AISTATS 2018
Artificial Intelligence and Statistics 2018
Apr 9 , 2018 - Apr 11, 2018, Lanzarote, Spain

Reviews For Paper

Paper ID 286

Title Clustering via Generalized Energy Statistics

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Overall Rating	Weak Accept
Detailed Comments	The main point of this paper is to clarify theoretical background of a clustering method based on energy statistics. Furthermore, the authors showed equivalence of the clustering method to kernel k-means clustering. The theoretical analysis is interesting and may offer some insights to the machine learning community. By combining Hartigan's method, the new clustering method was also proposed. However, the novelty of the proposed clustering method is small because k-means clustering based on Hartigan's method has been already proposed. Overall, I am still leaning to accepting this paper because the experimental results were fairly convincing. Minor comments: (1) What is the definition of accuracy in the clustering experiments? Adjusted Rand index? (2) Regarding Fig.1(d), why do the accuracies of k-means and GMM decrease as the sample size increases? The reason is that the model assumption in k-means and GMM is NOT satisfied? This is a bit counter intuitive in some statistical sense.
Summary of the paper and sumary of your comments	This paper provides a theoretical foundation of a clustering method based on energy statistics and shows some equivalence to kernel k-means clustering. Theoretical analysis is interesting, although the algorithmic novelty of the proposed clustering method is small. In total, I am still positive to this paper because the experimental results are convincing.

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Overall Rating	Weak Accept
Detailed Comments	1. The paper is well written and easy to follow. 2. While equation 3, is empirical estimate of expectation of the first term of equation 1, when X is samples from C_i and Y is samples from C_j, it is not clear the justification for equation 4 and 5. In particular for within energy dispersion (equation 40, what is the justification for a multiplicative factor n_j/2? Similarly, what the justification of the multiplicative factor in equation 5? Some discussion and possible derivation for these would be beneficial. 3. More experiments on real world datasets will bolster the claim of this paper.
Summary of the paper and sumary of your comments	In this paper, the authors presents clustering technique based on general energy statistics. The proposed formulation show connection to kernel methods and an iterative algorithm based on Hartigan's method is presented to solve the clustering methods. Experimental evaluations show that the proposed method performs

	better than, k-means, kernel k-means and spectral clustering algorithm. 1. The paper is well written and easy to follow. 2. While equation 3, is empirical estimate of expectation of the first term of equation 1, when X is samples from C_i and Y is samples from C_j, it is not clear the justification for equation 4 and 5. In particular for within energy dispersion (eqn 40, what is the justification for a multiplicative factor n_j/2? Similarly, what the justification of the multiplicative factor in eqn 5? Some discussion and possible derivation for these would be beneficial. 3. More experiments on real world datasets will bolster the claim of this paper.

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Overall Rating	Weak Reject
Detailed Comments	This paper proposes to formulate the clustering problem as energy statistic maximization and obtain local minima using Hartigen's method. This paper is clearly written, well organized, and easy to follow. However, I have several concerns about both theoretical and algorithmic sides in the following. About theoretical side, the new formulation using the energy statistic is interesting. However, the significance of the formulation is not clear as it seems to be a reformulation of existing known relationship between kernel k-means, spectral clustering, and graph partitioning. It would be more interesting and significance if the proposed formulation reveals unknown theoretical connections. About algorithmic side, the idea of using Hartigen's method is also interesting, but its empirical contribution is not clear. Since the proposed method is theoretically equivalent to kernel k-means, what makes the difference of accuracy between the proposed method and kernel k-means in Figure 3 (b),(d)? It is currently not clear and this point should be carefully examined by experiments. In addition, there are two issues in experiments: - There is no comparison on real-world data, which may provide more insight about the effectiveness of the proposed method. - Accuracy is often not appropriate as a measure of clustering performance. I recommend to use variation of information instead. (see: M. Meilă, Comparing clusterings-an information based distance, Journal of Multivariate Analysis, 98(5), 873-895, 2007.)
Summary of the paper and sumary of your comments	This paper proposes a new formulation of clustering and theoretically shows that the proposed formulation is a generalization of kernel k-means and spectral clustering. Although the paper has a novel contribution and well written overall, the significance is not clear and empirical comparison is not thorough.