Exploring Clustering Algorithms on Spatial Data in Video Games

Elise Fitsner, Department of Computer Science

July 10, 2025

This research was supported by the Valerie Kuehne Undergraduate Research Awards, University of Victoria

Supervised by Kevin Stanley and Reia Drucker

In this project, we looked at how unsupervised machine learning (specifically, clustering using Gaussian mixture models) can be used to determine player archetypes in Counter-Strike: Global Offensive (CSGO). We found that players on different sides play in very different ways; within each side, however, players seemed to be clustered more on their individual playstyle and/or the style of the round they were playing, as opposed to the map they were playing on.

1: FEATURES

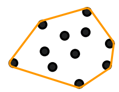
- Convex hull area
- Convex hull volume
- Fractal dimension
- Entropy constants
- Time spent alive
- Number of journey samples
- Number of dwell samples
- Alpha value for journey distribution
- Alpha value for dwell distribution

Using features computed from the ESTA-LAN dataset, we started with a dataset describing how individual players move while playing different games on different maps in CSGO. Each data point represents a player in a single round.

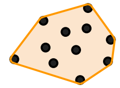
Remove players that were alive for less than 30 seconds

2: FILTERING

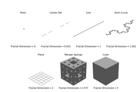
Convex hull area



Convex hull volume



Fractal dimension



Entropy constants

$$H(d,T) = \left(d^2\frac{C_1}{4T^2L} + \frac{C_2}{4T^2L} + 2d\frac{C_3}{4T^2L} + d\frac{C_4}{TL} + \frac{C_5}{TL}\right)^{-1}\log L$$

where

$$C_1 = \sum_{i=1}^{n}\frac{1}{{v_i^*}^2},\quad C_2 = \sum_{i=1}^{n} t_{di}^2,\quad C_3 = \sum_{i=1}^{n}\frac{t_{d_i}}{v_i^*},\quad C_4 = \sum_{i=1}^{n}\frac{1}{v_i^*},\quad C_5 = \sum_{i=1}^{n} t_{d_i}$$

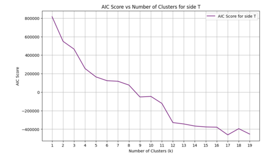
We determined that some features were too noisy to be used to cluster players, so we discarded them and chose to cluster on the features that were better for telling players apart.

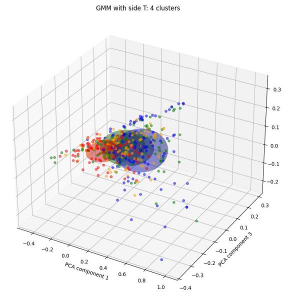
These features describe things like how much of the map a player explored, how quickly they moved, how often they dwelled, and how spatially complex their trajectory is.

Split into sides T and CT

3: CLUSTERING USING GAUSSIAN MIXTURE MODELS

T Side





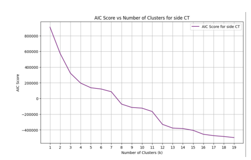
Cluster 0: Fast aggressive play
→ Dwelled the least, had the least spatially complex movement, moved the fastest

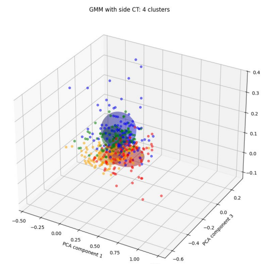
Cluster 1: Default playstyle/round
→ Explored a lot of the map, moving quickly, most spatially complex trajectory

Cluster 2: Lurkers
→ Dwelled the most, had a reasonable amount of exploration, moved the slowest

Cluster 3: Utility players
→ Dwell a lot, didn't explore as much, trajectory had some spatial complexity

CT Side





Cluster 0: Rotators
→ Explored the most, moved fairly fast, had the most spatially complex trajectory

Cluster 1: Site anchors
→ Dwelled the most, moved the slowest. Explored a lot of the map, but not in a spatially complex way

Cluster 2: Slow players/rounds
→ Explored the least, moved slowly and not in a spatially complex way

Cluster 3: Fast players/rounds
→ Moved the fastest, dwelled he least, did not explore a lot of the map

References

Rui Zhang et al. "Differentiating Population Spatial Behavior Using Representative Features of Geospatial Mobility (ReFGeM)". In: ACM Transactions on Spatial Algorithms and Systems 6.1 (Feb. 2020), pp. 1–25. doi: 10.1145/3362063. url: https://doi.org/10.1145/3362063.

Tuhin Paul, Kevin G. Stanley, and Nathaniel D. Osgood. "Multiscale entropy rate analysis of complex mobile agents". In: Royal Society Open Science 5.10 (Oct. 2018), p. 180488. issn: 2054- 5703. doi: 10.1098/rsos.180488. url: http://dx.doi.org/10.1098/rsos.180488.

Peter Xenopoulos and Claudio Silva. ESTA: An Esports Trajectory and Action Dataset. 2022. url: https://arxiv.org/abs/2209.09861.