EDH-LLM: Using NLP techniques to build card game decks

Theodore Hui
tchui@ucsd.edu

Linus Lin
l6lin@ucsd.edu

Evelyn Yee
eyee@ucsd.edu

Mentor: Jingbo Shang
jshang@ucsd.edu

UC San Diego™
HALICIOĞLU DATA SCIENCE INSTITUTE

Our website!

Introduction

Magic the Gathering is a deck-building card game where players collect cards and assemble custom decks to play against each other. Currently, there are **over 27,000** distinct Magic the Gathering cards to choose from.

In the popular **Elder Dragon Highlander (EDH)** format of play, players compete in a fast-paced, 1v1v1v1 competition to be the last man standing.

Valid decks for EDH must meet the following restrictions:

- Decks must contain a "**Commander**" card, which is the focal point of the deck and must be a Legendary Creature.
- There must be a total of 100 cards, which fit into the color typing of the commander.
- All cards in the deck must be unique, with the exception of basic land cards, which may be repeated.

Additionally, for optimal play, deck builders must consider:

- **Synergies:** Cards in a deck must interact well with each other and may create useful combos
- **Power Curve:** Decks must strike a balance between three main archetypes of cards: card draw, mana generation, and gameplan.
- **Rule Zero:** EDH is a casual format, so decks must be fun to play *with* and *against*, regarding their speed and power.

This combination of hard and soft objectives, and the large number of potential cards, deckbuilding a complex space to optimize over. We frame this problem as an **information retrieval** task, leveraging NLP tools to build well-rounded, competitive decks with **no human supervision**.

References

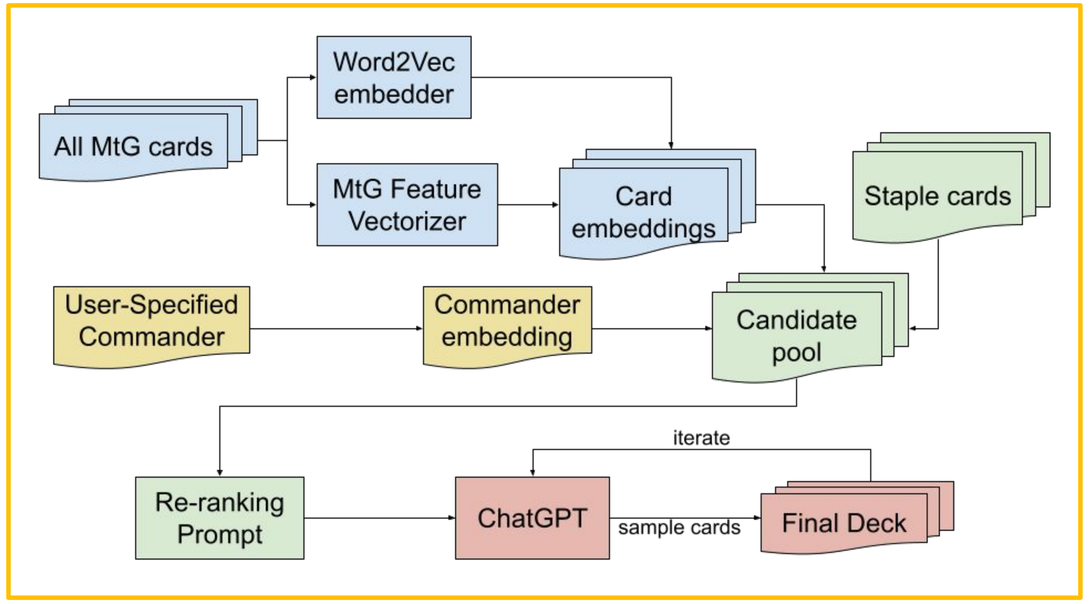
1. EDHREC. 2024. "EDHREC." URL: edhrec.com
2. Gavin. 2020. "My EDH Power level formula." Nov. URL: discipleofthevault.com/2020/11/18/my-edh-power-level-formula/
3. Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space.

Deckbuilding Pipeline

We frame the challenge of deck building as an **information retrieval task**, using the Commander card as the query/anchor to select the remaining composition of the deck.

To build a deck, given a specific commander, we:

1. Vectorize each card using Word2Vec[3] text embeddings and domain-specific features.
2. Filter the cards based on the commander's color typing.
3. Construct a **candidate pool** of 500 potential cards, using cosine similarity to the commander plus generic "staples," goodstuff cards that aren't necessarily synergistic but powerful in general.
4. Use ChatGPT to **re-rank** the candidate pool and iteratively sample cards for the final deck.



Evaluation

To empirically address the requirements of this building problem, we implemented two heuristic-based metrics for evaluation:

- **Synergy Heuristic:** We estimate the synergy between a pair of cards through a Bayesian probability measure, based on card co-occurrence in decklist data from EDHRec.[1] For each deck, we record the **average synergy** over all pairs of non-basic cards as well as the **Commander synergy** between each non-commander card and the commander.
- **Power Heuristic:** A custom heuristic for evaluating the power level of an EDH deck based on the distribution of functional card types during gameplay. A powerful deck should have a balance of a low average mana cost (**A**), card draw (**D**), deck searching/tutoring (**T**), mana ramp (**R**) and interaction/removal (**I**).[2]

$$\text{synergy}(a,b) = \log\left(\frac{freq(a,b)^2}{freq(a)freq(b)}\right)$$

Synergy Heuristic Formula

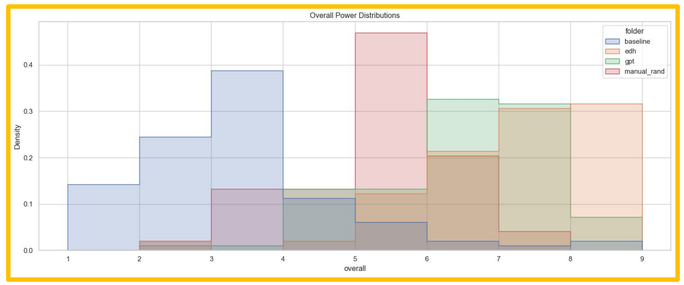
$$\frac{2}{A} + \frac{\frac{D}{2}+T+\frac{R}{2}}{2} + \frac{I}{20} = P$$

Power Heuristic Formula

Results

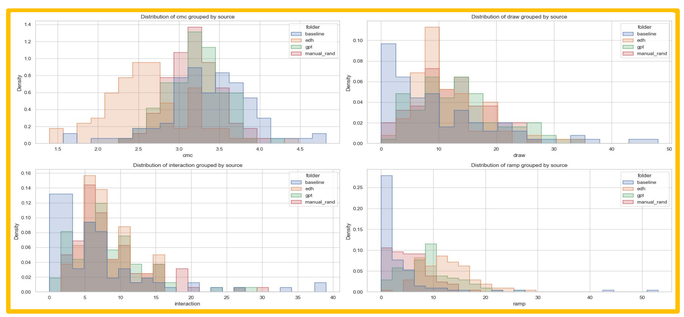
We evaluate our pipeline against three baselines for building EDH Decks:

- **EDHRec Baseline:** Greedily select the top cards which are most frequently played with the selected Commander, according to historical play data from EDHRec.[1]
- **Embedding-only Baseline:** Greedily select the top cards from each candidate pool, without any reranking by GPT-3.5.
- **Random Sample Baseline:** Randomly sample cards from each candidate pool.



Power Heuristic analysis:

- **Text similarity is not sufficient** indicator of deck strength
- **Manual feature engineering** improves card embedding utility
- **Historical play frequency creates strong decks**, even naively sampled.
- Final decks have **balanced performance** between EDHRec Baseline and Embedding-only Baseline



We inspect each individual component of the Power Heuristic formula:

- Most commonly played cards tend to have **low in-game cost** to play (CMC) and **high ramp**
- The Embedding-only baseline decks lack significant draw, interaction, and ramp

Conclusion

Historical play data seems to be the strongest predictor of deck quality in the EDH format of Magic the Gathering.

Text embedding techniques, combined with domain expertise and pre-trained LLM knowledge, can achieve similar results to historical results, with **much less labeled data**.

This information retrieval-based approach could be adapted to **other set-building tasks** which have a combination of hard restrictions and soft optimization objectives, like synergy.