Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27477

XGBoost-Based Synergistic Partner Recommendation in Strategy Games

Jinzhao He

University Hill Secondary School, Vancouver, Canada
jordihe0724@gmail.com

Abstract. Making a team with optimal synergy is essential for competitive success in strategy games like Pokémon. This study attempts an XGBoost-based recommendation model that forecasts synergistic Pokémon pairings utilizing numerical base stats and symbolic domain-specific features, incorporating type synergy scores and one-hot encodings of primary types. The paper applies 1,000 graded Gen9OU matches from Pokémon Showdown to produce a dataset as a binary classification task. The 49-dimensional advanced model, which has an Area Under the Curve of 0.824 and a precision of 0.766, shows the significance of type-based features in emulating team synergy. It provides notable performance improvements from the 12-dimensional baseline model. The planned strategy is compatible with the fundamental principles of Pokémon battles, which significantly affect outcomes. In order to enhance the accuracy and practical application, future work will likely investigate richer battle features, multi-core combinations, and larger datasets. Overall, this research underscores the potential of machine learning to systematically capture and predict strategic synergy in competitive gaming.

Keywords: Pokémon team building, XGBoost, Binary classification, Strategy recommendation

1. Introduction

For strategy games like Pokémon, the battle's winner may rely on the team's composition. A strong team needs offense, defense, and the capability to counter opponents. Traditional methods to build a team rely on player experience, expert advice, or trial-and-error processes. There is an expanding need for systems that may expertly suggest data-based teams as games grow less complex. Problems like recommendation lineups may be successfully rectified using supervised learning. XGBoost stands out from the existing models due to its stability, accuracy, and interpretability. It focuses well on numerical features like Pokémon attributes such as HP, attack power, defense power, and speed. This makes the decision-making process transparent. When accuracy and transparency are needed, XGBoost comes particularly adapted for this rendering. Its effectiveness may be proven by previous research. By both power and accuracy, Bentéjac and colleagues showed that XGBoost outclassed other gradient transforming models [1]. Zhang and Chen identified the necessity for explainability in building user trust over recommendation systems [2]. Wu et al. employed graph neural networks to explain the relation under the involved user-item interaction, but they said the cost was too high [3].

DOI: 10.54254/2755-2721/2025.LD27477

In the ability to predict user action, Purnamasari et al. utilized an integration of a used XGBoost, and the clustering has reflected the adaptability of the latter [4]. Conversely, there aren't many games that utilize directed learning to produce synergy-based recommendations. This fills the gap by producing an XGBoost-based system based on prior matches' records and estimates strong teammate compositions for Pokémon. By including 1 or 2 Pokémon, the system discovers the synergy patterns from the lineups that retain the high-winning rate and makes better possible recommendations. While most systems simply allow single-input support, this method allows single-core-based input and a candidate recommendation, which better represents actual team-building tactics [5-7]. The full model pipeline of the paper involves feature design, label building, training, and evaluation, and it demonstrates that XGBoost is efficient and scalable towards providing synergistic Pokémon lineup recommendations even on larger-scale teams such as triple-core configurations.

2. Methodology

2.1. Data collection

Two datasets are used for this project. The first is a battle log dataset, dragged from the Pokémon Showdown platform, which includes 1,000 ranked Gen9OU matches, incorporating full team compositions and outcomes. The second is a base stats dataset comprising over 300 Pokémon, each with reported numeric attributes and types. Team and name formatting is legitimized, and the data is preprocessed to plan for feature construction and modeling.

2.2. Feature engineering

The paper builds feature for the core Pokémon and the potential teammate (candidate) concerned to groom the synergy recommendation model. Each Pokémon is characterised by six numeric base stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. Those figures were given directly from the official base stats dataset and were normalized before processing into the models.

The candidate Pokémon (to become identified as a potential partner) and the core Pokémon (user-input or also on a team) make up each sample. Their base stats are chained to get a 12-dimensional numeric vector. Based on outdated type interaction charts, a type of synergy score is thus included in contemplating attribute compatibility. Furthermore, the paper incorporated 36 additional binary features by one-hot coding each Pokémon's primary type. Therefore, the final input vector totals 49 dimensions.

These pairs are selected from the losing team compositions from actual battle records to produce the positive samples. Negative samples are determined from losing teams or randomly sampling pairs that do not exist in the high-win-rate teams. This transforms the problem into a binary classification problem whose label is 1 for synergistic pairs and 0 otherwise.

Without taking the synergy score as a baseline comparison, the paper additionally utilized a standard pair classification model (consider 2.3). The paper constructed an alternative configuration that uses the same 12-dimensional base statistics (without the synergy score).

2.3. Model building

The paper modelled the Pokémon synergy recommendation problem as a binary classification task: given a core Pokémon and a candidate partner, detect whether the pairing makes a valid synergistic relationship. To study the consequences of various input characteristics on model performance, the paper constructed two XGBoost-based models. The baseline model produces a 12-dimensional input

DOI: 10.54254/2755-2721/2025.LD27477

vector only using the fundamental stats of the core Pokémon and the candidate Pokémon, specifically, the combination of their six (HP, Attack, Defense, Special Attack, Special Defense, and Speed) concatenations. A type synergy score (1D) and a one-hot encoding of each Pokémon's primary type (36D) are incorporated into the improved model to further enhance this. Forty-nine dimensions exist in the output vector. Both models were recruited employing an XGBoost classifier. The key hyperparameters were managed employing grid search and 5-fold cross-validation, with early stopping created to avoid overfitting, together with a maximum depth of 5 and a learning rate of 0.01, 100 epochs, a subsampling ratio, and a feature sampled ratio of 0.08. Both models were recruited using the same dataset, composed of samples of combinations of core Pokémon and candidate Pokémon, with a label of 1 marking good synergy and 0 marking poor synergy.

3. Results

3.1. Dataset description

The final training dataset is constructed from two-core Pokémon pairs. Each sample comprises three concatenated Pokémon feature vectors with six base statistics (HP, Attack, Defense, Special Attack, Special Defense, Speed) and one-hot type encodings. Positive and negative samples are made by missing team compositions or humanly created random pair compositions located on top-performing matches, separately, in the dataset. The paper sustains label balance by maintaining an equal set of negative and positive samples. Each sample is labelled 1 (synergistic) or 0 (non-synergistic). The resulting dataset is randomised before training and is utilized on five-fold cross-validation for assessment.

3.2. Evaluation metrics

Model performance is judged using accuracy, precision, recall, and AUC. These metrics are standard for binary classification and provide a balanced view of prediction quality. To ensure robustness, all results are weighted over a five-fold cross-validation.

3.3. Result & analysis

As shown in Figure1, the advanced model attained an accuracy of 0.758, precision of 0.766, recall of 0.801, and an AUC of 0.824 on the test set. The baseline model, which solely uses a 12-dimensional vector of base stats, employs substantially fewer features. Twelve numeric attributes, 36 one-hot encoded primary types, and a type synergy score based on the official type effectiveness chart are all included in the advanced model. Thanks to this enhanced feature representation, the model can more easily understand the underlying and candidate Pokémon's statistical patterns and semantic relationships.

The advanced model outperforms the baseline model in all other evaluation metrics. Accuracy improves from 0.555 to 0.758, precision from 0.555 to 0.766, and AUC from 0.532 to 0.824, signaling a substantial gain in classification performance and discriminative power. Despite having relatively high recall values, both models. 811 for the baseline vs. the advanced approach breaks a better balance between false positives and false negatives than the advanced model (0.0801). These results highlight the efficacy of incorporating domain-specific knowledge into feature design and show that mixing numeric stats with type-based synergy produces more accurate and interpretable team recommendations in strategic gameplay scenarios.

DOI: 10.54254/2755-2721/2025.LD27477

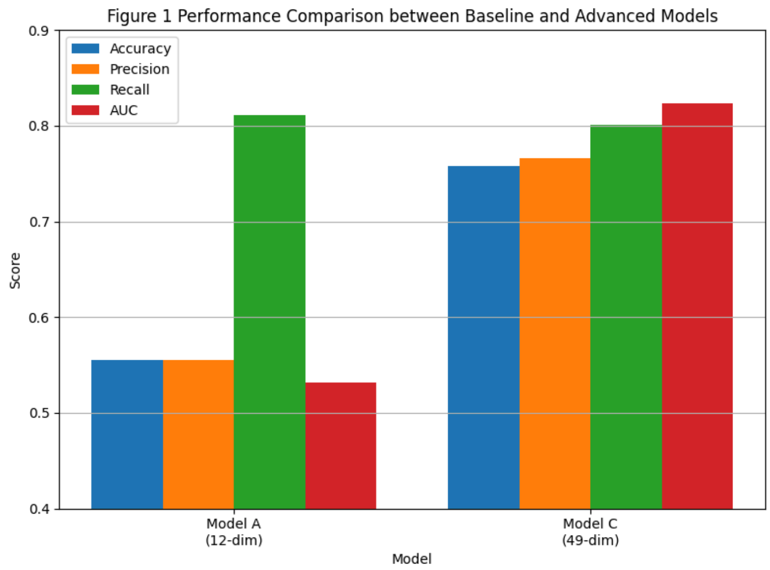

Figure 1. Performance comparison between Model A (12D) and Model C (49D) in terms of accuracy, precision, recall, and AUC

4. Discussion

This study indicates that incorporating teammate synergy recommendation models with domain-specific symbolic features, such as type synergy scores and one-hot encoded type representations, profoundly optimizes the performance of the Pokémon teammate synergy recommendation models. With an AUC of 0.824 and a precision of 0.766, the advanced model outpaces the baseline in key metrics, recognizing the need for type information when uncovering effective team combinations. These findings are in keeping with fundamental gameplay mechanics, which stipulate the outcome of battles. However, the current approach possesses a variety of limitations. Because the training data is limited to 1,000 matches from the Pokémon Showdown platform, this may not completely account for the range of competitor strategies.

Furthermore, synergy labels are posited directly from match outcomes, which may generate noise—for example, strong combinations may occur on losing teams due to misplays or unfavorable matchups. Also, the model maintains limited generalization ability when working with frequently used Pokémon or unconventional team structures. More in-depth battle features, such as move sets, natures, and item choices (e.g.), could be added to future works. For example, berries and status effects to better successfully capture synergy beyond static stats. Expanding the model to enable multi-core team structures (e.g.. Comprising turn-by-turn battle data, for instance, and employing tri-core combinations may improve its strategic understanding. Improving player-specific playstyle customization and extending the dataset with more diverse or simulated matches will help improve recommendation quality and gaming applicability. Moreover, the paper hope to leverage AI to enable rapid team lineups, and further extend its application to Nintendo Switch games [8-10].

5. Conclusion

This study develops an XGBoost-based synergy recommendation model for Pokémon team building. The model effectively captures core and candidate Pokémon compatibility by combining numerical base stats with domain-specific characteristics, such as one-hot encoded types and a type

DOI: 10.54254/2755-2721/2025.LD27477

synergy score. Results demonstrate that the advanced model dramatically outperforms the baseline in metrics like AUC and precision, authenticating the impact of combining type-related features. The strategy is essential to the game's strategic foundation, where type matchups are crucial in outcomes. This approach could also aid real-time teammate guidelines to help players establish more competitive teams in games. Future work will concentrate on extending the feature space and dataset size, incorporating multi-core team structures and dynamic battle elements such as movesets, held items, player strategies, and turn-by-turn interactions. These enhancements increase the model's personalization, interpretability, and adaptability in real gameplay scenarios.

References

[1] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54, 1937–1967. https: //doi.org/10.1007/s10462-020-09896-5
[2] Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval, 14(1), 1–101. https: //doi.org/10.1561/1500000056
[3] Wu, S., Sun, F., Zhang, W., & Zhu, J. (2020). Graph neural networks in recommender systems: A survey. arXiv preprint arXiv: 2011.02260. https: //arxiv.org/abs/2011.02260
[4] Purnamasari, S. D., Sutoyo, E., & Kurniawan, H. (2020). E-commerce product recommendations using XGBoost with user clusters and clickstream. International Journal of Advanced Trends in Computer Science and Engineering, 9(4), 5416–5421. https: //www.warse.org/IJATCSE/static/pdf/file/ijatcse258942020.pdf
[5] Cheng, W., Sun, Y., Zhang, Y., & Liu, J. (2023). Explainable recommendation with personalized review retrieval and aspect learning. Proceedings of the 61st Annual Meeting of the Association for Comrecommendation putational Linguistics (ACL). https: //aclanthology.org/2023.acl-long.4
[6] Ma, H., Zhang, H., Zhang, Y., & Liu, J. (2019). Jointly learning explainable rules for recommendation with knowledge graph. Proceedings of the 2019 World Wide Web Conference (WWW), 1210–1220. https: //doi.org/10.1145/3308558.3313590
[7] Hu, J. C., Hu, X. Y., Li, H. Z. (2020) Graph contrastive learning recommendation algorithm based on gradient aware graph enhancement [J/OL]. Computer Application Research, 1-10. https: //doi.org/10.19734/j.issn.1001-3695.2025.05.0172.
[8] Adomavicius, G., & Tuzhilin, A. (2022). A systematic review and research perspective on recommender systems. Journal of Big Data, 9, Article 92. https: //journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00592-5
[9] Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81, 84–90. https: //doi.org/10.1016/j.inffus.2021.11.011
[10] Chen, X., Zhang, Y., & Wen, J. R. (2022). Measuring “Why” in recommender systems: A comprehensive survey on the evaluation of explainable recommendation. arXiv preprint arXiv: 2202.06466. https: //arxiv.org/abs/2202.06466