Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27794

Exploring Behavioral Clustering for Player Matchmaking in Multiplayer Games

Haokai Tang

School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool, United Kingdom
sghtan11@liverpool.ac.uk

Abstract. Recent research highlights that beyond skill ratings, players' behavioral tendencies also play a crucial role in shaping gameplay experience. This study proposes a behavior-based clustering approach for matchmaking. It investigates how grouping players by behavioral similarity shapes interaction dynamics and system-level trade-offs. Using AI agents as proxies for players, four archetypal strategies—aggressive, defensive, cooperative, and non-cooperative—were modeled. Agents were clustered by behavioral features and matched either randomly (baseline) or by cluster similarity. The experiments, conducted over multiple runs in a simplified environment, compared outcomes at both the player and room levels using descriptive statistics and composite indices of diversity, balance, and richness. Results indicate that clustering changes interaction patterns rather than uniformly improving or degrading outcomes; notably, both defensive and aggressive agents showed reduced attack frequencies under similarity-based matching, suggesting environment-induced strategy adaptation. While the simulation is constrained and archetypes are simplified, the results suggest that behavior-aware matchmaking policies can optimize diversity, balance, and richness in practical deployment.

Keywords: Behavioral clustering, Player matchmaking, Multiplayer games, AI

1. Introduction

In multiplayer online games, matchmaking systems decisively shape player experience and fairness. However, objective measures alone are insufficient, as they are only reflect fairness when aligned with players' subjective experience [1]. Traditional methods, such as Elo or MMR, ensure some balance but have limitations in handling uncertainty and modeling team dynamics [2-3]. Moreover, they often overlook the complex behavioral patterns like cooperation, opportunism, or betrayal. Recent studies demonstrate that trust and betrayal dynamics in networked populations can strongly influence long-term interaction outcomes [4]. Ignoring these aspects risks mismatches that fail to capture actual player characteristics. Advances in artificial intelligence and behavioral analytics highlight the importance of modeling players not only through outcome-based indicators but also through observable behavioral data [5]. Features such as attack frequency, escape attempts, signaling for cooperation, and backstabbing reveal latent strategic tendencies and enable the identification of

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD27794

distinct player groups (e.g., aggressive, defensive, cooperative, non-cooperative), thus allowing for the construction of richer player profiles beyond traditional skill-based clustering.

This study explores behavioral clustering as an alternative to skill-based matchmaking using an agent - based simulation. Players perform actions like attack, escape, signal, and resource collection, representing four strategies. Behavioral features are extracted and clustered via unsupervised learning, then tested in matchmaking. The framework compares behavior-based grouping to random allocation regarding diversity, balance, and interaction richness, assessing if clustering enhances fairness and engagement and how matchmaking structures shape behavior.

2. Related work

Traditional matchmaking in competitive games uses skill-based rating systems like Elo, which represent player ability through relative skill levels in zero-sum settings. Yet Elo-based models face practical drawbacks: rapid rating adjustments can force matches between players of unequal ability, undermining fairness [6]. To address such issues, Di Fatta et al. proposed a Bayesian inference method that evaluates chess players not only by outcomes but also by the quality of moves, showing how probabilistic reasoning can capture ability through decision processes rather than final results [7]. Beyond outcome-based measures, clustering techniques have been employed to analyze player behavior. One study extracted behavioral features from gameplay and grouped players by play styles, further tracking how clusters evolved before and after game updates, thereby demonstrating how styles adapt to environmental changes [8]. Complementary work has sought to map detailed traits—such as caution, aggression, decisiveness, planning, and resource management—onto broader archetypes including Logistical, Tactical, and Strategic. These archetypes capture higher-level tendencies, with tactical players excelling at improvisation and strategic players favoring long-term reasoning, highlighting the diversity of decision-making across players [9].

Although these studies advance skill evaluation and behavior modeling, they neglect the feedback effects of matchmaking itself. This study investigates how grouping players by behavioral similarity reshapes interactions and dynamics in a controlled simulation.

3. Methodology

3.1. Research design

A tournament with 60 players and 300 matches (eight per room) is simulated to balance interaction diversity and computational tractability. Players are initialized as aggressive, defensive, cooperative, or non-cooperative, with action probabilities updated dynamically via Bayesian inference. On a grid-based map, they may attack, escape, signal, collect, or search, forming and breaking trust through reciprocal signaling and betrayal. Logs record kills, escapes, signals, resources, and backstabs. From these logs, behavioral features are extracted to represent outcomes and tendencies. Performance measures include win rate, kill/death ratio, escape rate, and resource efficiency; behavioral indicators cover attack, escape, and signal tendencies, aggression, and risk preference; social measures capture trust links, backstab rate, cooperation–betrayal balance, and social complexity; while complexity is reflected by strategy entropy, purity, and Bayesian confidence.

All features are normalized and compiled into a dataset for clustering using unsupervised algorithms (K-means, hierarchical clustering, Gaussian mixtures, spectral clustering). Validity is assessed by internal indices (silhouette, Davies–Bouldin, Calinski–Harabasz) and external indices (ARI, NMI) against initial types. PCA and t-SNE provide visualization. Two matchmaking modes

DOI: 10.54254/2755-2721/2025.LD27794

are then compared: random allocation and cluster-based grouping (cluster_loose), under identical settings. Simulation outputs include per-player and per-room statistics. Match quality is assessed across three dimensions, with an additional composite score. Diversity is measured by the normalized Shannon entropy of player types within each room:

$$D = -\frac{1}{\log K}\sum_{i=1}^{K} p_i \log p_i \tag{1}$$

Where K denotes the number of types and p_i the proportion of type i.

Competitive Balance is assessed by win rates across groups, as fairness of outcomes is a fundamental indicator in competitive settings:

$$B = 1 - \sigma(\text{winrate}) \tag{2}$$

Interaction Richness is defined by the variety of in-game dynamics:

$$R = \frac{0.2 \bullet \text{Kills} + 0.2 \bullet \text{Escapes} + 0.2 \bullet \text{Signals} + 0.2 \bullet \text{Backstabs} + 0.2 \bullet \text{TrustLinks}}{\text{Total_Actions}} \tag{3}$$

An Overall Quality Score is calculated as a weighted combination of these indices:

$$Q = 0.3 \bullet D + 0.3 \bullet B + 0.4 \bullet R \tag{4}$$

Weights are chosen to emphasize interaction richness (440%) as the central focus of this study, while ensuring that diversity and competitive balance are equally represented (30% each).

3.2. Experimental procedure

The experimental procedure begins by simulating tournament games with 60 players across 300 matches, generating detailed logs of player and room behaviors. First, these logs are processed to extract and normalize behavioral features, which are compiled into a dataset suitable for clustering. Second, unsupervised learning algorithms are applied to identify behavioral clusters, and their validity is assessed using both internal indices, such as silhouette score and Davies–Bouldin index, and external indices, including ARI and NMI. Then, two distinct matchmaking strategies—random allocation and cluster-based grouping—are implemented under identical tournament settings to ensure comparability. Finally, the outcomes produced by both modes are evaluated through diversity, competitive balance, and interaction richness, along with the composite overall score. The resulting datasets form the foundation for the comparative analysis presented in the Results section.

4. Results

4.1. Pre-clustering behavioral landscape

Before clustering, this study investigates player behaviors distribution. Figure 1 presents average action frequencies across player types, including kills, signals, backstabs, and wins.

DOI: 10.54254/2755-2721/2025.LD27794

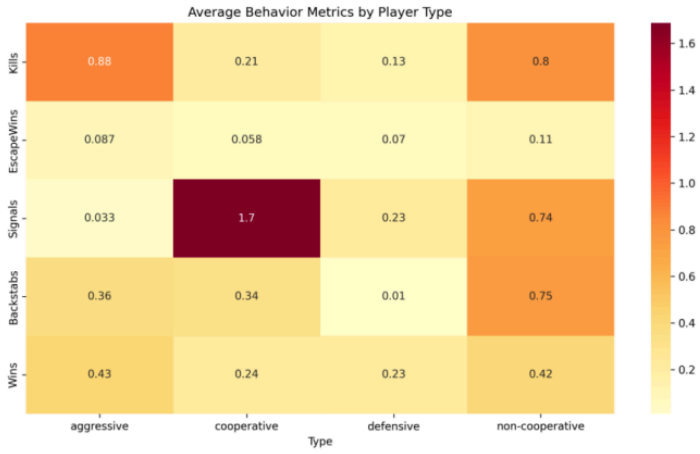

Figure 1. Heatmap of behavioral tendencies across player types (pre-clustering)

The heatmap provides a structural overview of tendencies, highlighting that cooperative players consistently dominate in signaling while aggressive players show higher kill frequencies. Complementing this, Figure 2 plots the distributions of strategy weights (AttackW, EscapeW, SignalW). The boxplots illustrate how different types occupy distinct regions in the strategic space: aggressive players lean toward offensive actions, defensive players concentrate on survival, and cooperative players engage in signaling more frequently. These pre-clustering observations serve as the behavioral baseline against which later clustering results are interpreted.

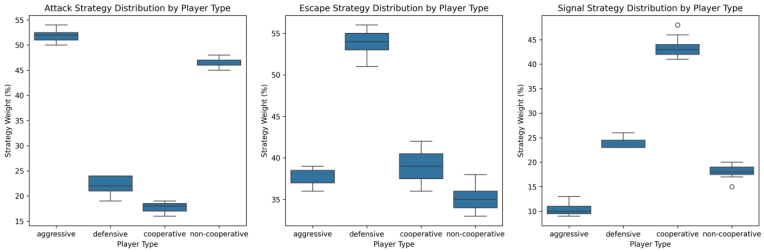

Figure 2. Strategy distribution across player types (pre-clustering)

4.2. Clustering outcomes

Behavioral clustering categorized the population into four distinct groups: Aggressive, Defensive, Cooperative, and Non-cooperative. Figure 3 presents a six-panel visualization of clustering quality and interpretability. The PCA projection demonstrates separability in the reduced space, while the aggression–defence scatter plot shows differences along key behavioral axes. The cluster distribution confirms that the four groups are well-balanced, preventing dominance by a single type.

DOI: 10.54254/2755-2721/2025.LD27794

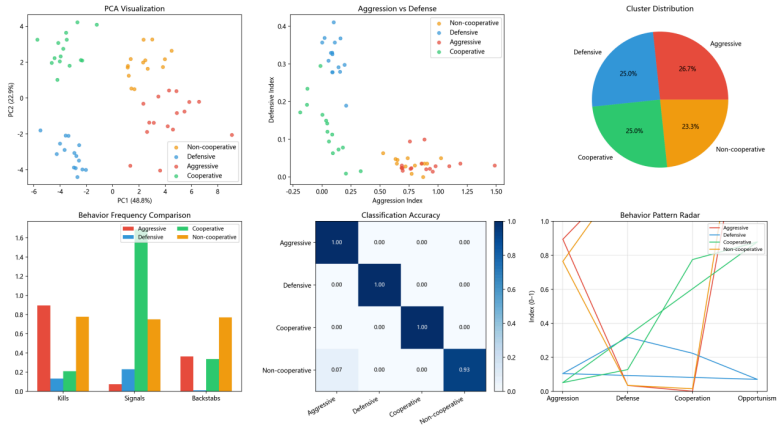

Figure 3. Six-panel visualization of clustering outcomes

Additional panels illustrate group characteristics: bar plots reveal the contrasting frequencies of kills, signals, and backstabs across clusters, radar charts highlight distinctive behavioral profiles, and the confusion matrix confirms a strong alignment between clustering outcomes and the original type labels. Together, these visualizations validate that clustering effectively captures heterogeneous player behaviors and provides a reliable basis for comparative matchmaking experiments.

4.3. Baseline vs. cluster: descriptive comparison

Table 1 reports per-player averages for wins, kills, deaths, signals, and backstabs across baseline and cluster-based matchmaking. The results show that clustering leads to a more uniform distribution of wins but also reduces the frequency of heterogeneous interactions. Specifically, signaling and backstabbing decrease substantially, while the number of kills also declines relative to the baseline condition. These reductions suggest that grouping similar players suppresses opportunities for cooperation, betrayal, and offensive strategies.

Table 1. Player-level comparison of baseline vs. cluste

metric	baseline_mean	cluster_mean	diff	diff_pct
Wins	13.083334	13.183334	0.099964	0.764331
Kills	21.116667	20.516667	-0.600000	-2.841358
Deaths	21.033335	20.433334	-0.600000	-2.852615
Signals	24.533335	26.450000	1.9166643	7.812500
Backstabs	12.666667	14.866667	2.200001	17.368421
Resources	98.316667	98.900000	0.583334	0.593321
EscapeWins	3.283334	3.400000	0.116667	3.553299

4.4. Match-quality metrics

The impact of clustering on overall match quality was assessed using three indices: diversity, competitive balance, and interaction richness. Table 2 summarizes the results along with the composite quality score. Diversity, measured by normalized Shannon entropy, decreased under clustering, reflecting the homogeneity of room composition. Interaction richness also declined, as the reduction in signaling and backstabbing limited the range of emergent events that could occur.

DOI: 10.54254/2755-2721/2025.LD27794

Table 2. Advanced metrics comparing baseline and cluster

level	metric	baseline_mean	cluster_mean	diff	diff_pct
per-player	Entropy	0.762746	0.766690	0.003945	0.517215
per-player	Variance	194.357334	216.137337	21.780000	11.206163
per-player	Richness	0.939998	0.916667	-0.023334	-2.482270
per-room	Entropy	0.812632	0.817672	0.005040	0.620209
per-room	Variance	4.507708	5.058333	0.550625	12.215187
per-room	Richness	0.891667	0.901667	0.009998	1.121495

Competitive balance displayed a mixed pattern: in some cases, clustering narrowed extreme disparities, but in others, it increased variance, producing instability rather than fairness. As a result, the composite score—which emphasizes richness as the most critical dimension—was consistently lower under clustering-based matchmaking. These findings confirm that clustering, while designed to improve balance, ultimately undermines emergent complexity and reduces overall quality.

4.5. Type-dominant rooms: an unexpected finding

When examining homogeneous groups of solely defensive or aggressive players, an unexpected pattern emerged. Initially, it was hypothesized that clustering would amplify type - specific behaviors: defensive groups would see more counterattacks due to mutual caution, and aggressive groups would witness heightened offensive activity as players reinforced each other's tendencies. However, the results defied these expectations.

As shown in Table 3, the average number of attack actions decreased in both defensive and aggressive groups under cluster-based matchmaking. Defensive players exhibited a sharp decline of 19.7% in attack frequency compared with the baseline, while aggressive players showed a more minor but still negative reduction of 2.4%. This trend was consistently observed across multiple experimental runs, confirming its stability. Thus, clustering actually suppressed offensive actions in both types of groups, rather than amplifying type - specific behaviors.

Table 3. Comparison of attack actions for defensive and aggressive players under baseline vs. cluster matchmaking

Player Type	Baseline Mean	Cluster Mean	Diff	Diff %
Defensive	95.4	76.6	-18.8	-19.7%
Aggressive	389.0	379.6	-9.4	-2.4%

This outcome highlights a critical feedback effect: when a room is composed entirely of defensive players, caution turns into collective passivity rather than increased counterplay; when composed entirely of aggressive players, mutual threat leads to hesitation instead of escalation. These counterintuitive shifts demonstrate that clustering does not merely reflect existing behavioral tendencies but actively reshapes them, producing dynamics that diverge from prior expectations.

In summary, clustering-based matchmaking introduces clear trade-offs. Pre-clustering visualizations confirmed structural differences among player types, and clustering outcomes revealed that heterogeneous behaviors could be grouped into four distinct profiles. Comparative experiments revealed that clustering balances wins but reduces behavioral diversity and richness, resulting in lower overall quality scores.

DOI: 10.54254/2755-2721/2025.LD27794

Type-dominant rooms further revealed robust feedback effects: defensive players showed a marked reduction in attacks, while aggressive players also decreased their offensive actions, though to a smaller extent. This consistent pattern across multiple runs demonstrates that even the most offensive group could not sustain its typical behavior in homogeneous environments.

5. Discussion

Players were divided into four discrete categories, and cluster-based matchmaking grouped same-type players within the same room. This design inevitably reduced diversity relative to random allocation. Real game environments, however, involve richer and more continuous behavioral spectra. A practical system would likely combine multiple behavioral dimensions and permit heterogeneous compositions, rather than enforcing strict homogeneity, to better reflect nuanced player profiles. Comparable large-scale clustering studies on commercial game telemetry have similarly demonstrated distinct player types with design implications [10].

The experiments show that clustering materially affects behavior. The initial expectation—that defensive rooms would increase counterattacks while aggressive rooms would amplify offensive actions—was contradicted: both groups exhibited reduced attack frequency under similarity-based matching. Defensive players showed a pronounced decline, and even aggressive players displayed a smaller but consistent reduction. The pattern reproduced across multiple runs, indicating robustness. These observations suggest that homogeneity constrains the available strategy space and suppresses rather than reinforces archetypal behaviors.

While behavior-based clustering can be used to shape gameplay dynamics, it does not by itself guarantee fairness. Depending solely on behavioral similarity may disadvantage certain groups. For practical deployment, behavior-aware policies are better combined with traditional skill-based mechanisms (e.g., Elo/MMR) to preserve competitive fairness while maintaining diversity and interaction richness.

6. Conclusion

This study evaluates the impact of behavior-based clustering on matchmaking and emergent dynamics by modeling four archetypal strategies with AI agents and comparing random allocation against similarity-based grouping across multiple runs. Measured at both player and room levels using diversity, competitive balance, and interaction richness (plus a composite score), clustering produced clearer separations among types and more uniform win distributions but simultaneously suppressed behavioral diversity and reduced interaction richness, yielding lower overall match quality. A consistent empirical pattern was that homogeneous rooms weakened, rather than amplified, archetypal behaviors: defensive agents markedly reduced attacks, and aggressive agents also showed a measurable decline.

However, this study is constrained by a simplified environment, restricted map design, limited player population and fixed match scale, and by the use of AI agents instead of human participants; artificial agents cannot fully capture the complexity, adaptability, and psychological factors of human play. Future research should test larger scales and more complex environments, include human or hybrid human–AI participants to strengthen external validity, and incorporate more nuanced features with continuous player representations. Dynamic clustering that updates profiles over time may better capture long-term strategy evolution and offer a more adaptive basis for matchmaking.

DOI: 10.54254/2755-2721/2025.LD27794

References

[1] Claypool, M., Decelle, J., Hall, G., & O'Donnell, L. (2015, October). Surrender at 20? Matchmaking in League of Legends. Paper presented at the IEEE Games Entertainment Media Conference (GEM), Toronto, ON, Canada.
[2] Delalleau, O., Contal, E., Thibodeau - Laufer, E., Chandias Ferrari, R., Bengio, Y., & Zhang, F. (2012, February). Beyond Skill Rating: Advanced Matchmaking in Ghost Recon Online. IEEE Transactions on Computational Intelligence and AI in Games, 167 - 177.
[3] Elbert, N., Flath, C. M., Klockmann, V., Kosse, F., von Schenk, A., & Stein, N. (2025, August). Matchmaking Beyond Elo: Why Teams are More Than the Sum of Individual Skill. Paper presented at the IEEE Conference on Games (CoG), Lisbon, Portugal.
[4] Shamsudin, D., Chew, L. M., & Yeng, O. L. (2022, September). Clustering Algorithms Analysis Based on Arcade Game Player Behavior. Paper presented at the IEEE Fifth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Laguna Hills, CA, USA.
[5] Chica, M., Chiong, R., Kirley, M., & Ishibuchi, H. (2017, November). A Networked N - Player Trust Game and Its Evolutionary Dynamics. IEEE Transactions on Evolutionary Computation, 22(6), 866 - 878.
[6] Elsayed, M., Hamdy, A., & Mostafa, A. (2022, December). Tibia Player Profiling and Modeling using Machine Learning. Paper presented at the International Conference on Computer and Applications (ICCA), Cairo, Egypt.
[7] Reyes, M. J. R., Dicreto, E., Santos, E. G. D., Limbag, D. F. P., & Sampedro, G. A. (2025, January). EloMetrics: Advanced Outcome Prediction for Chess Matches with Elo Ratings and Logistic Regression. Paper presented at the International Conference on Electronics, Information, and Communication (ICEIC), Osaka, Japan.
[8] Suznjevic, M., Matijasevic, M., & Konfic, J. (2015, December). Application context based algorithm for player skill evaluation in MOBA games. Paper presented at the International Workshop on Network and Systems Support for Games (NetGames), Zagreb, Croatia.
[9] Kwon, H., Jeong, W., Kim, D. - W., & Yang, S. - I. (2018, October). Clustering Player Behavioral Data and Improving Performance of Churn Prediction from Mobile Game. Paper presented at the International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South).
[10] Drachen, A., Sifa, R., Bauckhage, C., & Thurau, C. (2012, September). Guns, swords and data: Clustering of player behavior in computer games in the wild. Paper presented at the IEEE Conference on Computational Intelligence and Games (CIG), Granada, Spain.