NeurIPS Tuesday

Yuchen Zeng

12/10/2019

1 Morning Poster Session

1.1 Multi-marginal Wasserstein GAN

Motivation: Existing methods neglect to jointly optimize the multi-marginal distance among domains, and thus cannot get guarantee the generalization performance and may lead to distribution mismatching issue. Moreover, existing methods ignore correlations among target domains, and are hard to fully capture information to improve the performance.

Summary: They propose a novel MWGAN to optimize the multi-marginal distance among different domains. They utilize the multi-marginal Wasserstein distance as the objective function to train the GAN. They define and analyze the generalization performance of MWGAN for the multiple domain translation set. They also perform extensive experiments to demostrate the effectiveness of MWGAN on balanced and imbalanced translation set.

Comment: This is a well-executed paper. The formulation is clear and the flow of ideas is natural. Theoretical analysis are provided. Extensive experiments with adequate comparison on both toy and real-world datasets demonstrate the effectiveness of the proposed method.

1.2 k-Means Clustering of Lines for Big Data

Summary: This work focus on performing k-means clustering on lines. They propose (two constant factor approximation) k-line mean algorithm to compute a set of k centers (points) in $\mathbb{R}^d$ that minimizes the sum of squared distances. The amount of data required by the merge-and-reduce corest tree is logarithmic in the number of lines. This is a straightforward generalization of the k-means problem where the input is a set of n points instead of lines. The results generalized for other distance function such as k-median, m-estimators, or ignoring farthest m lines from the given centers to handle outliers. Their contributions can be summarized as:

- Formulates a generalization of k-means clustering to lines. (strong)
- Provides both theoretical analysis and numerical algorithm. (strong)
- Performs simulations to compare the proposed method with alternative method. (less significant)
- Performs simulation to analyze the relationship between memory and coreset tree size over time, coreset construction time and machines in cloud. (less significant)

Comments: Their work has no constraints on the dimension. Therefore, it has wide application. I believe this can be a quite impactful paper.

1.3 Learning-Based Low-Rank Approximations

Summary: For a given $n \times d$ matrix A and a parameter k, the goal of this paper is to find a rank-k matrix B' such that $\|A - B'\|_F \leq (1+\epsilon)\|A - A_k\|_F$, where ε is a given constant, A_k is the best rank-k approximation to A. The existing Sketching-based approach with random sketch matrices requires fixed random sketch matrices for all inputs and do not exploit the structure of the input matrices. The main contribution of this paper is that they propose a learned sketching-based approach to find a learned sketch S that minimizes the total loss $\sum_i ||A_i - \text{SCW}(S, A_i)||_F$. The mixed sketch (SCW with learned + random) achieves better empirical performance and obatain worst-case bounds, compared with SCW with random sktech matrix R and SCW with learned sketch matrix S. They perform simulations to compare their methods with two alternative methods. Their method achieves lower test error and decrease quicker.

Comments: In this paper they introduce a learning-based approach to sketching algorithms for computing low-rank decompositions. They provide theoretical analysis. Experiments are performed to show the improvement their algorithm achieves in both test error and running time. In addition, this work can be easily extended to higher-order tensors.

Improvement: Express the novelty of the proposed method compared to other methods more clearly.

2 Afternoon Poster Session

2.1 Nonconvex Low-Rank Symmetric Tensor Completion from Noisy Data

Summary: Suppose T^* is a superposition of r rank-one symmetric tensors: $\boldsymbol{T}^\star = \sum_{i=1}^r \boldsymbol{u}_i^\star \otimes \boldsymbol{u}_i^\star \otimes \boldsymbol{u}_i^\star \in \mathbb{R}^{d\times d\times d}$. The goal of this paper is to recover tensor factors $\{\boldsymbol{u}_s^*\}_{s=1}^r$ and $\boldsymbol{T}^*$ from partial noisy samples $T_{i,j,k} = T_{i,j,k}^\star + E_{i,j,k}, \quad (i,j,k) \in \Omega \subset \{1,\ldots,d\}^3$. In other words, the goal of this paper is to recover the tensor with missing data. They develop the least squared estimator, obtain the theoretical accuracy guarantee. However, the challenge is that the mean squared error function is highly nonconvex. They propose a two-stage nonconvex algorithm to solve the problem. They utilize spectral initialization to get a good starting point. They can minimize the mean square error because the loss function is locally convex. Their algorithm achieves lowest computation complexity, compared with existed methods.

Comments: Combining local concavity and good starting point to solve non-convex problem is a clever idea. Theoretical analysis and numerical algorithm for the proposed estimator are provided.

2.2 Iterative Least Trimmed Squares for Mixed Linear Regression

Summary: Given n samples $\{(x_i,y_i)\}_{i=1}^n$, $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^d$, least trimmed squares estimator

$$\widehat{\theta}_{\mathrm{LTS}} = \arg\min_{\theta} \min_{S:|S|=[\pi n]} \sum_{i\in S} \left(y_i - \langle x_i, \theta\rangle\right)^2$$

is a robust estimator for linear regression, compared with least square estimator

$$\widehat{\theta}_{\mathrm{LS}} = \arg\min_{\theta} \sum_{i\in[n]} \left(y_i - \langle x_i, \theta\rangle\right)^2 .$$

This paper develop an iterative least trimmed squares estimator

$$S_t = \arg \min_{S:|S|=\lfloor \tau n\rfloor} \sum_{i\in S} \left(y_i - \langle x_i, \theta_t\rangle\right)^2, \quad \theta_{t+1} = \arg\min_{\theta} \sum_{i\in S_t} \left(y_i - \langle x_i, \theta\rangle\right)^2 .$$

The iterative least trimmed squared estimator involves alternating between (a) selecting the subset of samples with lowest current loss, and (b) re-fitting the linear model only on that subset. In this paper they provide rigorous theoretical evidence that the proposed estimator obtains state-of-art results for a specific simple setting: mixed linear regression with corruptions. The corresponding algorithms are provided.

Comments: The flow of ideas is natural. Theoretical analysis and numerical algorithm for the proposed estimator are provided.

Improvement: Perform the simulations and real data analysis to demonstrate the effectiveness of the proposed estimator empirically.

2.3 Large-scale Optimal Transport Map Estimation using Projection Pursuit

Summary: This paper studies the estimation of large-scale optimal transport map (OTM), which is a well-known challenging problem owing to the curse of dimensionality. Existing literature approximate the large-scale suffer from slow or none convergence in practice due to the nature of randomly selected projection directions. They propose an algorithm to utilize dimension reduction methods such as Principle Component Analysis (PCA) first, then calculate the Wasserstein distance for selected dimensions, respectively. Theoretical analysis and numerical experiments are provided. Adequate comparison with other methods in CPU time and mean convergence time for estimating the Wasserstein distance.

Improvement: Express the novelty of the proposed method compared to other methods more clearly.

2.4 Face Reconstruction from Voice using Generative Adversarial Networks

Summary: The contribution of this paper can be summarized as:

- Introduce a new task of generating faces from voice in voice profiling. (strong)
- Propose a simple but effective framework based on generative adversarial networks. (strong)
- Propose to quantitatively evaluate the generated faces by using a cross-modal matching task. (less significant)

Comment: It is a novel and interesting topic.

2.5 Beyond Alternating Updates for Matrix Factorization with Inertial Bregman Proximal Gradient Algorithms

Summary: Given $\boldsymbol{A} \in \mathbb{R}^{M \times N}$, find $\boldsymbol{U} \in \mathbb{R}^{M \times K}$, $\boldsymbol{Z} \in \mathbb{R}^{K \times N}$ such that $\boldsymbol{A} \approx \boldsymbol{U}\boldsymbol{Z}$ by solving

$$\min_{\boldsymbol{U} \in \mathcal{U}, \boldsymbol{Z} \in \mathcal{Z}} \left\{ \Psi \equiv \frac{1}{2} \|\boldsymbol{A} - \boldsymbol{U}\boldsymbol{Z}\|_F^2 + \mathcal{R}_1(\boldsymbol{U}) + \mathcal{R}_2(\boldsymbol{Z}) \right\}.$$

The optimization problem is non-convex. In this paper, they exploit the Bregman distance by proposing a novel Bregman distance for matrix factorization problems. A non-alternating algorithms are proposed. The superior performance of their non-alternating schemes in terms of speed and objective value at the limit point are observed in experiments.

Improvement: Express the novelty of the proposed method compared to other methods more clearly.