Predicting Human Card Selection in *Magic: The Gathering* with **Contextual Preference Ranking**

Timo Bertram, Johannes Fürnkranz, Martin Müller

Accepted at IEEE Conference on Games 2021

Content

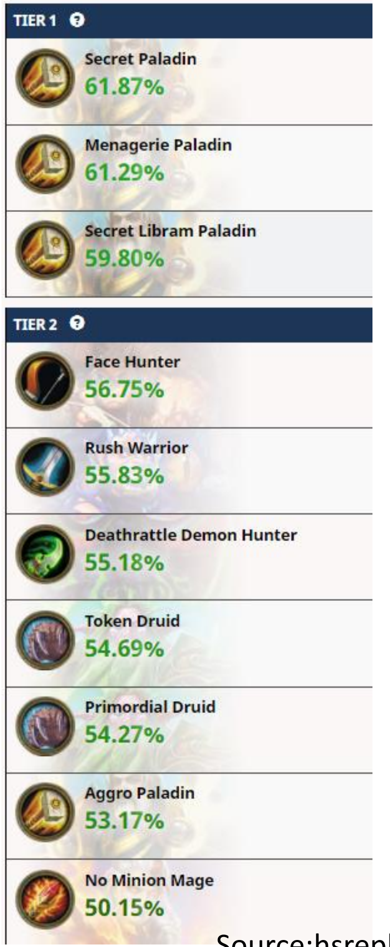
1) General problem
2) Application to card games
3) Magic: The Gathering
4) Dataset
5) Contextual Preference Ranking Method
6) Results
7) Limitations

General problem

- Select optimal subset of set
- Select optimal item to add to set
- Main problem: Evaluation of set ≠ sum of items
  - Don’t want 11 goalkeepers, band with only guitarists

Application to card games

- Collectible Card Game (MTG, Hearthstone,..)
- Select set of cards (deck) to play with
- Deck building is integral part of success



Source:hsreplay.net

Magic: The Gathering

- One of the most widely played collectible card games
- Strength of card is highly dependent on context
- Cards generally have one of five colors
- More different colors makes decks less consistent
- Cards have different rarities which decide their chance of occurrence

Source: Wizards of the Coast

Drafting

- Cards are structured in *packs* of 15 cards
- At each step, a set of cards N is offered and a single one is chosen to add to the deck of the player
- Unchosen cards are given to next player
- Number of cards to chose from varies from 15 to 1 throughout draft
- 45 cards are chosen per player in a single draft
- Goal of this research: Predict the card chosen by a human player

Data

- Collection of Magic: The Gathering drafts
- Data collected from website where humans can simulate drafting

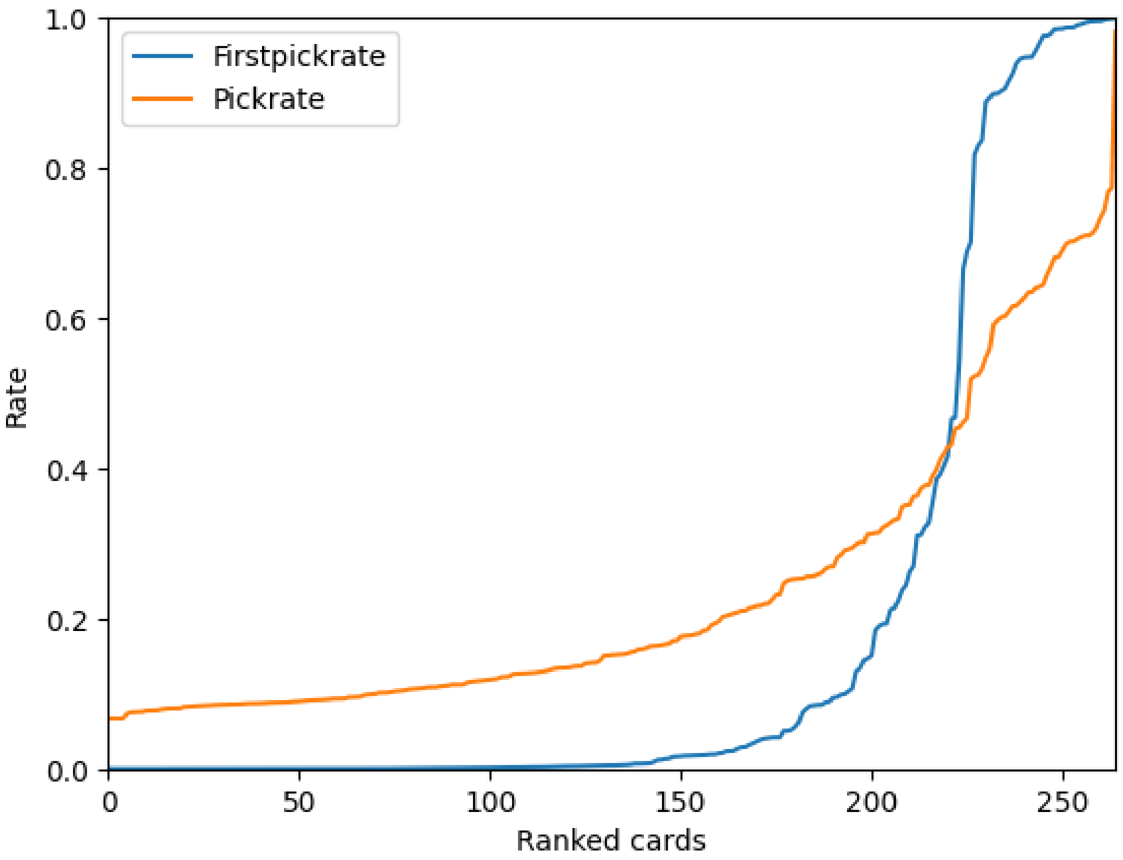
Source: draftsim.com

Data

- 107.949 drafts
- 8 players per draft
- 265 different cards in dataset
- Highly imbalanced quantity of cards

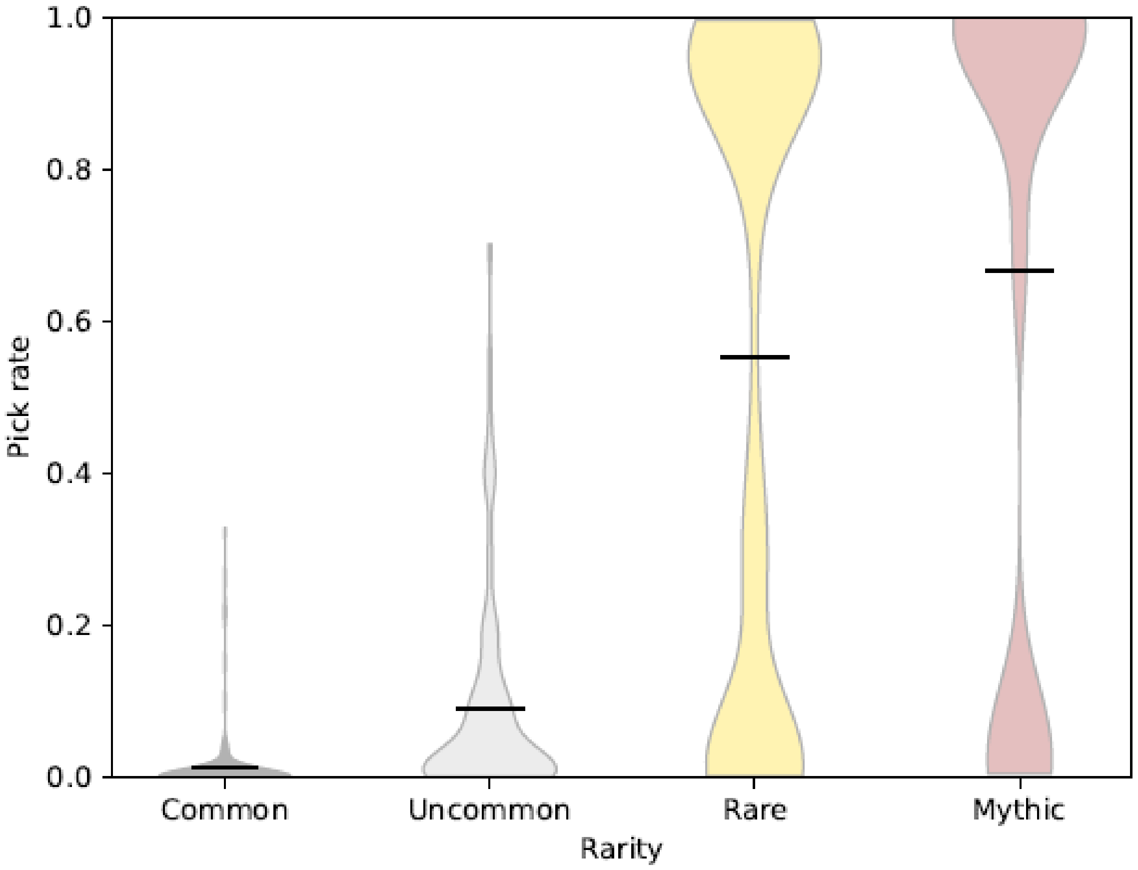
Data

Firstpickrate: Chance of card being chosen as the very first card



Pickrate: Chance of card being chosen when offered

Data


1.0
0.8
0.6
0.4
0.2
0.0
Pick rate
Common
Uncommon
Rare
Mythic
Rarity

Contextual Preference Ranking

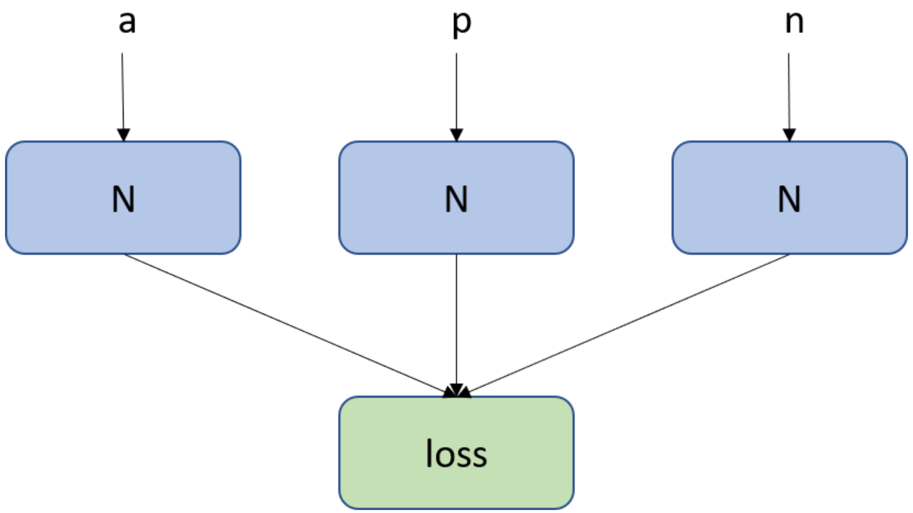
- Train neural network to predict human decisions
- Training data uses preferences between cards given previously chosen cards

Contextual Preference Ranking

- Much larger training data than directly predicting best cards
- Takes into account which cards are available
- Cards/sets are encoded one-hot

Training

- **Siamese architecture**, multiple forward passes through the network with different inputs (anchor, positive, negative) to compute their embeddings
- **Embedding**: D-dimensional output of network



Training

- **Triplet loss**: L(a,p,n) = max(d(a,p)- d(a,n) + m,0)

decrease

increase

- Novelty: Image recognition defines distances as similarity. We define distance as how good it fits

Testing

- Ranking over all choices instead of pairwise comparisons
- Compute distance of all choices to anchor
- Rank based on minimizing distance (i.e. maximize how well it fits the context)

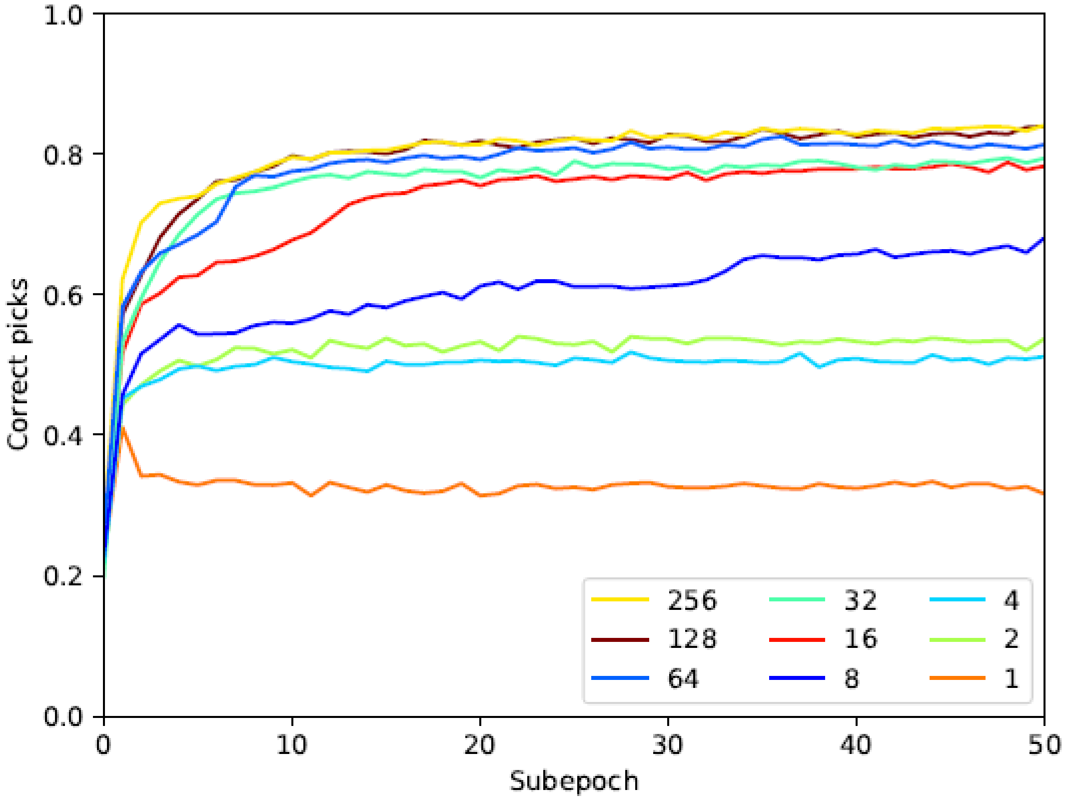
Performance

- We report two measures, mean testing top-one accuracy and mean testing pick distance

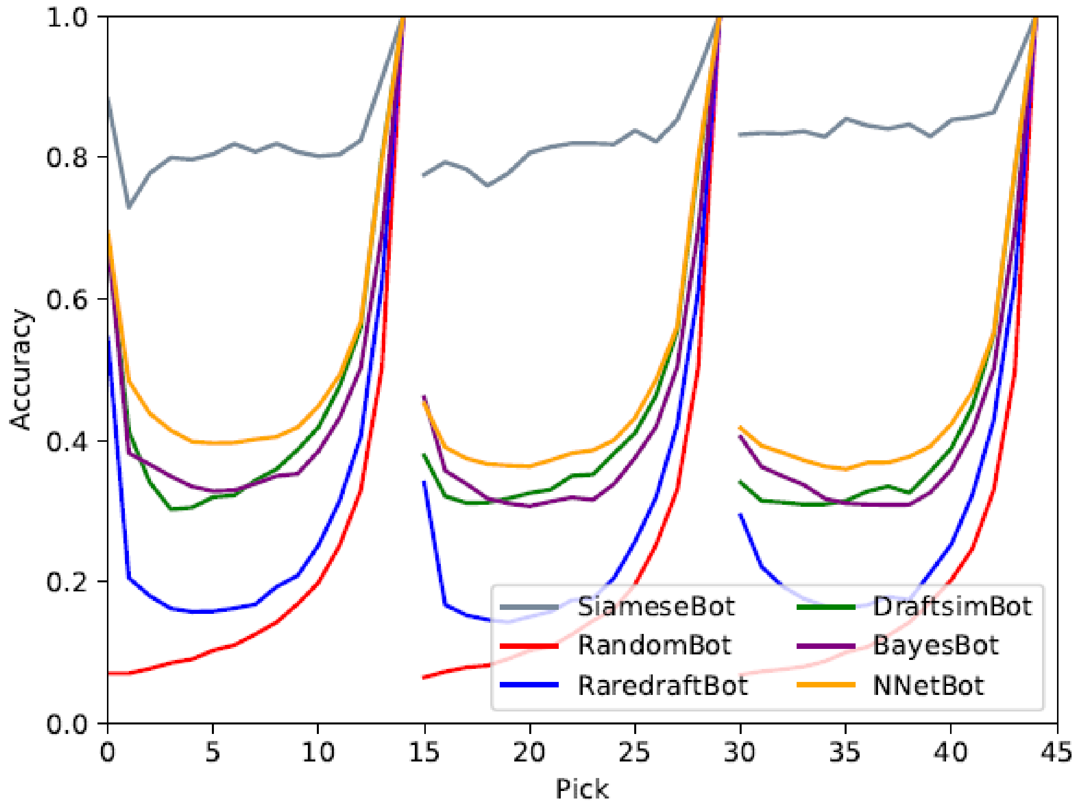
PERFORMANCE OF PROPOSED AGENT TO PREVIOUSLY SEEN:
[1] HEURISTIC AGENTS [16] [2] TRAINED AGENTS [16] [3] THIS WORK

Agent	MTTA (%)	MTPD
RandomBot[1]	22.15	NA
RaredraftBot[1]	30.53	2.62
DraftsimBot[1]	44.54	1.62
BayesBot[2]	43.35	1.74
NNetBot[2]	48.67	1.48
SiameseBot[3], D=2	**53.69**	**0.98**
SiameseBot[3], D=256	**83.78**	**0.2476**

Performance



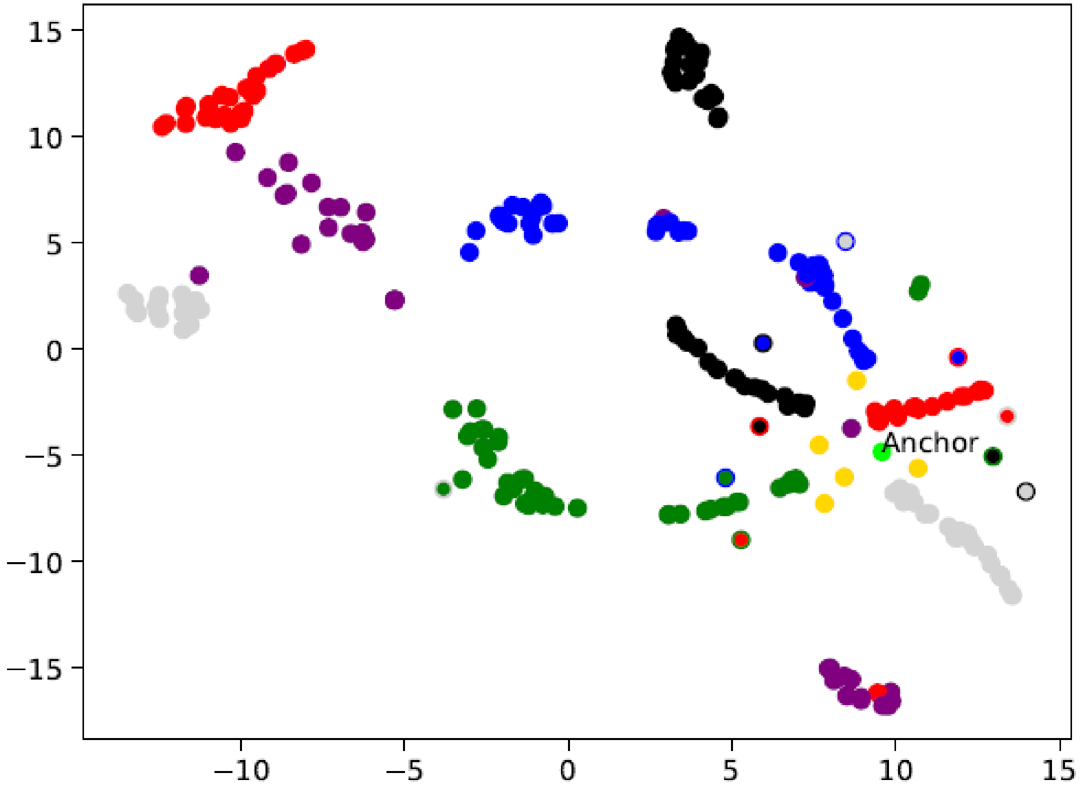
Performance



Embedding

- Apart from using distances to rank choices, we can visualize embeddings of cards
- Compute D-dimensional embedding of each individual card
- Compute 2-dimensional visualization with t-SNE
- Color each card in its color in the game

Embedding



Limitations

- Comparisons between Rares and Mythics are not occurring in dataset => ranking impossible

Card	Siamese	DraftSim	CFB	FPR
Spit Flame (R)	1	18	22	17
Leonin Warleader (R)	2	15	4	8
Goblin Trashmaster (R)	3	51	112	32
Ajani, Adversary of Tyrants (M)	4	7	5	1
Djinn of Wishes (R)	5	14	6	14
Tezzeret, Artifice Master (M)	20	1	2	3
Resplendent Angel (M)	30	9	1	2
Murder (U)	12	21	9	39

- Larger training data leads to longer training time

Considerations

- Evaluate cards/decks based on playing the game
- Moving from one-hot encoding to feature-based one
- Different domains (pregame selection, game playing, subset selection apart from games)

Summary

- Novel method for set addition problem
- Triplet loss using context, positive and negative
- Generating ranking of choices based on distances
- Achieved much higher accuracy than previously
- Approach is sensitive to embedding dimension
- Resulting embedding makes intuitive sense and can give intuition about data