6685, pp. 554–557, Jun. 1998.

[32] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[33] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. Providence, RI, USA: AMS, 2010.

[34] D. Gunther, R. A. Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny, "Characterizing molecular interactions in chemical systems," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2476–2485, Dec. 2014.

[35] T. Dey, "Protein classification with improved topological data analysis," in *Proc. 18th Int. Workshop Algorithms Bioinf.*, no. 6, 2018, pp. 1–6.

[36] A. McGovern, N. C. Hiers, M. Collier, D. J. Gagne, II, and R. A. Brown, "Spatiotemporal relational probability trees: An introduction," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 935–940.

[37] A. McGovern, N. Troutman, R. A. Brown, J. K. Williams, and J. Abernethy, "Enhanced spatiotemporal relational probability trees and forests," *Data Mining Knowl. Discovery*, vol. 26, no. 2, pp. 398–433, Mar. 2012.

[38] R. Frank, M. Ester, and A. Knobbe, "A multi-relational approach to spatial classification," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2009, pp. 309–318.

[39] W. Ding, T. Stepiński, and J. Salazar, "Discovery of geospatial discriminating patterns from remote sensing datasets," in *Proc. SIAM Int. Conf. Data Mining*, Dec. 2013, pp. 425–436.

[40] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, Mar. 2007.

[41] R. H. Chan, Chung-Wa, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479–1485, Oct. 2005.

[42] D. Brownrigg, "The weighted median filter," *Commun. ACM*, vol. 27, no. 8, pp. 807–818, Aug. 1984.

[43] H. Hwang and R. A. Haddad, "Adaptive median filters: New algorithms and results," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 499–502, Apr. 1995.

[44] S. Esakkirajan, T. Veerakumar, A. N. Subramanyam, and C. H. PremChand, "Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 287–290, May 2011.

[45] A. Puissant, J. Hirsch, and C. Weber, "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 26, no. 4, pp. 733–745, Aug. 2006.

[46] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, and J. Corcoran, "Focal-test-based spatial decision tree learning: A summary of results," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 320–329.

[47] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, and J. Corcoran, "Focal-test-based spatial decision tree learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 6, pp. 1547–1559, Jun. 2015.

[48] Z. Jiang, S. Shekhar, P. Mohan, J. Knight, and J. Corcoran, "Learning spatial decision tree for geographical classification: A summary of results," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2012, pp. 390–393.

[49] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[50] G. Hay and G. Castilla, "Geographic object-based image analysis (GEOBIA): A new name for a new discipline," in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 75–89.

[51] Z. Jiang, Y. Li, S. Shekhar, L. Rampi, and J. Knight, "Spatial ensemble learning for heterogeneous geographic data with class ambiguity: A summary of results," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst. (SIGSPATIAL)*, 2017, pp. 23–1–23–10.

[52] Z. Jiang, A. M. Sainju, Y. Li, S. Shekhar, and J. Knight, "Spatial ensemble learning for heterogeneous geographic data with class ambiguity," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 4, pp. 1–25, Aug. 2019.

[53] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[54] H. Wang and Z. Li, "Region representation learning via mobility flow," in *Proc. the ACM Conf. Inf. Knowl. Manage. (CIKM)*, 2017, pp. 237–246.

[55] Z. Yao, Y. Fu, B. Liu, W. Hu, and H. Xiong, "Representing urban functions through zone embedding with human mobility patterns," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3919–3925.

[56] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowl.-Based Syst.*, vol. 151, pp. 78–94, Jul. 2018.

[57] C. Peng, Z. Kang, Y. Hu, J. Cheng, and Q. Cheng, "Nonnegative matrix factorization with integrated graph and feature learning," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, pp. 1–29, Jan. 2017.

[58] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>

[59] R. Urtasun, "Learning deep structured models," in *Proc. ICML*, 2015, pp. 1785–1794.

[60] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[61] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "DeepSat: A learning framework for satellite imagery," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2015, p. 37.

[62] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[63] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 2588–2595.

[64] Z. Yuan, X. Zhou, and T. Yang, "Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 984–992.

[65] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1436–1444.

[66] F. Wu, Z. Li, W.-C. Lee, H. Wang, and Z. Huang, "Semantic annotation of mobility data using social media," in *Proc. 24th Int. Conf. World Wide Web (WWW)*, 2015, pp. 1253–1263.

[67] F. Wu and Z. Li, "Where did you go: Personalized annotation of mobility records," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 589–598.

[68] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE Trans. Big Data*, vol. 1, no. 1, pp. 16–34, Mar. 2015.

[69] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, nos. 1–2, pp. 1–305, 2007, doi: [10.1561/2200000001](https://doi.org/10.1561/2200000001).

[70] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[71] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[72] G. F. Cooper, "The computational complexity of probabilistic inference using Bayesian belief networks," *Artif. Intell.*, vol. 42, nos. 2–3, pp. 393–405, Mar. 1990.

[73] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B, Methodol.*, vol. 48, no. 3, pp. 259–279, Dec. 2018.

[74] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[75] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. London, U.K.: Springer, 2009.

[76] F. V. Jensen, *An Introduction to Bayesian Networks*, vol. 210. London, U.K.: UCL Press, 1996.

[77] M. Richardson and P. Domingos, "Markov logic networks," *Mach. Learn.*, vol. 62, nos. 1–2, pp. 107–136, 2006.

[78] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 4. Dordrecht, The Netherlands: Springer, 2013.

[79] S. Chawla, S. Shekhar, W. Wu, and U. Ozesmi, "Modeling spatial dependencies for mining geospatial data," in *Proc. SIAM Int. Conf. Data Mining*, Dec. 2013, pp. 1–17.

[80] S. Shekhar, P. R. Schrater, R. R. Vatsavai, W. Wu, and S. Chawla, "Spatial contextual classification and prediction models for mining geospatial data," *IEEE Trans. Multimedia*, vol. 4, no. 2, pp. 174–188, Jun. 2002.

[81] Q. Fu, A. Banerjee, S. Liess, and P. K. Snyder, "Drought detection of the last century: An MRF-based approach," in *Proc. SIAM Int. Conf. Data Mining*, Dec. 2013, pp. 24–34.

[82] D. Brook, "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems," *Biometrika*, vol. 51, no. 3/4, p. 481, Dec. 1964.

[83] P. Clifford, "Markov random fields in statistics," *Disorder in Physical Systems: A Volume in Honour of John Hammersley*. Cambridge, U.K., 1990, pp. 19–32.

[84] L. Anselin, *Spatial Econometrics: Methods and Models*. Dordrecht, Netherlands: Kluwer, 1988.

[85] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL, USA: CRC Press, 2014.

[86] P. A. Viton, "Notes on spatial econometric models," *City regional planning*, vol. 870, no. 03, pp. 9–10, 2010.

[87] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2001, pp. 282–289.

[88] C.-H. Lee, R. Greiner, and O. R. Zaïane, "Efficient spatial classification using decoupled conditional random fields," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (PKDD)*, 2006, pp. 272–283.

[89] S. Kumar and M. Hebert, "Discriminative fields for modeling spatial dependencies in natural images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1531–1538.

[90] C.-H. Lee, R. Greiner, and M. W. Schmidt, "Support vector random fields for spatial classification," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (PKDD)*, 2005, pp. 121–132.

[91] J. Mur and A. Angulo, "The spatial durbin model and the common factor tests," *Spatial Econ. Anal.*, vol. 1, no. 2, pp. 207–226, Nov. 2006.

[92] K. Subbian and A. Banerjee, "Climate multi-model regression using spatial smoothing," in *Proc. SIAM Int. Conf. Data Mining*, Dec. 2013, pp. 324–332.

[93] A. Karpatne, A. Khandelwal, S. Boriah, and V. Kumar, "Predictive learning in the presence of heterogeneity and limited training data," in *Proc. SIAM Int. Conf. Data Mining*, Philadelphia, PA, USA, Apr. 2014, pp. 253–261.

[94] J. Stoeckel and G. Fung, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," in *Proc. ICDM*, 2005, pp. 410–417.

[95] A. Okabe and K. Sugihara, *Spatial Analysis Along Networks: Statistical and Computational Methods*. Hoboken, NJ, USA: Wiley, 2012.

[96] J. Paul Lesage and W. Polasek, "Incorporating transportation network structure in spatial econometric models of commodity flows," *Spatial Econ. Anal.*, vol. 3, no. 2, pp. 225–245, Jun. 2008.

[97] D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong, "An experimental comparison of ordinary and Universal Kriging and inverse distance weighting," *Math. Geol.*, vol. 31, no. 4, pp. 375–390, May 1999.

[98] I. Kweon, "Extracting topographic terrain features from elevation maps," *Comput. Vis. Image Understand.*, vol. 59, no. 2, pp. 171–182, Mar. 1994.

[99] D. P. Kroese, T. Taimre, and Z. I. Botev, "Feature-based statistical analysis of combustion simulation data," in *Handbook of Monte Carlo Methods*. New York, NY, USA: IEEE, 2011, pp. 301–345. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118014967.ch8>

[100] G. Weber, P.-T. Bremer, and V. Pascucci, "Topological landscapes: A terrain metaphor for scientific data," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1416–1423, Nov. 2007.

[101] M. Xie, Z. Jiang, and A. M. Sainju, "Geographical hidden Markov tree for flood extent mapping," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2018, pp. 2545–2554.

[102] Z. Jiang, M. Xie, and A. M. Sainju, "Geographical hidden Markov tree," *IEEE Trans. Knowl. Data Eng.*, to be published.

[103] A. M. Sainju, W. He, and Z. Jiang, "Spatial classification with limited observations based on physics-aware structural constraint," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 1–8.

[104] Z. Jiang and A. M. Sainju, "Hidden Markov contour tree: A spatial structured model for hydrological applications," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2019, pp. 1–9.

[105] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, Oct. 2019.

[106] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*. [Online]. Available: <http://arxiv.org/abs/1707.01926>

[107] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," 2018, *arXiv:1812.08434*. [Online]. Available: <http://arxiv.org/abs/1812.08434>

[108] Y. Li, "Building more expressive structured models," Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 2017. [Online]. Available: https://www.cs.toronto.edu/%7B~7Dyujiali/papers/phd_thesis.pdf

[109] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," 2018, *arXiv:1812.04202*. [Online]. Available: <http://arxiv.org/abs/1812.04202>

[110] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," 2019, *arXiv:1901.00596*. [Online]. Available: <http://arxiv.org/abs/1901.00596>

[111] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[112] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: <http://arxiv.org/abs/1609.02907>

[113] D. K. Duvenaud, D. Maclaurin, J. Iparragirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2224–2232.

[114] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1993–2001.

[115] K. Sheng Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," 2015, *arXiv:1503.00075*. [Online]. Available: <http://arxiv.org/abs/1503.00075>

[116] V. Zayats and M. Ostendorf, "Conversation modeling on reddit using a graph-structured LSTM," *Trans. Assoc. Comput. Linguistics*, vol. 6, pp. 121–132, Dec. 2018.

[117] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-T. Yih, "Cross-sentence N-ary relation extraction with graph LSTMs," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 101–115, Dec. 2017.

[118] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with graph LSTM," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 125–143.

[119] Y. Yuan, X. Liang, X. Wang, D.-Y. Yeung, and A. Gupta, "Temporal dynamic graph LSTM for action-driven video object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1801–1810.

[120] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*. [Online]. Available: <http://arxiv.org/abs/1710.10903>

[121] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[122] H. Lakkaraju, S. H. Bach, and J. Leskovec, "Interpretable decision sets: A joint framework for description and prediction," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1675–1684.

[123] A. Vellido, J. D. Martín-Guerrero, and P. J. G. Lisboa, "Making machine learning models interpretable," in *Proc. ESANN*, vol. 12, 2012, pp. 163–172.

[124] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," 2018, *arXiv:1808.00033*. [Online]. Available: <http://arxiv.org/abs/1808.00033>

[125] Z. C. Lipton, "The mythos of model interpretability," 2016, *arXiv:1606.03490*. [Online]. Available: <http://arxiv.org/abs/1606.03490>

[126] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*. [Online]. Available: <http://arxiv.org/abs/1702.08608>

[127] D. Monroe, "AI, explain yourself," *Commun. ACM*, vol. 61, no. 11, pp. 11–13, Oct. 2018.

[128] D. Gunning, "Explainable artificial intelligence (XAI)," in *Proc. Defense Adv. Res. Projects Agency (DARPA)*, 2017, pp. 1–36.

[129] C. Chen, K. Lin, C. Rudin, Y. Shaposhnik, S. Wang, and T. Wang, "An interpretable model with globally consistent explanations for credit risk," 2018, *arXiv:1811.12615*. [Online]. Available: <http://arxiv.org/abs/1811.12615>

[130] J. Gao, N. Liu, M. Lawley, and X. Hu, "An interpretable classification framework for information extraction from online healthcare forums," *J. Healthcare Eng.*, vol. 2017, Aug. 2017, Art. no. 2460174.

[131] M. Wu, M. C. Hughes, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi-Velez, "Beyond sparsity: Tree regularization of deep models for interpretability," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1670–1678.

[132] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," *Artif. Intell. Med.*, vol. 16, no. 2, pp. 149–169, Jun. 1999.

[133] D. Malioutov and K. Varshney, "Exact rule learning via Boolean compressed sensing," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 765–773.

[134] N. El Bekri, J. Kling, and M. F. Huber, "A study on trust in black box models and post-hoc explanations," in *Proc. Int. Workshop Soft Comput. Models Ind. Environ. Appl.* Cham, Switzerland: Springer, 2019, pp. 35–46.

[135] G. Peake and J. Wang, "Explanation mining: Post hoc interpretability of latent factor models for recommendation systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 2060–2069.

[136] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Interpretable deep models for ICU outcome prediction," in *Proc. AMIA Annu. Symp.*, 2016, p. 371.

[137] Y. Zhang, N. Liu, S. Ji, J. Caverlee, and X. Hu, "An interpretable neural model with interactive stepwise influence," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2019, pp. 528–540.

[138] E. Horel, V. Mison, T. Xiong, K. Giesecke, and L. Mangu, "Sensitivity based neural networks explanations," in *Proc. NIPS Workshop Challenges Opportunities AI Financial Services, Impact Fairness, Explainability, Accuracy, Privacy*, 2018. [Online]. Available: <http://arxiv.org/abs/1812.01029>

[139] C. Modarres, M. Ibrahim, M. Louie, and J. Paisley, "Towards explainable deep learning for credit lending: A case study," in *Proc. Workshop Challenges Opportunities AI Financial Services: Impact Fairness, Explainability, Accuracy, Privacy (NIPS)*, 2018, pp. 1–8. [Online]. Available: <http://arxiv.org/abs/1811.06471>

[140] N. Liu, M. Du, and X. Hu, "Representation interpretation with spatial encoding and multimodal analytics," in *Proc. 12th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2019, pp. 60–68.

[141] F. Yang, N. Liu, S. Wang, and X. Hu, "Towards interpretation of recommender systems with sorted explanation paths," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 667–676.

[142] N. Liu, X. Huang, J. Li, and X. Hu, "On interpretation of network embedding via taxonomy induction," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2018, pp. 1812–1820.

[143] I. Unceta, J. Nin, and O. Pujol, "Towards global explanations for credit risk scoring," in *Proc. Workshop Challenges Opportunities AI Financial Services, Impact Fairness, Explainability, Accuracy, Privacy (NIPS)*, 2018, pp. 1–5. [Online]. Available: <http://arxiv.org/abs/1811.07698>

[144] N. Muralidhar, M. R. Islam, M. Marwah, A. Karpatne, and N. Ramakrishnan, "Incorporating prior domain knowledge into deep neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 36–45.

[145] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, 2019, pp. 558–566.

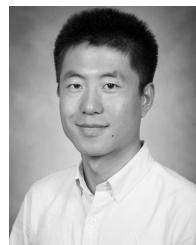
[146] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. 15th Conf. Uncertainty Artif. Intell.* San Mateo, CA, USA: Morgan Kaufmann, 1999, pp. 467–475.

[147] M. J. Beal, *Variational Algorithms for Approximate Bayesian Inference*. London, U.K.: Univ. London, 2003.

[148] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London, U.K.: Chapman & Hall, 2006.

[149] H. T. Son and C. Jones, *Graph Neural Networks With Efficient Tensor Operations in CUDA/GPU and Graphflow Deep Learning Framework in C++ for Quantum Chemistry*. Accessed: Nov. 1, 2019. [Online]. Available: <http://people.cs.uchicago.edu/~hytruongson/CCN-GraphFlow.pdf>

[150] L. Ma, Z. Yang, Y. Miao, J. Xue, M. Wu, L. Zhou, and Y. Dai, "Towards efficient large-scale graph neural network computing," 2018, *arXiv:1810.08403*. [Online]. Available: <http://arxiv.org/abs/1810.08403>



ZHE JIANG (Member, IEEE) received the B.E. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2010, and the Ph.D. degree in computer science from the University of Minnesota, Twin Cities, in 2016. He is currently an Assistant Professor with the Department of Computer Science, The University of Alabama, Tuscaloosa. His research interests include data mining and machine learning, particularly spatial and spatio-temporal data mining, spatial database, geographic information systems, and their interdisciplinary applications in earth science, transportation, public safety, and so on.

• • •