Review for

"Sparse logistic tensor decomposition for binary data"

This work proposes a tensor decomposition method for sparse, low-rank, and binary tensors. The authors apply the majorization-minimization (MM) approach to solve the corresponding regularized likelihood optimization problem.

Overall, I feel the contribution is incremental, since this paper generalizes previous works (Allen, 2012; Sun et al., 2017) to the binary case via standard MM approach. Theoretically, no algorithm accuracy compared with true parameters is provided. Numerically, the experiments are inadequate, including lack of comparison with competitive method for the same problem, unfair comparison, and other issues. Hence, I am on the fence with respect to this paper.

1. (Novelty) Novelty is the biggest concern of this work. Though the combination of sparse tensor decomposition with MM algorithm is new, the sparsity framework from Allen (2012); Sun et al. (2017) and the MM approach themselves are quite classical and the combination is natural. Theoretically, the main theorems are straightforward, and algorithm accuracy compared with true parameters are not provided. Numerically, due to the lack of comparison with competitive method Kolda and Hong (2020), the empirical performance of proposed method with state-of-art is unknown. It is worthwhile to list the unique contributions clearly and thus emphasize the novelty.

2. (Theory) Current theoretical results do not guarantee the accuracy of the algorithm outputs.

   Specifically:

   - Theorems 1, 2, and 4 show the algorithm convergence to the local minimum. However, the performance of local minimum compared with true parameters is not provided. The number of iterations or the speed for the algorithm to achieve local minimum is also unknown. Therefore, we actually have no accuracy guarantees on the algorithm outputs.

   - Theorems 2 and 3 focus only on the rank-1 optimization. The convergence property for the case of $R > 1$ and multiple L initialization is not easily to extended. More analyses on general cases will be helpful.

   - Some theorem statements are vague for me. In Theorems 2 and 3, true parameters (if I understand correctly, no explanations in the statement) u, v, w are involved in the solutions $\hat{u}, \hat{v}, \hat{w}$. Is the solution $(\hat{u}, \hat{v}, \hat{w})$ the joint optimizer over the triplet (u, v, w) in (9) or the marginal optimizer when the other two vectors are fixed? In Theorems 1 and 4, is there any rigorous definition of "majorize"? In what sense, the quadratic error "majorizes" the negative likelihood?

3. (Model) More discussions on the offset μ are necessary. The extra offset μ brings the concern to the model identifiability and the extra computational cost. It is beneficial to explain the theoretical or empirical necessity of including an offset in the model, provide a rigorous model identifiability theorem, and compare the computational cost with and without offset.

4. (Experiment) Several problems arise in the numerical experiments.

   - (Comparison with competitive method) The method in Kolda and Hong (2020) tackles the same sparse, low-rank, binary tensor decomposition problem. This competitive method should be involved in all numerical and theoretical comparison. It is also helpful if the comparison with Kolda and Hong (2020) is non-applicable.

   - (Unfair comparison) In the comparison of Table 1, TSP and TTP are given true sparsity parameters. This setting is unfair to the other method without prior knowledge on sparsity. The parameter tuning procedures or the error from parameter tuning should be involved to make a fair comparison.

   - (Computational time) To compare the computational efficiency, it is helpful to report the theoretical computational complexity of the proposed MM-based method and compare with others. Also, in Table 2, it is helpful to report the time for tuning hyper-parameters including sparsity, rank R, number of initialization L, and the time to finish the initialization. The time reports for parameter tuning are critical for real-life applications.

   - (Model misspecification) In practice, with perturbed observations, we do not know the sparsity information. It is helpful to test the robustness of the proposed method to the model misspecification. For example, what are the performances of TSP and TTP with dense low-rank tensors and how about the other methods?

   - (Guidance) Both ℓ_1 and ℓ_0 regulation are proposed to address sparsity, and multiple parameter tuning (BIC/AIC, cross validation, experienced deviance) methods are provided. Practical guidance to choose the sparsity regulation and the tuning method should be provided.

References

Allen, G. (2012). Sparse higher-order principal components analysis. In *Artificial Intelligence and Statistics*, pages 27–36. PMLR.

Kolda, T. G. and Hong, D. (2020). Stochastic gradients for large-scale tensor decomposition. *SIAM Journal on Mathematics of Data Science*, 2(4):1066–1095.

Sun, W. W., Lu, J., Liu, H., and Cheng, G. (2017). Provable sparse tensor decomposition. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):899–916.