

Melnykov, Igor; Melnykov, Volodymyr

A note on the formal implementation of the K -means algorithm with hard positive and negative constraints. (English) [Zbl 07300773](#)

J. Classif. 37, No. 3, 789-809 (2020)

Summary: The paper discusses a new approach for incorporating hard constraints into the K -means algorithm for semi-supervised clustering. An analytic modification of the objective function of K -means is proposed that has not been previously considered in the literature.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)

Keywords:

K -means; semi-supervised clustering; hard constraints

Software:

MixSim

Full Text: [DOI](#)

References:

- [1] Barbier, G.; Zafarani, R.; Gao, H.; Fung, G.; Liu, H., Maximizing benefits from crowdsourced data, Computational and Mathematical Organization Theory, 18, 257-279 (2012) · [doi:10.1007/s10588-012-9121-2](#)
- [2] Basu, S., Banerjee, A., Mooney, R. (2002). Semi-supervised clustering by seeding. In Proceedings of the 19th international conference on machine learning (pp. 19-26).
- [3] Basu, S., Banerjee, A., Mooney, R. (2004). Active semi-supervision for pairwise constrained clustering. In Proceedings of the SIAM international conference on data mining.
- [4] Basu, S.; Davidson, I.; Wagstaff, K., Constrained clustering: advances in algorithms, theory, and applications (2008), Boca Raton: CRC Press, Boca Raton · [Zbl 1142.68005](#) · [doi:10.1201/9781584889977](#)
- [5] Bilenko, M., \& Mooney, J.R. (2003). Adaptive duplicate detection using learnable string similarity measures. In International conference on knowledge discovery and data mining (pp. 39-48).
- [6] Celeux, G.; Govaert, G., Comparison of the mixture and the classification maximum likelihood in cluster analysis, Journal of Statistical Computation and Simulation, 47, 127-146 (1993) · [doi:10.1080/00949659308811525](#)
- [7] Covões, TF; Hruschka, ER; Ghosh, J., A study of k-means-based algorithms for constrained clustering, Intelligent Data Analysis, 17, 485-505 (2013) · [doi:10.3233/IDA-130590](#)
- [8] Davidson, I., \& Ravi, S. (2005). Clustering with constraints: feasibility issues and the k-means algorithm. In Proceedings of the 2005 SIAM international conference on data mining (pp. 138-149): SIAM.
- [9] Dempster, AP; Laird, NM; Rubin, DB, Maximum likelihood for incomplete data via the EM algorithm (with discussion), Journal of the Royal Statistical Society, Series B, 39, 1-38 (1977) · [Zbl 0364.62022](#)
- [10] DeSarbo, WS; Mahajan, V., Constrained classification: the use of a priori information in cluster analysis, Psychometrika, 49, 187-215 (1984) · [Zbl 0554.62050](#) · [doi:10.1007/BF02294172](#)
- [11] Dinler, D., \& Tural, M.K. (2016). A survey of constrained clustering. In Unsupervised learning algorithms (pp. 207-235): Springer. · [Zbl 1344.90030](#)
- [12] Fatehi, K.; Bozorgi, A.; Zahedi, MS; Asgarian, E., Improving semi-supervised constrained k-means clustering method using user feedback, Journal of Computing and Security, 1, 273-261 (2014)
- [13] Gu, L., \& Lu, X. (2012). Semi-supervised subtractive clustering by seeding. In 2012 9th international conference on fuzzy systems and knowledge discovery (pp. 738-741): IEEE.
- [14] Hennig, C.; Meila, M.; Murtagh, F.; Rocci, R., Handbook of cluster analysis (2015), Boca Raton: CRC Press, Boca Raton · [doi:10.1201/b19706](#)
- [15] Liu, H., \& Fu, Y. (2015). Clustering with partition level side information. In 2015 IEEE international conference on data mining (pp. 877-882): IEEE.
- [16] Maitra, R.; Melnykov, V., Simulating data to study performance of finite mixture modeling and clustering algorithms, Journal of Computational and Graphical Statistics, 19, 354-376 (2010) · [doi:10.1198/jcgs.2009.08054](#)

- [17] McLachlan, G.; Peel, D., *Finite mixture models* (2000), New York: Wiley, New York · [Zbl 0963.62061](#) · [doi:10.1002/0471721182](#)
- [18] Melnykov, V.; Chen, W-C; Maitra, R., *Mixsim: an R package for simulating data to study performance of clustering algorithms*, *Journal of Statistical Software*, 51, 1-25 (2012) · [doi:10.18637/jss.v051.i12](#)
- [19] Melnykov, V.; Melnykov, I.; Michael, S., *Semi-supervised model-based clustering with positive and negative constraints*, *Advances in data analysis and classification*, 10, 327-349 (2016) · [Zbl 1414.62255](#) · [doi:10.1007/s11634-015-0200-3](#)
- [20] Nimmo, DWR; Herrmann, SJ; Sublette, JE; Melnykov, IV; Helland, LK; Romine, JA; Carsella, JS; Herrmann-Hoesing, LM; Turner, JA; Vanden Heuvel, BD, *Occurrence of Chironomid species (Diptera: Chironomidae) in the high Se-78 concentrations and high pH of Fountain Creek Watershed, Colorado, USA*, *Western North American Naturalist*, 78, 39-64-26 (2018) · [doi:10.3398/064.078.0106](#)
- [21] Ruiz, C.; Spiliopoulou, M.; Menasalvas, E., *Density-based semi-supervised clustering*, *Data Mining and Knowledge Discovery*, 21, 345-370 (2010) · [doi:10.1007/s10618-009-0157-y](#)
- [22] Śmieja, M.; Wiercioch, M., *Constrained clustering with a complex cluster structure*, *Advances in Data Analysis and Classification*, 11, 493-518 (2017) · [Zbl 1414.62273](#) · [doi:10.1007/s11634-016-0254-x](#)
- [23] Steinley, D.; Brusco, MJ, *Evaluating mixture modeling for clustering: recommendations and cautions*, *Psychological Methods*, 16, 63 (2011) · [doi:10.1037/a0022673](#)
- [24] Wagstaff, K., Cardie, C., Rogers, S., Schroedl, S. (2001). *Constrained K-means clustering with background knowledge*. In *Proceedings of the eighteenth international conference on machine learning (ICML-2001)* (pp. 577-584).
- [25] Wang, X., Wang, C., Shen, J. (2011). *Semi-supervised K-means clustering by optimizing initial cluster centers*. In *International conference on web information systems and mining* (pp. 178-187): Springer.
- [26] Yu, Z.; Luo, P.; You, J.; Wong, H-S; Leung, H.; Wu, S.; Zhang, J.; Han, G., *Incremental semi-supervised clustering ensemble for high dimensional data clustering*, *IEEE Transactions on Knowledge and Data Engineering*, 28, 701-714 (2015) · [doi:10.1109/TKDE.2015.2499200](#)
- [27] Zhigang, C., Xuan, L., Fan, Y. (2013). *Constrained k-means with external information*. In *2013 8th International conference on computer science & education* (pp. 490-493): IEEE.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.