PIKA – Program for Imagining a Klustering Algorithm

Team Rocket

CS6242 – Data Visual Analytics

27 February 2020

Team Rocket

York Delloyd
Project Manager
ydelloyd3@gatech.edu

Yongquan Tan
Back-End Developer
ytan49@gatech.edu

Hongji Wang
Front-End Developer
hwang766@gatech.edu

Philip Edwards
Data Scientist
pedwards34@gatech.edu

Bowen Yang
QA and Architect
byang63@gatech.edu

Georgia Tech
CREATING THE NEXT

Heilmeier Questions

- Objective
  - To allow for users to see similar living areas based off of different customizable factors
- Current Approaches
  - Ranking cities based off of a set criteria
- What's New?
  - A new way for users to find a new home
  - Move away from rankings to clustering to allow users to make more informed decisions
  - Allow for users to customize attributes they care about
- Affected Parties
  - Anyone who is moving or looking for a new place to live
- Application Impact
  - Users will be report similar satisfaction with their new living areas and lower moving stress
- Cost/Benefit
  - $20,880 development cost
  - $1,200/year in server maintenance
  - Invaluable peace of mine and satisfaction
- Time Estimation
  - 8 hours/week X 6 weeks
- Checkpoints
  - Midterm – local application with frameworks
  - Final – Accurate visualization and interactivity for the end user

Literature Survey

- Opportunity and Background
  - Rankings were done using different attributes which have contradictory results [4]
  - Urban areas can be classified [8]
  - Cities on the east and west coast tend to move together in housing vacancies [2]

- Implementation
  - Most previous implementations focused on ranking locations based off a variety of algorithms listed below[7].
  - Manual clustering based upon a limited number of attributes[5, 6, 15]. Clustering has a wide variety of uses as demonstrated in a 2017 study where it was used to define different urban morphological zones [16].
  - K-means on categorized variables by applying advanced different dissimilarity measures[11].
  - K-modes, which takes the mode of a cluster instead of mean [10, 13, 3].
  - Clustering algorithms such as spectral clustering which treats data points as nodes on a graph [9, 10]. This could prove exceptionally important as we understand the “shape” of the clusters.
  - An alternative approach outside of clustering is to derive an index from our attributes, much like how a Constant-Utility Cost of Living index is derived by L.R Klein in 1947 [14]

Plan of Activities

Design

- Start up Django Backend
- Start up React Front End
- Connect application
- Building the dataset and normalize

Develop

- Develop clustering algorithms
- Develop back-end routes and connection to clustering algorithm
- Develop front-end interface where users can interact with the data

Deploy

- Front-end unit testing
- Back-end unit testing
- Application deployment
- User acceptance testing

References

[1] How We Rank the Best Places to Live & Retire. (n.d.). Retrieved from https://realestate.usnews.com/places/methodology

[2] Goetzmann, W. N., & Wachter, S. M. (1995). Clustering Methods for Real Estate Portfolios. Real Estate Economics, 23(3), 271–310. doi: 10.1111/1540-6229.00666

[3] Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. DMKD, 3(8), 34-39.

[4] Florida, R., Mellander, C., & Rentfrow, P. J. (2013). The Happiness of Cities. Regional Studies, 47(4), 613–627. doi: 10.1080/00343404.2011.589830

[5] Clustering Minnesota Cities. (2005, September 22). Retrieved from https://www.lcc.leg.mn/lga/Background/clustermethodology.pdf

[6] Geltner, D., MacGregor, B. D., & Schwann, G. M. (2003). Appraisal smoothing and price discovery in real estate markets. Urban Studies, 40(5-6), 1047-1064.

[7] Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. Organizational behavior and human decision processes, 39(1), 84-97.

[8] Francisco J. Goerlich Gisbert, Isidro Cantarino Martí & Eric Gielen (2017) Clustering cities through urban metrics analysis, Journal of Urban Design, 22:5, 689-708, DOI:10.1080/13574809.2017.1305882

[9] Ng, A. Y., Jordan, M. I., & Weiss, Y. (n.d.). On Spectral Clustering: Analysis and an algorithm. Retrieved from http://papers.nips.cc/paper/2092-on-spectral-clustering-analysis-and-an-algorithm.pdf

[10] R. Xu and D. C. Wunsch, "Survey of Clustering Algorithms," IEEE Transactions on Neural Networks, Institute of Electrical and Electronics Engineers (IEEE), May 2005.

https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=1763&context=ele_comeng_facwork

[11] Chandrasekhar, T. , Thangavel, K. , Elayaraja, E. Effective clustering algorithms for gene expression data. Int J Comput Appl. 2011; 32(4): 25–9.

[12] Beata Calka, “Estimating Residential Property Values on the Basis of Clustering and Geostatistics ”, Received: 7 February 2019; Accepted: 21 March 2019; Published: 24 March 2019

[13] Aristidis Likas, et al. “Pattern Recognition.” The Global K-Means Clustering Algorithm, vol. 36, no. 2, 2003, pp. 451–61.

[14] L. R. Klein , and H. Rubin. A Constant-Utility Index of the Cost of Living. Oxford University Press, 1947–1948, p. 4.

[15] Findlay, Allan, et al. “Where to Live in Britain in 1988.” Copyright © 1988 Published by Elsevier Ltd.

[16] Goerlich Gisbert, F. j., Martí, I. C., & Gielen, E. (2017, April 10). Clustering cities through urban metrics analysis. Retrieved from https://doi.org/10.1080/13574809.2017.1305882