Team: pokemon battle AI

Erdős Institute Data Science Boot Camp May-Summer 2024 — Executive Summary

Group members: Izabella Freitas, Guoqing Zhang, Tianyu Zhu, Hongyi Shen, Mary Ann Collins

Project Objective: We aim to create an AI model that can predict the likely winners of competitive Pokémon Video Game Championships (VGC) battles, utilizing replay records from *Pokémon Showdown!*[1], a popular online battle simulator. This format is a highly competitive, 'doubles battle' where each player selects four out of six Pokémon to battle against their opponent, with two Pokémon active for each player at a time. By reconstructing these replay records into text files, we can extract turn-by-turn information from over 10,000+ battles and curate a database to train our classification models.

Dataset

Data Gathering:

- Battles following the [Gen 9] VGC 2024 Reg G format were scraped from replays on *Pokémon Showdown!*
- Replays records were reconstructed into text-based .log files containing the information for each battle, sorted by turn; a total of 14,474 unique battles were collected
- Data not available from battle logs, such as the base stats and typing for each Pokémon, was sourced from Smogon[2] (the creator of *Pokémon Showdown!*) as .json files
- Data from battle logs and .json files were parsed into dataframes and then combined into one database

Data Processing:

- The database was filtered to remove battles that lasted less than five turns or ended in a tie
- Battles were sorted by rating so only battles between players above a certain decent ranking are considered for training
- All battle parameters were transformed into a 492-dimensional vector that was used for training the models described below. This vectorized dataset contained the conditions for each battle (weather, field, terrain) as well as the attributes of both player's pokémon played during the match

Modeling Approach: Several different classification models were generated[3]. For each model, the main Key Performance Indicator is the accuracy of the predicted winner compared to the actual winner of a battle. Below is a description for the most interesting models (*see the Table for the accuracy rate for all models*).

TABLE: MODEL ACCURACY RATE	
Baseline Model	0.500
k-Nearest Neighbors	0.513
Weighted Neural Network	**0.550**
Gaussian NaiveBayes	0.580
Neural Network	**0.600**
Quadratic Discriminant Analysis	0.616
Logistic Regression	0.656
Support Vector Classification	0.671
Linear Discriminant Analysis	0.672
Decision Tree Classifier	0.675
XGBoost	0.715
Random Forest Classifier	**0.766**
Extra Trees Classifier	**0.780**

1. Baseline Model: The baseline model used is a coin-flip with 50% chance of predicting the correct winner.

2. RandomForest Model: The features used for this model only considered the attributes of the four Pokémon (two Pokémon per player) that started at turn one, from the vectorized dataset. The initial accuracy in predicting the correct winner was 61%. When we revised this model to consider the attributes of all Pokémon from every turn of the battle as well as the battle conditions from the whole vectorized data, the accuracy increased to 76%.

3. ExtraTree Model: We also generated an ExtraTree Classifier, using the initial features from the RandomForest Model (only considering the attributes of the starting Pokémon). The initial accuracy in predicting the correct winner was the same as the first RandomForest Model (61%). When we used the revised feature set that considers the attributes of all Pokémon and the battle conditions throughout the match, the accuracy increased to 78% and performed slightly better than the RandomForest Model.

4. Neural Net (NN) Models: The first iteration of the NN achieved a 61% success rate in predictions. However, it became apparent that this model was heavily influenced by the total amount of HP of all remaining Pokémon on each side. To reduce the influence of the remaining total HP, the model was adjusted. However the accuracy rate for this Weighted NN Model modestly dropped to 55%.

Conclusions: Compared to our baseline model (50% accuracy), our other classification models performed better, with an accuracy of 55% - 78% to predict the correct winner. The best overall model was the ExtraTree Classifier, which incorporated the whole vectorized dataset, including the battle conditions as well as attributes for each Pokémon as features, and was trained on every turn for each battle. While neither of our Neural Networks performed as well as the ExtraTree Classifier, we were able to use the NN for more in-depth team analysis and explore the impact the most popular Pokémon have on the metagame.

Modeling Insights:
In addition to predicting the winner of each battle, we also explored how individual Pokémon impact the battle outcomes by modifying the input vectors—essentially swapping one Pokémon for another and observing changes in the model's output. This analysis revealed that Chi-Yu and Gholdengo, two Pokémon known for their medium speed, moderate durability, and powerful special attacks, play similar roles when on a team. This analysis also revealed the unique advantages of certain Pokémon. For example, although Incinaroar and Calyrex-Ice are two of the most-played Pokémon, they serve completely different roles; Calyrex-Ice is a strong Physical Attacker while Incinaroar is the strongest supportive Pokémon.

Modeling Challenges:
We encountered a number of challenges; mainly concerning dataset variability and the computational costs of training various models. In addition, competitive Pokémon battling involves a significant level of uncertainty due to random chance events, information unavailable to both players, and a highly variable set of Pokémon. Such factors were difficult to fully account for in our models, as reflected in each of their accuracy rates, yet highlights the complexity of the game and dataset.

Future Directions & Applications: We believe this method of analyzing Pokémon battles holds promising potential for future applications in broader competitive strategy development and training. By understanding the roles and impact of specific Pokémon more deeply, players can refine team building and battle tactics more strategically. This model could also be used by *Pokémon Showdown!* and official Pokémon tournaments organizers (such as Nintendo and The Pokémon Company) to track changes in the metagame as new rulesets are released. Lastly, this approach could be adapted to other complex systems where numerous variables influence outcomes, such as other strategic turn-based game theory simulations.

---

[1] “Pokémon Showdown! Battle Simulator” https://replay.pokemonshowdown.com/
[2] “Smogon University pokemon-showdown” https://github.com/smogon/pokemon-showdown/tree/master/data
[3] “GitHub Repository for pokemon-battle-AI” https://github.com/maryanncollins/pokemon-battle-AI