ICML 2018
International Conference on Machine Learning 2018
July 10 to July 15 2018, Stockholm, Sweden

Reviews For Paper

Paper ID 1805

Title Clustering via Generalized Energy Statistics

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Summary of the paper (Summarize the main claims/contributions of the paper.)	The paper explores the optimization problem underlying energy based clustering formulation. An iterative algorithm based on Hartigan's method has been proposed. It finds a local maxima of the underlying optimization problem (which is a quadratically constrained quadratic program that is NP-hard to solve). Experiments done on one real dataset and synthetic datasets show that the proposed algorithm is better than k-means, kernel k-means and GMM. It also matches up to the performance of spectral clustering. The paper also emphasizes the model free nature of the proposed approach.
Clarity (Assess the clarity of the presentation and reproducibility of the results.)	Above Average
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Below Average
Correctness (Is the paper technically correct?)	Paper is technically correct
Overall Rating	Weak accept
Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	In an attempt to verify the numbers presented in the paper, an experiment which performs clustering on three concentric circles using spectral clustering algorithm was constructed. For the right choice of the affinity matrix (RBF kernel with appropriate sigma), a score > 0.95 was obtained which is far away from the reported score (used the sklearn's implementation of spectral clustering). The work done is very incremental and it lacks significant new contributions.
Reviewer confidence	Reviewer is knowledgeable

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Summary of the paper (Summarize the main claims/contributions of the paper.)	The authors have submitted a clearly written manuscript that shows a mathematical derivation of energy clustering and reveals relationships to existing kernel k-means and spectral clustering cost functions. Then they propose to solve the cost function via Hartigan's method rather than Lloyd's method, which has theoretical benefits as well as seemingly some empirical ones.

Clarity (Assess the clarity of the presentation and reproducibility of the results.)	Above Average
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Below Average
Correctness (Is the paper technically correct?)	Paper is technically correct
Overall Rating	Weak reject
Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	First, I think that the paper should be rewritten to clarify the contributions of the paper. It seems that the two main contributions of the paper are to (i) reveal relationships between energy clustering and existing methods and (ii) extend Hartigan's method to the kernel case. In the abstract the author's claim that they "propose a principled approach to clustering based on energy statistics theory;" however, they have been previous efforts on energy clustering (including papers cited in this manuscript), so this claim needs to be clarified. My reading of the paper is that only new relationships have been revealed, which is a less grandiose than the original claim. Second, there is significant effort devoted to revealing relationships to energy clustering, and it is implied that this is algorithmically important. However, in the end it seems that Hartigan's method could of been applied to any of the previously existing frameworks for similar clustering approaches. The primary advantage seems to be that one could generalize to any semimetric of a negative type. When would this have an advantage? The rationale behind the motivation for examining the energy statistics rather than existing cost functions needs to be clarified. Third, the extension of Hartigan's method is nice. I would suggest that the authors address the novelty more completely. As it is written, it is a straightforward extension of a previously proven technique already shown useful in very related clustering applications. What is the key novelty in this development? Finally, the application to real data seems to come without any conclusion. What have the new methods helped us learn? It's unclear if there's practically any difference here over existing clustering algorithms
Reviewer confidence	Reviewer is an expert

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Summary of the paper (Summarize the main claims/contributions of the paper.)	The paper introduces an approach to clustering based on the theory of energy statistics, which is called energy clustering under general semi-metric ρ. One of the contributions of the paper is to obtain the equivalence between energy clustering optimization and kernel K-means of the corresponding positive semi-definite Gram matrix.
Clarity (Assess the clarity of the	Excellent (Easy to follow)

presentation and reproducibility of the results.)	
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Above Average
Correctness (Is the paper technically correct?)	Paper is technically correct
Overall Rating	Weak reject
Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	Even though the paper is rather interesting, the contribution of the paper is quite marginal given the previous work using energy statistics theory for clustering purpose (e.g, Li (2015) for Euclidean case). In particular, it is an extension of Li (2015) work with Euclidean distance to the setting of general semi-metric ρ through a connection with RKHS. Here are some of my detailed comments regarding the paper: (1) The arguments that the authors use Hartingan's method over Lloyd's method for the inference is not convincing to me. In particular, the authors rely on the advantages of Hartigan's method over Lloyd's method (e.g, Theorem 4 and Theorem 5 in the paper with Euclidean distance) on two aspects: Hartigan's method may potentially escape local optima of Lloy's method as well as perform better than Lloyd's method in high dimension. I am not sure whether these advantages will hold for general positive definite kernel of kernel k-means. The authors may need to think about establishing the results of Theorem 4 and Theorem 5 under general kernel for more convincing argument when they use Hartingan's method instead of Lloyd's method. (2) For the experiment results in Figure 1 with model (24), I think it is unfair to compare the accuracy of GMM with $\mathcal{E}^{H}$-clustering as the true model (24) is clearly very different from GMM. Did you compare the accuracy result of clustering with parameters learned from EM of (24) and $\mathcal{E}^{H}$-clustering? (3) Minor typos: --- Line 238: Should remover i next to the equation with $\Delta Q^{j \to l}$. --- Line 253: Miss Q in the formulation of $\Delta^{j \to l}(x_{i})$.
Reviewer confidence	Reviewer is an expert