

成果一.

闫桂英研究员与合作者的论文 Affine Automorphism Group of Polar Codes 被 IEEE TRANSACTIONS ON INFORMATION THEORY 接收发表。

摘要: The automorphism ensemble (AE) decoding framework for polar codes attracts much attention recently. It decodes multiple permuted codewords with successive cancellation (SC) decoders in parallel and hence has lower latency compared to successive cancellation list (SCL) decoding. However, the AE decoding framework is ineffective for permutations falling into the lower-triangular affine (LTA) automorphism group, as they are invariant under SC decoding. Therefore, the block lower-triangular affine (BLTA) group was discovered to achieve better AE decoding performance. However, the equivalence of the BLTA group and the complete affine automorphism group was unresolved. Additionally, some automorphisms in BLTA group are also SC-invariant, thus are redundant in AE decoding. In this paper, we prove that BLTA group coincides with the complete automorphisms of decreasing polar codes that can be formulated as affine transformations. Also, we find a necessary and sufficient condition related to the block lower-triangular structure of transformation matrices to identify SC-invariant automorphisms. Furthermore, We present an algorithm that efficiently identifies all SC-invariant affine automorphisms under specific constructions.

论文链接: <http://dx.doi.org/10.1109/TIT.2024.3429277>

成果二.

张波研究员与合作者的论文 A KERNEL MACHINE LEARNING FOR INVERSE SOURCE AND SCATTERING PROBLEMS 被 SIAM JOURNAL ON NUMERICAL ANALYSIS 接收发表。

摘要: In this work we connect machine learning techniques, in particular kernel machine learning, to inverse source and scattering problems. We show the proposed kernel machine learning has demonstrated generalization capability and has a rigorous mathematical foundation. The proposed learning is based on the Mercer kernel, the reproducing kernel Hilbert space, the kernel trick, as well as the mathematical theory of inverse source and scattering theory, and the restricted Fourier integral operator. The kernel machine learns a multilayer neural network which outputs an ϵ -neighborhood average of the unknown or its nonlinear transformation. We then apply the general architecture to the multifrequency inverse source problem for a fixed observation direction and the Born inverse medium scattering problem. We establish a mathematically justified kernel machine indicator with demonstrated capability in both shape identification and parameter identification, under very general assumptions on the physical unknowns. More importantly, stability estimates are established in the case of both noiseless and noisy measurement data. Of central importance is the interplay between a restricted Fourier integral operator and a corresponding Sturm--Liouville differential operator. Several numerical examples are presented to demonstrate the capability of the proposed kernel machine learning.

论文链接: <http://dx.doi.org/10.1137/23M1597381>

成果三.

张世华研究员与合作者的论文 Statistical batch-aware embedded integration, dimension reduction, and alignment for spatial transcriptomics 被 BIOINFORMATICS 接收发表。

摘要: Motivation: Spatial transcriptomics (ST) technologies provide richer insights into the molecular characteristics of cells by simultaneously measuring gene expression profiles and their relative locations. However, each slice can only contain limited biological variation, and since there are almost always non-negligible batch effects across different slices, integrating numerous slices to account for batch effects and locations is not straightforward. Performing multi-slice integration, dimensionality reduction, and other downstream analyses separately often results in suboptimal embeddings for technical artifacts and biological variations. Joint modeling integrating these steps can enhance our understanding of the complex interplay between technical artifacts and biological signals, leading to more accurate and insightful results. Results: In this context, we propose a hierarchical hidden Markov random field model STADIA to reduce batch effects, extract common biological patterns across multiple ST slices, and simultaneously identify spatial domains. We demonstrate the effectiveness of STADIA using five datasets from different species (human and mouse), various organs (brain, skin, and liver), and diverse platforms (10x Visium, ST, and SliceseqV2). STADIA can capture common tissue structures across multiple slices and preserve slice-specific biological signals. In addition, STADIA outperforms the other three competing methods (PRECAST, fastMNN, and Harmony) in terms of the balance between batch mixing and spatial domain identification, and it demonstrates the advantage of joint modeling when compared to STAGATE and GraphST.

论文链接: <http://dx.doi.org/10.1093/bioinformatics/btae611>

成果四.

骆顺龙研究员与合作者的论文 Characterizing bipartite states with vanishing basis-dependent correlations 被 PHYSICAL REVIEW A 接收发表。

摘要: Since both coherence and quantum correlations arise from the superposition principle and can be regarded as resources in quantum information tasks, it is of significance to investigate the interplay between them from different perspectives. In this work we focus on the basis-dependent correlations in a bipartite state defined by the coherence difference between global state and local state relative to a local basis and characterize bipartite states with vanishing basis-dependent correlations. Using the relative entropy of coherence, the structure of such states has been determined by Yadin et al. [Phys. Rev. X 6, 041028 (2016)], which we call block-diagonal product states here. The first result of this work is to demonstrate that the set of block-diagonal product states can also be characterized by the property of possessing vanishing basis-dependent correlations using the coherence measure based on skew information. As a by-product

of this result, we describe the structure of quantum ensembles saturating the convexity inequality in the resource theory of coherence using the coherence measure based on skew information, which may be of independent interest. Next, we characterize the set of bipartite states with vanishing basis-dependent correlations using the l_1 norm of coherence, and show that the set of block-diagonal product states is a subset of it. Furthermore, we provide an operational interpretation of block-diagonal product states in an interference model. Finally, we compare the amount of basis-dependent correlations using the three mentioned coherence measures through several examples such as Werner states, isotropic states, Bell-diagonal states, and a family of classical-quantum states.

论文链接: <http://dx.doi.org/10.1103/PhysRevA.110.022418>

成果五.

杨钊副研究员与合作者的论文 Existence and Stability of Nonmonotone Hydraulic Shocks for the Saint Venant Equations of Inclined Thin-Film Flow 被 ARCHIVE FOR RATIONAL MECHANICS AND ANALYSIS 接收发表。

摘要: Extending the work of Yang-Zumbrun for the hydrodynamically stable case of Froude number $F < 2$, we categorize completely the existence and convective stability of hydraulic shock profiles of the Saint Venant equations of inclined thin film flow. Moreover, we confirm by numerical experiment that asymptotic dynamics for general Riemann data is given in the hydrodynamic instability regime by either stable hydraulic shock waves, or a pattern consisting of an invading roll wave front separated by a finite terminating Lax shock from a constant state at plus infinity. Notably, profiles, and existence and stability diagrams, are all rigorously obtained by mathematical analysis and explicit calculation.

论文链接: <http://dx.doi.org/10.1007/s00205-024-02033-4>

成果六.

马志明研究员与合作者的论文 Pre-training with fractional denoising to enhance molecular property prediction 被 NATURE MACHINE INTELLIGENCE 接收发表。

摘要: Deep learning methods have been considered promising for accelerating molecular screening in drug discovery and material design. Due to the limited availability of labelled data, various self-supervised molecular pre-training methods have been presented. Although many existing methods utilize common pre-training tasks in computer vision and natural language processing, they often overlook the fundamental physical principles governing molecules. In contrast, applying denoising in pre-training can be interpreted as an equivalent force learning, but the limited noise distribution introduces bias into the molecular distribution. To address this issue, we introduce a molecular pre-training framework called fractional denoising, which decouples noise design from the constraints imposed by force learning equivalence. In this way, the noise becomes customizable, allowing for incorporating chemical priors to substantially improve the molecular distribution modelling. Experiments

demonstrate that our framework consistently outperforms existing methods, establishing state-of-the-art results across force prediction, quantum chemical properties and binding affinity tasks. The refined noise design enhances force accuracy and sampling coverage, which contribute to the creation of physically consistent molecular representations, ultimately leading to superior predictive performance.

论文链接: <http://dx.doi.org/10.1038/s42256-024-00900-z>

成果七.

夏旭助理研究员的论文 Fate of localization features in a one-dimensional non-Hermitian flat-band lattice with quasiperiodic modulations 被 NEW JOURNAL OF PHYSICS 接收发表。

摘要: We investigate the influence of quasiperiodic modulations on one-dimensional non-Hermitian diamond lattices with an artificial magnetic flux θ that possess flat bands. Our study shows that the symmetry of these modulations and the magnetic flux θ play a pivotal role in shaping the localization properties of the system. When $\theta = 0$, the non-Hermitian lattice exhibits a single flat band in the crystalline case, and symmetric as well as antisymmetric modulations can induce accurate mobility edges. In contrast, when $\theta = \pi$, the clean diamond lattice manifests three dispersionless bands referred to as an 'all-band-flat' (ABF) structure, irrespective of the non-Hermitian parameter. The ABF structure restricts the transition from delocalized to localized states, as all states remain localized for any finite symmetric modulation. Our numerical calculations further unveil that the ABF system subjected to antisymmetric modulations exhibits multifractal-to-localized edges. Multifractal states are predominantly concentrated in the internal region of the spectrum. Additionally, we explore the case where θ lies within the range of $(0, \pi)$, revealing a diverse array of complex localization features. Finally, we propose a classical electrical circuit scheme to realize the non-Hermitian flat-band chain with quasiperiodic modulations.

论文链接: <http://dx.doi.org/10.1088/1367-2630/ad7529>

成果八.

骆顺龙研究员与合作者的论文 Quantum resource theory of coding for error correction 被 PHYSICAL REVIEW A 接收发表。

摘要: Error-correction codes are central for fault-tolerant information processing. Here we develop a rigorous framework to describe various coding models based on quantum resource theory of superchannels. We find, by treating codings as superchannels, a hierarchy of coding models can be established, including the entanglement assisted or unassisted settings, and their local versions. We show that these coding models can be used to classify error-correction codes and accommodate different computation and communication settings depending on the data type, side channels, and pre-/postprocessing. We believe the coding hierarchy could also inspire new coding models and error-correction methods.

论文链接: <http://dx.doi.org/10.1103/PhysRevA.110.032413>

成果九.

张世华研究员与合作者的论文 GeneCompass: deciphering universal gene regulatory mechanisms with a knowledge-informed cross-species foundation model 被 CELL RESEARCH 接收发表。

摘要: Deciphering universal gene regulatory mechanisms in diverse organisms holds great potential for advancing our knowledge of fundamental life processes and facilitating clinical applications. However, the traditional research paradigm primarily focuses on individual model organisms and does not integrate various cell types across species. Recent breakthroughs in single-cell sequencing and deep learning techniques present an unprecedented opportunity to address this challenge. In this study, we built an extensive dataset of over 120 million human and mouse single-cell transcriptomes. After data preprocessing, we obtained 101,768,420 single-cell transcriptomes and developed a knowledge-informed cross-species foundation model, named GeneCompass. During pre-training, GeneCompass effectively integrated four types of prior biological knowledge to enhance our understanding of gene regulatory mechanisms in a self-supervised manner. By fine-tuning for multiple downstream tasks, GeneCompass outperformed state-of-the-art models in diverse applications for a single species and unlocked new realms of cross-species biological investigations. We also employed GeneCompass to search for key factors associated with cell fate transition and showed that the predicted candidate genes could successfully induce the differentiation of human embryonic stem cells into the gonadal fate. Overall, GeneCompass demonstrates the advantages of using artificial intelligence technology to decipher universal gene regulatory mechanisms and shows tremendous potential for accelerating the discovery of critical cell fate regulators and candidate drug targets.

论文链接: <http://dx.doi.org/10.1038/s41422-024-01034-y>

成果十.

李梅博士后与合作者的论文 Sample Average Approximation for Conditional Stochastic Optimization with Dependent Data 被 International Conference on Machine Learning, ICML 2024 接收发表。

摘要: Conditional Stochastic Optimization (CSO) is a powerful modelling paradigm for optimization under uncertainty. The existing literature on CSO is mainly based on the independence assumption of data, which shows that the solution of CSO is asymptotically consistent and enjoys a finite sample guarantee. The independence assumption, however, does not typically hold in many important applications with dependence patterns, such as time series analysis, operational control, and reinforcement learning. In this paper, we aim to fill this gap and consider a Sample Average Approximation (SAA) for CSO with dependent data. Leveraging covariance inequalities and independent block sampling technique, we provide theoretical guarantees of SAA for CSO with dependent data. In particular, we show that SAA for CSO retains asymptotic consistency and a finite sample guarantee under mild

conditions. In addition, we establish the sample complexity $O(d/E^4)$ of SAA for CSO, which is shown to be of the same order as independent cases. Through experiments on several applications, we verify the theoretical results and demonstrate that dependence does not degrade the performance of the SAA approach in real data applications.

论文链接: <https://openreview.net/pdf?id=YuGnRORkJm>

成果十一.

周川副研究员与合作者的论文 Graph Neural Stochastic Diffusion for Estimating Uncertainty in Node Classification 被 International Conference on Machine Learning, ICML 2024 接收发表。

摘要: Graph neural networks (GNNs) have advanced the state of the art in various domains. Despite their remarkable success, the uncertainty estimation of GNN predictions remains under-explored, which limits their practical applications especially in risk-sensitive areas. Current works suffer from either intractable posteriors or inflexible prior specifications, leading to sub-optimal empirical results. In this paper, we present graph neural stochastic diffusion (GNSD), a novel framework for estimating predictive uncertainty on graphs by establishing theoretical connections between GNNs and stochastic partial differential equation. GNSD represents a GNN-based parameterization of the proposed graph stochastic diffusion equation which includes a Q-Wiener process to model the stochastic evolution of node representations. GNSD introduces a drift network to guarantee accurate prediction and a stochastic forcing network to model the propagation of epistemic uncertainty among nodes. Extensive experiments are conducted on multiple detection tasks, demonstrating that GNSD yields the superior performance over existing strong approaches.

论文链接: <https://openreview.net/pdf?id=xJUhgvm2u8>

成果十二.

周川副研究员与合作者的论文 Neural Jump-Diffusion Temporal Point Processes 被 International Conference on Machine Learning, ICML 2024 接收发表。

摘要: We present a novel perspective on temporal point processes (TPPs) by reformulating their intensity processes as solutions to stochastic differential equations (SDEs). In particular, we first prove the equivalent SDE formulations of several classical TPPs, including Poisson processes, Hawkes processes, and self-correcting processes. Based on these proofs, we introduce a unified TPP framework called Neural Jump-Diffusion Temporal Point Process (NJDTTP), whose intensity process is governed by a neural jump-diffusion SDE (NJDSDE) where the drift, diffusion, and jump coefficient functions are parameterized by neural networks. Compared to previous works, NJDTTP exhibits model flexibility in capturing intensity dynamics without relying on any specific functional form, and provides theoretical guarantees regarding the existence and uniqueness of the solution to the proposed NJDSDE. Experiments on both synthetic and real-world datasets demonstrate that NJDTTP is capable of capturing

the dynamics of intensity processes in different scenarios and significantly outperforms the state-of-the-art TPP models in prediction tasks.

论文链接: <https://openreview.net/pdf?id=d1P6GtRzuV>

成果十三.

马志明研究员与合作者的论文 Rethinking Specificity in SBDD: Leveraging Delta Score and Energy-Guided Diffusion 被 International Conference on Machine Learning, ICML 2024 接收发表。

摘要: In the field of Structure-based Drug Design (SBDD), deep learning-based generative models have achieved outstanding performance in terms of docking score. However, further study shows that the existing molecular generative methods and docking scores both have lacked consideration in terms of specificity, which means that generated molecules bind to almost every protein pocket with high affinity. To address this, we introduce the Delta Score, a new metric for evaluating the specificity of molecular binding. To further incorporate this insight for generation, we develop an innovative energy-guided approach using contrastive learning, with active compounds as decoys, to direct generative models toward creating molecules with high specificity. Our empirical results show that this method not only enhances the delta score but also maintains or improves traditional docking scores, successfully bridging the gap between SBDD and real-world needs.

论文链接: <https://openreview.net/pdf?id=8WSNl2XA9r>

成果十四.

马志明研究员与合作者的论文 UniCorn: A Unified Contrastive Learning Approach for Multi-view Molecular Representation Learning 被 International Conference on Machine Learning, ICML 2024 接收发表。

摘要: Recently, a noticeable trend has emerged in developing pre-trained foundation models in the domains of CV and NLP. However, for molecular pre-training, there lacks a universal model capable of effectively applying to various categories of molecular tasks, since existing prevalent pretraining methods exhibit effectiveness for specific types of downstream tasks. Furthermore, the lack of profound understanding of existing pre-training methods, including 2D graph masking, 2D-3D contrastive learning, and 3D denoising, hampers the advancement of molecular foundation models. In this work, we provide a unified comprehension of existing pre-training methods through the lens of contrastive learning. Thus their distinctions lie in clustering different views of molecules, which is shown beneficial to specific downstream tasks. To achieve a complete and general-purpose molecular representation, we propose a novel pretraining framework, named UniCorn, that inherits the merits of the three methods, depicting molecular views in three different levels. SOTA performance across quantum, physicochemical, and biological tasks, along with comprehensive ablation study, validate the universality and effectiveness of UniCorn.

论文链接: <https://openreview.net/pdf?id=2NfpFwJfKu>

