

Weighted k -Centers & Lloyd's Algorithm

Taylor Killian & Leonhard Spiegelberg

February 22, 2016

Abstract

When determining the optimal location of distribution centers (e.g. warehouses, regional retail, or neighborhood commercial property), consideration must be made to provide effective customer service. It is assumed that individuals prefer the distance from them to any distribution center be minimized. To select the optimal location of distribution centers that minimize the average person's distance from any center, Lloyd's algorithm was modified to solve a variant of weighted k -centers. This augmented clustering algorithm is used to simulate a variety of distribution paradigms (differentiated by the weighting function used) using the 2010 U.S. Census population data for Massachusetts which provides a dense decision surface to optimize over. Results reaffirm the intuition that dense areas of high population are optimal locations to place centers. The effects of various weighting functions and population filterings on center locations are discussed.

1 Introduction

Serving the maximum number of people at minimum expense is a canonical problem in the retail and commercial industries. A company must balance placing their distribution centers near as many people as possible with the cost associated with opening and maintaining a large number of centers. A general formulation of this problem known as the metric k -center problem asks for a placement of n distribution centers that minimizes the maximum distance from any point to a center [1]. In this paper, we examine the variant in which each point has a weight that affects its distance from the center point; for this report the weight is considered to be a function of the population at each point. We also seek to minimize the sum of the maximum distances rather than the maximum individual distance. Efficient algorithms to solve this problem exist in one dimension [7]. For the two-dimensional case, we adapt Lloyd's algorithm, an algorithm commonly used to find evenly spaced points and subsets within a given set of data, to work on the 2010 U.S. Census data from the state of Massachusetts by modifying the clustering step.

If no attempt is made to avoid local extrema, then this problem is easy to solve for one center. In section ??, we give a solution for one center using compass search. However, compass search is not viable for a large number of points

because the number of required directions to search grows quickly. In section ??, we give a solution method for an arbitrary number of centers using the variant on Lloyd’s algorithm. In section 4, we use our algorithm to optimize placement of distribution centers in the population data of Massachusetts, subject to different constraints. We considered the use of various weighting functions centers, as well as the use of population filtering.

2 Background Information

2.1 Background

Here we’ll want to add some information about Sparse regression, set functions and definitions of sub modular and super modular, Dictionary Selection, Greedy Methods, etc.

2.2 Related Work

A simple overview of prior work

3 Methods

Place holder section to begin to outline our work

Algorithm 1 Notional algorithm

```

Figure out optimal arrangement of rows and columns of input data
while there’s still time in the semester do
  if  $f(p_k + s_k \lambda_i) < f(p_k)$  for some  $\lambda_i$  then
     $p_{k+1} = p_k + s_k \lambda_i$ 
     $s_{k+1} = s_k$ 
  else
     $p_{k+1} = p_k$ 
     $s_{k+1} = \alpha s_k$ 
  end if
end while
return VICTORIOUS

```

4 Experiments

Hold for applying our methods

4.1 Results

The results of our algorithms against others

5 Conclusions & Future Work

5.1 Conclusions

We will have done it!

5.2 Future Work

Rule the world

References

- [1] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C., Silverman, R., and Wu, A. Y (2002). An Efficient k -Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), pp. 881-892.
- [2] Lloyd, S. P (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* 28(2), pp. 129-137.
- [3] U.S. Census Bureau. *TIGER/Line with Selected Demographic and Economic Data: Population & Housing Unit Counts*. Retrieved from <https://www.census.gov/geo/maps-data/data/tiger-data.html>
- [4] Kolda, T. G., Lewis, R. M., and Torczon, V (2003). Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods. *SIAM Review*, 45(3), pp. 385-482.
- [5] Davis, C. (1954). Theory of Positive Linear Dependence. *Journal of American Mathematics*, 76(4), pp. 733-746.
- [6] *Python Shapefile Library*. Github repository. Retrieved from <https://github.com/GeospatialPython/pyshp>
- [7] Chen, D. Z., Li, J., and Wang, H. Efficient Algorithms for the One-Dimensional k -Center Problem (2015). *Theoretical Computer Science* 592, pp. 135-142.
- [8] Rossiter, K. (2011, July 20). What are census blocks? [Web log post]. Retrieved from <http://blogs.census.gov/2011/07/20/what-are-census-blocks/>
- [9] Ackerman, M., Ben-David, S., Brânzei, S., and Loker, D (2012). Weighted Clustering. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- [10] Bourne, D., and Roper, S (2014). Centroidal Power Diagrams, Lloyd's Algorithm, and Applications to Optimal Location Problems. *SIAM Journal on Numerical Analysis*, 53(6).