White Paper

Geometry and Learning from Data in 3D and Beyond

IPAM Long Program, Spring 2019

Author List (in alphabetical order)

Pradeep Kr. Banerjee, Sumukh Bansal, İlke Demir, Minh Ha Quang, Lin Huang, Ruben Hühnerbein, Scott C. James, Oleg Kachan, Louis Ly, Ola Marius Lysaker, Samee Maharjan, Anton Mallasto, Guido Montúfar, Kai Sandfort, Stefan C. Schonsheck, Pablo Suárez-Serrato, Katarína Tóthová, Yuguang Wang, Jia Le Xian, Rui Xiang, Hongkai Zhao

June 13, 2019

Table of Contents

Executive Summary 4

Introduction 6

3D Shape Analysis 8
- Introduction 8
- 3D Deep Learning for Shape Understanding 8
  - Shape Representations 10
  - Kernels and Architectures 10
  - Applications 12
- Geometry for Deep Learning 12
- Holistic and Scalable 3D Deep Learning Systems 13
- Expressive Shape Representations: From Procedural to Deep Generative Models 13
- Beyond 3D: Convolution on Manifold via Isometric Embeddings in Higher Dimensions 14

Graphs and Data 15
- Introduction 15
- Applied Harmonic Analysis on Graphs 16
- Spectral Based Graph Neural Networks 16
- Spatial-Based Graph Neural Networks 17
  - Graph Attention Mechanism 17
  - Graph Generation 18
- Generalization of GNNs and Applications 19

Optimal Transport / Wasserstein Information Geometry 19
- Introduction 20
- Optimal Transport in ML Models 20
- Computational Aspects 21
- Connection with Other Traditional Statistics/ML Methods 22
- Optimal Transport in Computer Graphics 23

Applications and Practical Matters 23
- Introduction 23
- Open Research Topics 24
  - Neural Machine Translation 24
  - ML in Medical Fields 24
  - Social Impacts of ML 25

Executive Summary

As technology advances, there is an ever-increasing demand to acquire, analyze, and generate 3D data. These necessarily large data sets must be amenable to efficient processing, analysis, and implementation in a variety of settings such as multi-dimensional modeling, high-resolution visualization, medical imaging, and the entertainment industry. Beyond 3D shapes, understanding and learning high-dimensional geometric structures is an active area of research. Given the goals of this program, the Core Participants identified four areas of particular interest: (1) 3D shape analysis, (2) graphs and data, (3) optimal transport and Wasserstein information geometry, and (4) practical matters.

3D shape modeling and analysis is critical in efforts to digitize and replicate the world without losing the core geometric information. Several applications like 3D content generation, shape modeling, animation, and manufacturing necessitate novel shape-analysis approaches. Fortunately, recent advances in deep-learning architectures (e.g., Convolutional Neural Networks) have facilitated solutions to many difficult problems in the 2D domain. In this program, a primary motivation was to adapt these advances to the 3D domain and thereby bridge the gap between traditional 3D shape analysis and deep learning methods. An important question emerged: *In addition to using machine learning to understand geometry, how can geometry be used to understand machine learning (ML)?* We believe that formulations around strong shape priors and shape properties will play a crucial role in understanding complex neural networks. Moreover, it will be essential to develop holistic shape-understanding systems where the analysis goes beyond the 3D geometry to including texture, color, material, and semantic information.

Graph-based analysis methods have been increasingly used for large-scale pattern recognition. Graph neural networks can learn representations of nodes, edges, subgraphs, whole graphs, and spaces of graphs. The generality of these networks allow for neural graph representation learning to be applied to many domains where standard techniques fail. This is especially true for problems that involve heterogeneous data. Graph neural networks explicitly learn lower-dimensional graph representations for less computational cost than classical dimensionality-reduction algorithms. Properly defining convolution-like graph operations is fundamental to formulating these networks and is an active area of research. Furthermore, incorporation of topological data-analysis tools (e.g., the study of persistent homologies) may geometrically motivate neural-network architectures.

Optimal transport provides a geometric framework for the study of probability distributions by extending the geometry of the sample space. It defines a similarity measures between (high-dimensional) distributions, providing geometric tools for navigating the space of probability distributions. In contrast to information theoretical divergences that do not consider the geometry of the sample space, this is an inherent consideration in optimal transport. In particular, Wasserstein metrics extend distance functions of sample spaces to distances between distributions. Solution of optimal transport problems is amenable to ML because the approach can be applied to minimize the distance between a probabilistic model and a data population; this should be further investigated. Moreover, potential associations between the geometric structures from optimal transport and other ML methods, such as kernel methods, should be explored.

Finally, the group espoused rigorously deriving the mathematical formalisms required to explain, describe, and predict the performance of ML models. For example, robust mathematical analyses would help to characterize aspects of the behavior of ML (e.g., why transfer learning works, how dropout is so effective in reducing overfitting, etc.). Moreover, because important decisions will increasingly be made with the assistance of AI, ML models must be verified and validated to build confidence in their estimates and predictions (e.g., machine-learning-assisted medical diagnoses and treatments). Perhaps a more thorough mathematical understanding of ML will help address important societal issues identified by the group. As AI systems become increasingly prevalent in everyday life and infrastructure, the social impacts of adversarial attacks, loss of privacy and anonymity, and intentional manipulation must be forefront. Risks include abuse of natural language processors (fake news), subversion of security systems (altered facial recognition), malicious and adversarial attacks on infrastructure (vulnerability of the power grid), and perhaps most importantly, job displacement. It is incumbent upon experts in our field to inform policy makers and decision makers so that proper regulations can be developed and laws enacted to protect individuals, societies, and economies.

Overall, participants in this program thoughtfully identified and outlined some of the fields and tools with the potential to yield significant impacts at the intersection of mathematics and ML. These fields include differential geometry, topology, probability theory, information geometry, optimal transport, partial differential equations, harmonic analysis, graph theory, combinatorics, functional analysis, linear algebra, and optimization, among others.

Introduction

With the striking advances in technology, computing power, and the deluge of data from sensors, experiments, simulations, the Internet, and social media, comes the urgent need for novel mathematical theoretical and computational methods to process and analyze big data and extract desired information. Although evolving techniques, such as deep learning, have achieved impressive performance in computer vision, natural language processing, and speech analysis, these tasks have mainly focused on data that lie in Euclidean domains. The mathematical and computational tools underlying these techniques, such as convolution, resampling, multiscale decomposition, and locality, are well-defined and benefit from powerful computational hardware like GPUs and TPUs. However, many essential data and applications deal with non-Euclidean domains for which deep learning methods were not originally designed. Examples include 3D point clouds and 3D shapes in computer vision, functional MRI signals on the brain's structural connectivity network, the DNA of the gene regulatory network in genomics, drug design in quantum chemistry, neutrino detection in high-energy particle physics, and knowledge graphs for common-sense understanding of visual scenes. These opportunities and challenges motivated this IPAM long program, the main theme and goal of which was to learn, explore, and exploit geometric structures and features underlying the problem and to leverage them to more fully understand the data.

As a concrete example and specific application, the program started with 3D shape processing and analysis. 3D modeling is becoming ubiquitous due to fast acquisition and frequent use of 3D data, such as in laser structured light scanning, remote sensing, 3D printing, 3D cameras, 3D prototyping and other novel fabrication methods, virtual and augmented reality, to name a few. New technologies, such as depth sensors and 3D scanners integrated into smart phones and personal computers, make imminent the transition from 2D descriptions of the world, or images, to 3D. Compared to images, 3D shapes are geometric objects that are much more challenging to represent and analyze from both mathematical and computational perspectives. This makes higher-level tasks in 3D shape analysis and understanding, such as registration, comparison, recognition, classification, and learning, even more difficult. During the program and the first two workshops, recent progress and developments based on geometric theory, such as conformal and quasi-conformal maps, heat kernel and diffusion geometry, and Laplace-Beltrami spectral analysis, were reported and discussed. In particular, these approaches provide efficient computational techniques for extracting local and global intrinsic features and structures that are invariant under various transformations or embeddings. Moreover, recent advances in both supervised and unsupervised ML for data analysis can be used to develop robust and distinctive data features as well as in application-specific tasks such as recognition and classification. However, there are still significant knowledge gaps and questions that remain to be addressed. Combining mathematical understanding and tools for analyzing geometric surfaces with ML techniques will lead to more powerful and effective ways of training a computer to learn application-specific tasks.

The other major motivation and theme for this program was to learn and exploit geometry for big data beyond 3D. For example, a good mathematical metric can characterize similarities and distinguish dissimilarities. Simultaneously, features must be designed that remain invariant under certain transformations or group actions. When these features are used as input or desired properties are incorporated into learning structures and algorithms, the accuracy, efficiency, and interpretability of the entire process can be significantly enhanced. Another important task is to generalize neural networks to arbitrary geometric domains like graphs and manifolds. Fundamental operations such as convolution, coarsening, multi-resolution, and causality have been redefined through spectral and spatial approaches. Recent techniques for non-Euclidean data analysis show promising results for applications in many fields. During this program, various state-of-the-art methods for integrating geometry, modeling, and learning were investigated and presented. Important topics and techniques discussed and studied included construction of operations and network structures that possess desired properties, dimension reduction, manifold learning, metric design, and the information-geometry perspective of probabilistic and statistical properties.

Commensurately, there was one workshop on the geometry of big data and another on deep geometric learning of big data and applications.

The overall goal of this program was to bring together domain experts, including junior and senior researchers, with graduate students in mathematics, computer science, statistics, and data science to: (1) discuss and present the current state of their fields to facilitate new developments, (2) establish new research directions and applications in these fields, (3) build new connections and collaborations among participants, and (4) train young minds. Below is a more detailed summary of research activities and progress made during this program.

3D Shape Analysis

Introduction

Digitizing and replicating the world without losing the core geometric information has been the primary motivation for 3D shape modeling and analysis methods. Several applications such as 3D content generation, shape modeling, animation, and manufacturing create the demand for novel shape processing approaches. On the other hand, recent advances in deep learning (DL) architectures, especially in Convolutional Neural Networks (CNNs), have facilitated solutions to many difficult problems in the 2D domain. In this program, one of our main motivations was to adopt these advances for the 3D domain and to bridge the gap between traditional 3D shape analysis and DL methods.

3D Deep Learning for Shape Understanding

Several IPAM research directions were focused at the intersection of 3D shape analysis and DL. There is a vast variety of 3D shape representations with important properties extensively studied for traditional shape processing. For example, voxel sets have been widely used in 3D reconstruction due to their structured nature, which allows efficient occupancy and occlusion computations. Point clouds, on the other hand, can be converted to matrix representations for feature mining and statistical analysis. Meshes and triangle soups contain topological information that enables formulating mathematical properties for subspaces of shapes. Exploring these and developing new representations were some of the fundamental motivations for this program. These also establish a foundation for novel shape representations for DL.

Another motivation was to reformulate key operations (such as convolution and pooling, which are the backbone of network models) exercised in 2D DL settings to 3D DL architectures. 2D convolutions (and Fourier Transformations as their spectral correspondence) have been well-exercised for images, such as dilated, strided, and pyramidal convolutions. Going beyond images, 3D convolutions for voxels are similar to 2D convolutions with a cuboid kernel instead of a square kernel. However, for unstructured representations such as points, graphs, and meshes, novel operations like point convolutions, graph convolutions, and spectral convolutions are formulated, depending on some conventions of the underlying geometry.

In addition to these operations, the program explored novel loss functions in shape and metric spaces in accordance with the proposed representations and operations. For structured representations, distance metrics are easier to derive and formulate because the direction and neighborhood information is known. For unstructured representations, unit distances, connectivity, neighborhood, and ordering are all additional constraints to solve for. For example, Chamfer Distance (CD), used for evaluating similarity of point sets, is symmetric and invariant to the size of and permutations in point sets.

Apart from theoretical explorations discussed above, the participants of the program would also like to understand the relationship between 3D shapes and DL architectures from three more application-specific perspectives by investigating: (1) the role of shapes in traditional DL applications such as recognition, segmentation, registration, etc. with an emphasis on generative models, (2) sensor-fusion techniques (i.e., using images and point clouds for 3D geometry recovery, or learning shape properties jointly from the topological and geometric properties) to recover and emphasize shape information with joint- or transfer-learning approaches, and (3) using embedded 3D shape properties (i.e., planarity, angle preservation, manifoldness, curvature, class templates, etc.) to gain insight into complex DL systems.

Recent advances in geometric DL have shown promising results in a variety of challenging problems. Shape matching is an example where learning-based methods, in conjunction with geometric methods, perform with previously unattainable accuracies on existing datasets. In the case of supervised DL on clean and synthetic datasets, state-of-the-art methods achieve near-perfect accuracies. To explore these difficult shape-processing problems, the current research focus of the program is divided into three parts: (1) shape representations, (2) kernels and architectures, and (3) applications.

Shape Representations

To advance the geometric understanding of shape-analysis tasks, traditional as well as DL approaches were studied in the context of shape representations such as voxels, point clouds, graphs, and meshes. To fortify the mathematical foundations for 3D DL, the study of nonlinear shape representations was also carried out.

First, functional maps constitute a representation primarily used for shape matching and morphing operations. Defining a bijective mapping between two manifolds, this mapping induces a natural transformation of derived quantities, such as functions on the meshes. Having differentiable measures make functional maps suitable for DL. Second, Lie bodies is one of the domain-specific representations discussed during the program. It is a novel shape representation defining deformations directly on triangles, using a new Lie group of deformations. The similarity between meshes in this domain is easily measured on the manifold due to the Riemannian structure. Finally, derived or combined representations were explored and experimented upon, such as shape atlases (learnable parametrization of 2D square patches to cover a surface), structured implicit functions (parametrized spherical units to cover a volume), part templates (oriented and parametrized bounding boxes to cover all components of a shape), polycubes (like voxels, but textured and multi-scale with padding on the boundary), and grammar-based representations (rule-based parametrized collection of terminals). It has been observed that the parametrization and hierarchical nature of these representations is a key feature in generative models (e.g., variational autoencoders (VAEs) and generative adversarial networks (GANs)).

Kernels and Architectures

Kernels. Particular interest was directed toward the exploration of novel convolutional kernels and network architectures. Particularly interesting novel convolutions and kernels include pullback convolutions (defined on toric surfaces with a homeomorphism), edge convolutions (defined on the *k*-nearest neighbourhood graph of each point in a point cloud), graph convolutions (see Section "Graphs and Data), spectral smoothing filters (as in SplineNet), geodesic kernels (circular kernels in point clouds and graphs with different orderings), parametrized Gaussian kernels, and isotropic filters.

Architectures. In addition to kernels, novel network architectures were also explored such as string-based VAEs to encode tiling grammars, conditional mesh VAEs, foldingNet with a special point-cloud folding operator in the decoder, P2PNet to transform surface points to global points using a Siamese network with a geometric loss and cross regularization, PUNet with hierarchical point feature embeddings, CayleyNet as a spectral CNN with rational coefficients, PTCNN expressing convolution on the manifold as parallel transport, and various Graph Convolutional Networks (GCNNs), such as GCCN, ACNN, DynGCNN, MDCGNN, and MoNet.

Eigen-function spectra. Another interesting topic was the shape signatures obtained by Laplace-Beltrami (LB) spectra. It is proved that by taking the spectra of eigenvalues of its LB operator, or spherical harmonics, it is possible to compute a numerical signature of any 2D or 3D manifold. Because the spectrum is isometry invariant, it is independent of the object's representation including parametrization and spatial position. However the eigenfunctions in this spectra are not ordered, which led to many open questions in state-of-the-art LB-based 3D shape analysis approaches. Eigenfunction alignment, for example, matches the LB basis of two shapes using best or first *k* elements in the basis through minimization. Basis pursuit is a synthesis and analysis approach used in convolutional sparse modeling to find a layer-based family of filters. Basis synchronization is also used for the same purpose in Spectral Transformer Networks to overcome this issue. LB expects the manifolds to be isometric, so isometric manifold deformation is another problem that can be explored by isospectralization. Various regularization techniques and generalizations to non-manifold shapes were also discussed.

Graph pooling. The equivalent of spatial pooling operation for CNNs is defined as graph pooling for GCNNs, which requires coarsening the graph. As this problem corresponds to graph partitioning, traditional methods as edge collapsing, heavy edge matching, and balanced cuts are used as graph-pooling operations. This works fine for structured graphs with indexing; however, it becomes computationally expensive for dynamic graphs because re-indexing is needed after each coarsening.

Loss functions. Finally, novel loss functions (similar to Earth Mover's Distance (EMD) and CD for points) that combine novel kernels and architectures were studied. Soft error loss, for example, is measured on a (bijective) soft correspondence matrix with geodesic distances between Riemannian manifolds of two shapes.

Applications

3D shape analysis has numerous applications across various industries, several of which were discussed during the program. Shape-matching and registration approaches have been widely explored in CT and MRI analysis in healthcare. 3D simulations in VR are also useful for patient education, surgery simulation, and telemedicine. Digitization of 3D shapes, animation, motion capture, and character design in the entertainment industry, and augmented reality and virtual reality (AR/VR) applications are in high demand. Popular topics in discussion that are directly related to the program included point cloud registration, segmentation, and classification approaches used in: (1) Simultaneous Localization and Mapping (SLAM) and semantic inference on 3D data, (2) perception systems in modern robotic systems and in autonomous/assistive driving, and (3) shape morphing for novel shape synthesis and manufacturing. Applications of 3D shape generators were also discussed for facial animation, content generation, and visual question answering. As both the scientific and industrial focus gradually moves from "2D" to "3D," 3D shape analysis is gaining more attention and more efforts are needed in the aforementioned domains.

Geometry for Deep Learning

Taking a step back from the specific research topics outlined in the previous sections, some general open questions were posed for future discussions. One question emerging from the program was how to use geometry for understanding machine learning, in contrast to using ML for understanding geometry. For example, in a 2D setting, visualizing outputs of CNN layers has provided new insights into ML. However, to what extent applications involving 3D shapes can provide a deeper understanding of ML is unclear. Formulations around strong shape priors and shape properties are expected to play a crucial role in understanding complex neural networks. Understanding the intrinsic dimensionality of geometry objects will potentially decide the size of learning systems.

Holistic and Scalable 3D Deep Learning Systems

Another future direction is to embed shape processing algorithms in low-power devices with memory and computational constraints. Processing 3D shapes is a complex task due to its increased dimensionality, i.e., voxel nets are greatly restricted in resolution and by computational power because of 3D kernels and the discretization process. Another potential future direction is to develop holistic shape understanding systems where the analysis goes beyond 3D geometry to include texture, color, material, and semantic information. To form an integrated DL system with feature-specific layers using all information in the most suitable way is a major challenge. Joint learning methods can be introduced to combine extracting descriptive features from those non-geometric properties of the data or to learn different embeddings of shapes that contain those features implicitly.

Expressive Shape Representations: From Procedural to Deep Generative Models

Structured content generation has been an important topic in AR/VR, games, simulations, animation, and architecture. Traditional approaches employ manual modeling, content retrieval from massive model databases, or procedural modeling (PM - the process of generating geometry using a grammar). However, these approaches are costly, need domain expertise, have limited generative power, and lack control over creation.

Proceduralization and inverse procedural modeling (IPM) algorithms have been proposed to overcome the aforementioned drawbacks. IPM is an optimization over a procedural representation given a target model. The derivation space of the grammar becomes the search space for IPM, looking for the derivation that best fits the target 3D model. Proceduralization, on the other hand, extracts the grammar directly from the geometry without any assumption on the underlying grammar. Both of these approaches incorporate several shape analysis, geometry processing, and optimization algorithms for different shape representations. Proceduralization outputs shape grammars, which are hierarchical and parameterized descriptions of 3D shapes, objects, or scenes.

On the other hand, deep generative models for 3D shapes exist for images, voxels, point clouds, and surface meshes, outputting shapes in the representation of the input domain or in the latent space, both of which are not suitable for controlled shape synthesis tasks. In conjunction with the main motivations of this program, those networks were studied to define and obtain expressive shape representations, posing the question: *Can we formulate and learn an explicit or implicit shape representation that mimics procedural models, following their parametric and hierarchical nature?*

The first step toward formulating such representations for DL is proposed as a global shape understanding problem: learning skeletons from shapes as parametric representations. Although these skeletons satisfy the parametrization requirement, other requirements such as hierarchical subdivision and capturing patterns as rules remain unfulfilled.

Overarching open questions to model novel network architectures that learn expressive shape representations can be encapsulated as: (1) how to embed explicit grammars as implicit differentiable features, (2) how to jointly learn terminals (geometric unit structures) and rules (patterns and distributions), and (3) how to formulate loss functions to maximize the generative power of the model.

Beyond 3D: Convolution on Manifold via Isometric Embeddings in Higher Dimensions

There are several current 3D CNN architectures that use data embedded in Euclidean 3D space such as VoxNet, V-Net, PointNet, and VoxelNet. An architecture that incorporates normal directions to a voxel object is NormalNet. To reduce memory requirements and improve performance, sparsification of 3D data is implemented in OctNet. Submanifolds have been used in a CNN architecture called SparseConvNet, which has linear cost with respect to the number of active sites, with considerable computational economies even while maintaining state-of-the-art performance.

A proposal to extend these ideas to higher dimensions was developed and presented during this long program. This direction incorporates methods from high-dimensional differential geometry, namely, isometric embeddings and the notion of reach of an embedding. In this way, spatial convolution can be defined and generalized to deal with geometric structures on higher dimensional manifolds.

Historically, the theoretical guarantee of finding an isometric embedding of a smooth, closed Riemannian manifold was first solved by John Nash. Other useful strategies realized embeddings into L^2 and recent advances improved these approaches using heat kernels and eigenvector fields of connection Laplacians.

Implementations of these theoretical ideas involve, among many others, eigenvalues of the LB operator, nearly isometric embeddings via relaxation, using KDE and local PCA, and considering strengthenings of Whitney embeddings to produce almost isometric embeddings. There is a significant trade-off when lower dimensions are used for the target space of the embedding. Relaxation conditions allow for embeddings into finite-dimensional normed spaces, however, here the embedding dimension grows exponentially with respect to the manifold's intrinsic dimension, d. Using a Nash-type embedding, the ambient dimension grows quadratically in d.

A major challenge in this direction is to leverage the theoretical advantages afforded by these isometric embedding techniques with the computational cost of sweeping d-dimensional tensors spatially, as these convolutions are defined in the ambient space.

Graphs and Data

Introduction

CNNs have been extremely successful in many high-dimensional regression and classification tasks on Euclidean domains. Recently, several generalizations to graph structured data have attracted increased attention due to their potential for use in pattern recognition for extremely large-scale problems with minimal assumptions on the data. Properly defining graph convolution (or a convolution-like operation) is a fundamental ingredient in formulating these networks. Over the past six years there have been several new approaches and many further refinements to achieve this. Another key problem in understanding Graph Neural Networks (GNNs) is understanding the generation of the underlying graphs that the networks work on. Beyond these challenges, critical insights have been gained by considering Cartesian grids and triangulated meshes as special cases of graphs with regular connectivity and edge weights. Investigating how to adapt elements of classical signal processing from these regular graphs into more general ones has been and will continue to be an important research avenue.

Applied Harmonic Analysis on Graphs

Graphs are particularly powerful and flexible in capturing irregularity at both global and local scales. While many operator techniques (such as multiscale transforms) are well understood on regular grids, a big open question is their extension to irregular structures such as graphs. On the other hand, the impact of explicit graph construction on the properties of such generalized objects is thoroughly understood. The irregular structure of some given graph so far also introduces ambiguity in how to choose, for specific purposes, a suitable definition of distance in the vertex domain or a favorable basis of the function space used for signals on graphs.

We expect the field of spectral graph theory to significantly mature with a deeper understanding of these issues and the tradeoffs between desired properties in the vertex domain and in the graph spectral domain. Moreover, sometimes it is either impossible to directly translate a numerical scheme available in a Euclidean space, or on a regular grid to the general graph setting, or the resulting counterpart method is significantly more expensive to compute. Hence, to make spectral approaches to graph problems feasible and attractive, **r**esearch is still needed for devising efficient (approximate) numerical methods.

Spectral Based Graph Neural Networks

Directly defining spatial convolution is extremely difficult on graphs because it is not obvious how to define a local filter on different vertices with different connectivities that have the desired weight-sharing and shift-equivariant properties. Spectral approaches avoid this by using the graph Fourier transform to define a convolution in the spectral domain. The standard basis is composed of the eigenfunctions of the graph Laplacian. The locality of filters on the spatial domain can be enforced by requiring the filter to be smooth in the frequency domain. Additionally, the eigensystem can be truncated to reduce computational costs. Recently, approaches based on other graph transforms such as the Haar transform have also been considered in this framework. The Haar basis is a sparse and localized orthonormal system for graph-function spaces built on a coarse-grained chain of the graph under which the graph convolution is defined accordingly for GNNs. The sparsity and locality of the Haar basis allow Fast Haar Transforms on graphs by which an efficient evaluation of Haar convolution between the graph signals and the filters can be achieved.

Spectral approaches are particularly promising because they connect the traditional fields of graph theory, harmonic analysis, and operator theory, all of which have been studied extensively and have yielded profound insights on their own. Thus, we expect a variety of improvements from further exploration and justification of connections between these fields, which can be formulated using a graph representation.

Spatial-Based Graph Neural Networks

Locally connected networks, sometimes called message-passing networks, have been developed as an alternative approach to defining convolution-like operators on graphs. In these techniques, an adjacency matrix is used as a mask for a traditional fully connected layer. Then for each layer, at each node the hidden state is updated by an accumulation of the neighboring states. Several techniques for this accumulation operation have been proposed, but the most common are various weighted averages with the weights coming from the adjacency matrix. Recently, several authors have proposed ways to learn these accumulations as discussed in the next section.

Graph Attention Mechanism

Nowadays, the attention mechanism is one of the most important tools in ML. It was first introduced in the field of machine translation and has since been used in vision applications. In the standard message-passing framework, accumulation weights are fixed, but to achieve more powerful representations, an attention mechanism has been introduced.

The graph attention mechanism, on the other hand, can be understood as a function that can be used to determine the connecting weight between nodes in a given graph based on the hidden state of each node. In most cases, this function is modeled by a neural network that is learned within the overlying model. Here, the overlying model is a GNN, which aggregates information from neighboring nodes. An important question is: *Is a particular connection in a graph important for a given task?* The attention mechanism answers this question by predicting a weight for this connection. Therefore, by using the attention mechanism, we learn a new weighted adjacency matrix.

This procedure assumes that the structure of the graph has been given, but what if no graph structure is available? The graph structure could be formulated by learning the attention weights for a fully connected graph, but this is extremely computationally expensive. Moreover, the higher the connectivity of the graph, the more expensive the computation of the attention mechanism.

Graph Generation

In some applications, such as social network analysis, the underlying graph is readily available. However, in other fields such as point cloud registration or medical imaging, obtaining this graph is non-trivial. Currently, the standard method to connect a set of data points into a graph is to define some metric on the data space and perform either a *k*-nearest neighbor search (to connect each data point to its *k* nearest neighbors) or a range search (connect each data point to all other points within some radius). The choice of metric is critical because choosing different distance measures yields significantly different graphs. For image data, there are well-studied metrics such as Euclidean pixel distance, cosine distance, and EMD, but this is not generally true, especially when the data are heterogeneous. Because attention networks only estimate attention coefficients for nodes within a fixed number of hops, selecting a good initialization is vital.

Recently, there have been a few approaches proposed to learn a graph generation model — all of which have important contributions, none of which is perfect. Methods based on sampling adjacency matrices from Bernoulli random variables are computationally expensive and do not currently allow for the user to specify properties of the generated graphs (such as connectedness). Methods based on auto-encoder frameworks require large training sets. In cases where the “best” graph representation of the data is unknown, it is unlikely that there will be a training set. In the future, we believe that graph generation will become an even more important part of the graph analysis pipeline and we anticipate that the study of such generative models will be an important cornerstone of future research in this field.

Generalization of GNNs and Applications

Because GNNs can learn representations of nodes, edges, subgraphs, and whole graphs, there is a connection between the neural graph representation learning and the field of dimensionality reduction. Most dimensionality reduction models build a neighborhood graph of data points, which is often weighted and then implicitly used to embed data points in lower-dimensional space. GNNs learn lower-dimensional representations of graphs (or subcomponents) explicitly, often with less computational cost than classical dimensionality-reduction algorithms that rely on matrix-factorization approaches. Skip-gram-like models, which are shallow neural networks, embed nodes of the graph based on sampled random walks. It would be interesting to characterize how biased random walks, which account for various geometric and topological properties of the graph and the manifold it triangulates, affect the embeddings.

Another promising field of research is generalizing GNNs to more general and higher-order structures such as hypergraphs and simplicial complexes. Here, we anticipate the incorporation of tools from topological data analysis into geometrically motivated neural network architectures will be an active and fruitful area of research in the near future. An important question is how to apply persistent homology or combine it with other statistics/analysis/differential geometry methods to measure and improve the performance of GNNs. This also involves subjects from optimal transport and Wasserstein information geometry.

Optimal Transport / Wasserstein Information Geometry

Introduction

Optimal transport (OT) provides a geometric framework for the study of probability distributions by extending the geometry of the sample space. OT can answer questions such as how far away two distributions are from one another, what an average probability distribution looks like, and whether two distributions are in the same direction relative to a vantage point. Such framework fits ML well, as learning can often be cast as minimizing a function of distance between a model and a data population. In this case especially, OT facilitates learning from probability distributions. The defining property of OT distances is their ability to consider the geometry of the sample space upon which the distributions are defined. This is in contrast to information theoretical divergences that do not take the geometry of sample space into account. A particular example of OT distances is given by the Wasserstein metrics, which extend distance functions of sample spaces to distances between distributions.

Information geometry takes a geometric perspective to learning. The typical structures here are information divergences and the Fisher-Rao metric, which plays the role of a Riemannian metric. This well-established field has found multiple applications in statistics and ML. One aim of information geometry is to study the loss of information via geometric objects.

Wasserstein information geometry bridges these two approaches by studying the questions arising in information geometry through constructions in OT. It develops new theories, methods, and applications, which incorporate aspects of both data space and model space, in a way that is not possible with other approaches. This area contributes to addressing the current deficit and increasing demand for mathematically sound approaches in ML.

Optimal Transport in ML Models

Geometry of Data via Optimal Transport. Learning, or generalizing from examples, amounts to discovering regularities from data and discarding noise. The amount of data and computations required to accomplish this is strongly affected by the amount of prior knowledge that can be incorporated into the learning algorithm. OT can define neighborhoods in data space of natural images, which correlate well with human perception of semantic vicinity. This can be used to define generative models that better reflect the natural variability of data. During the IPAM program, we discussed implementations of such approaches in state-of-the-art generative models called GANs. Another application can be found in the training of discriminative models that are robust to in-class data variability. This is particularly important in the context of adversarial robustness. During the program, the core participants submitted articles on these topics and had abstracts accepted in major conferences including International Conference on Machine Learning and were invited for workshop presentations at Computer Vision and Pattern Recognition.

Geometry of Hypothesis Space via Optimal Transport. Besides the geometry of data, Information Geometry, the geometry of the space of hypotheses, is also important. Local optimization strategies over hypothesis spaces are strongly affected by their underlying geometry. OT can define geometric structures on hypothesis spaces, which can be used to determine the direction of steepest improvement (depending on how distances are measured in hypothesis space). During the program, several such papers were submitted to important conferences and special journal issues including Geometric Science of Information (GSI), Deep Learning Information Geometry, and Deep Learning Theory-MPI.

Computational Aspects

Learning is often formulated as the minimization of a discrepancy measure between an observed behavior and a model. The quality of a solution depends on the selected discrepancy measure, which can be defined using OT. In practice, such measures are evaluated only approximately. For example, one considers iterative optimization methods where at each iteration the loss function is evaluated over only a subset of the data (stochastic gradient descent) or where only an approximation is considered (e.g., factorized versions of a Riemannian metric).

Wasserstein GANs are popular examples of OT applied using stochastic gradient descent in the domain of big data. For example, the goal might be to learn from an empirical distribution, such as a population of pictures of celebrities. Wasserstein GANs can estimate the OT distance between the generator and the target distribution through dual formulation of OT, which can be cast as maximizing the expected value of the sum of two functions called Kantorovich potentials. However, these potentials must satisfy certain non-trivial conditions, which impose the greatest challenges for OT in GANs. In the current literature, constraints are heuristically incorporated and validation of the resulting OT distances is lacking.

Finally, in practice we are only able to consider a subset of candidates for the optimal Kantorovich potentials consisting of neural networks such as multilayer perceptrons or CNNs. This poses two questions: (1) *How well these families are approximating the real OT distance?* (2) *Is the OT distance optimal for learning a generator in the GAN setting or could the restriction to specific function classes improve the discriminator?*

The entropic regularization of OT allows us to take advantage of computational power without compromising theoretical properties of the Wasserstein distance. Indeed, it allows us to define the so-called Sinkhorn divergence, which metrizes the weak-* topology. In the context of GANs, the Kantorovich formulation of the entropic transport problem defines an unconstrained optimization problem that resolves the above-mentioned difficulties of satisfying the constraints on the Kantorovich potentials.

Further generalizations will be investigated, including the extension of this approach in the multi-marginal case, i.e., an OT problem with more than two marginals.

Connection with Other Traditional Statistics/ML Methods

It would be useful to find connections between the geometrical structures, such as the distances between probability distributions, obtained in the framework of OT and those obtained in other mathematical areas such as information geometry and information theory. From a practical viewpoint, this can potentially lead to new computational methods where the distances are optimally tuned for a particular application.

For example, through entropic relaxation of OT, Sinkhorn divergences can be defined. Sinkhorn divergences provide an intriguing connection between OT distances and MMD measures because varying the magnitude of relaxation interpolates between the two. As another example, in Procrustes analysis, a parametrized family of Riemannian metrics and distances can be defined to unify the Wasserstein-Riemannian metric and distance with the Log-Euclidean Riemannian metric and distance in the Gaussian case, both in the finite- and infinite-dimensional settings (work accepted at GSI 2019). Two intriguing research directions along this line include: (1) extension of the previous formulations to the non-Gaussian setting and (2) unification of the Wasserstein-Riemannian metric and distance with the Fisher-Rao metric and distance in information geometry.

We have also explored the connections between the geometrical structures from OT with other traditional ML methods, such as kernel methods. As an example, we have obtained closed-form formulae for the Wasserstein-Riemannian distance between *reproducing kernel Hilbert spaces* covariance operators, potentially leading to new kernel algorithms employing these operators.

Optimal Transport in Computer Graphics

OT has been applied to: (1) geometry processing for computing soft maps between meshes or shape interpolation and (2) rendering for the design of bi-directional reflectance distribution functions. It can also be applied to physical simulation for generating vector field interpolations between simulated frames. Defining the ground metric on triangular meshes for the Wasserstein distance is a challenging problem. Computing machine learning Wasserstein isometric embeddings using ML methods is being explored to compute a distance between meshes more accurately and efficiently. More work needs to be done in computing the Wasserstein distance on vector fields of physical simulations and applying OT to generate interpolation frames and to perform latent space simulations.

Applications and Practical Matters

Introduction

Further theoretical and mathematically based analysis is required to describe, explain, and predict the performance of ML models. This is vital for decision making because if life-altering decisions are to be assisted with ML models, there must be verified and validated confidence in the output estimates and predictions (e.g., ML-assisted medical diagnosis and treatment).

Open Research Topics

Neural Machine Translation

In the long program, we discussed ML in natural language processing (NLP). ML has been used in NLP, especially for translation. Although the progress on neural machine translation has been impressive, there remains a clear gap between neural machine translation and human translation. Based on domain expertise in translation of the program participants, several options for improving neural machine translation were identified: (1) analysis of tone and emotion, (2) paraphrase translation, (3) characterizing the environmental context, (4) event-based context understanding, (5) underlying intent of the language, (6) relationships between speakers (e.g., hierarchical), and (7) multi-skill aspects (speaking, writing, and listening).

In contemporary neural machine translation, the most successful ML models include: RNNs, attention-based transformers, and combinations of knowledge graphs and reinforcement learning. Semantic-position-prediction-based NLP ML models have difficulties understanding word contexts; however, in combination with a knowledge-graph model, stable representations of words can be produced. In the future, there is the exciting potential of using knowledge graphs leveraged with semantic position prediction to develop a system to perform human-level translation.

ML in Medical Fields

Another scientific arena wide open to ML is medicine, which was a major topic in the long program. However, most current ML models and architectures are not readily suited for medical applications. Major challenges identified by the program members include: (1) lack of ML models for small datasets, especially those suited for transfer learning, (2) lack of high-quality data (medical images are often noisy or corrupted), (3) memory limitations (because of high-resolution 3D medical images), (4) precision requirements (sub-voxel accuracy is often needed), (5) verification and validation (V&V) of ML models, and (6) continuously learning ML models. For example, most ML models do not work well with small datasets; however, due to privacy concerns, regulations, costs, and incidents of disease, only relatively small datasets are available in many medical applications. Although transfer learning has been used to improve model development (sometimes to a surprising degree), most image-processing models are pre-trained on natural rather than medical images. ML models pre-trained solely on medical images focusing on certain organs or specific diseases could improve predictive performance.

V&V of all ML models, especially those in medical applications, has significant implications for governmental agencies, physicians, and relevant stakeholders who evaluate AI applications in medicine. A continuously learning model is an ML model in which an algorithm is able to adapt to new data. This situation is common in the medical arena because patient healthcare situation changes over time. A more fully developed continuously learning ML model may have the potential of performing on-par with human doctors who continuously adapt diagnoses according to the ever-changing health situation of the patient.

Social Impacts of ML

In the long program, state-of-the-art facial recognition methods in the form of reconstruction of 3D shapes from pictures, parametrizations using curvature flows, and image recovery from occluded images were presented. Also, ML applications in NLP and other fields were extensively discussed. As ML systems become evermore prevalent in everyday life and infrastructure, program participants were concerned about the social impacts of adversarial attacks, privacy , anonymity, and abuse. Some of the issues discussed during the sessions were:

(1) **Abuse of NLP**. OpenAI's GPT-2 transformer network was trained with approximately 1.5 billion parameters on a corpus with over 40 Gigabytes of text. Given a prompt, GPT-2 can synthesize paragraphs of coherent, realistic text which matches the style and content of the prompt. The full model was originally deemed too risky to release publicly due to potential for misuse, e.g., the software can write blogs and press releases that could be used to spread misinformation.

(2) **Risks from Visual Recognition Technology**. A nefarious agent could subvert facial recognition systems to gain unauthorized access to confidential information. Moreover, unexpected inputs can also confuse ML models as they can only be expected to perform properly on data with a distribution similar to that of the

training set. In a recent fatal accident, an autonomous car crashed into a truck trailer, which it confused for a harmless overhead road sign.

(3) **Flawed Data-collection Pipeline**. Supervised ML requires labeled training data, which are usually crowd-sourced. Malicious agents can inject purposefully-incorrect labels into the dataset and disrupt the training process. More alarmingly, even without having control of the labels, agents can poison training data to cause the model to fail at test time. Recent work demonstrated that training data can be reverse-engineered from a model, potentially violating privacy rights. For example, faces of individuals can be recovered from facial recognition models. *How can ML models be improved to defend against such attacks?*

(4) **Economic Impacts of ML**. As AI systems mature, there is growing concern that they will render some existing jobs obsolete. For example, 25% of all jobs in the US are based at least partially on the transportation sector. *How will autonomous vehicles displace these employees? Can it be avoided?* Alternatively, AI will create new jobs. *But will these new jobs sufficiently offset those which they replaced?* For example, displaced workers may lack the necessary skills and experience to perform the duties of the newly-created jobs. *Can policy be enacted to help?*

These topics were discussed throughout the program and the participants agree that these issues must be continuously considered as ML models advance. Additional issues will certainly emerge and it is important to remain vigilant.