Level-0 Models for Predicting Human Behavior in Games

Artificial Intelligence for Games (2019)
Prof. Dr. Köthe

Michael Hartmann

August 30, 2019

Abstract

This report is about the research of James R. Wright and Kevin Leyton-Brown concerning improved Level-0 models to accurately predict how humans will behave in games. The main focus of this report lies on the findings presented in the paper "Level-0 Models for Predicting Human Behavior in Games" which was published in 2019.
The basic concepts of these new models as well as their performance and the game theoretic models they seek to improve will be presented in this report.
All the research was done on a dataset called the *ALL10 Dataset* which included various unrepeated simultaneous move normal form games played by humans.
All the numbers and figures in the following report are derived from training and testing on this dataset by the original authors.

Contents

1 Introduction

1.1 Motivation

In most strategic situations that can be described as a normal-form game it is quite simple to find the best possible solution or a Nash equilibrium. But the idea behind behavioural game theory is not to describe optimal behaviour but rather the way actual people reason and decide when faced with a strategic decision.
Modelling this human decision making process is the goal of iterative models such as *Level-K* and *Quantal Cognitive Hierarchy*. Both models require a specification of non-strategic behaviour (Level-0) as a starting point for the iterative reasoning.
It is a common practice to define the behaviour of these Level-0 agents simply as uniform randomization over actions. The reasoning behind this is that it is generally believed that these non-strategic agents are merely a starting point for the reasoning of higher level agents and not part of the predictive model itself.
The research of Kevin Leyton-Brown and James R. Wright shows that this belief should be questioned and that defining a more accurate model of the non-strategic behaviour of humans can actually significantly improve the predictions made by these models.

1.2 Level-K

The Level-K framework defines a hierarchy of levels of reasoning. The bottom level, Level-0, is the non-strategic behaviour level. The definition of non-strategic behaviour is in principle up to the modeler as long as it does not include responding to explicit beliefs about other agents behaviour. The most commonly used behaviour for Level-0 is uniform randomization over actions.
A Level-1 Agent will assume that every other agent is a Level-0 Agent and will play the best response to the expected behaviour. The same goes for Level-2 agents they will assume that the whole population consists solely of Level-1 agents.
In theory there is no upper boundary for this hierarchy but it is generally advised to limit it to a reasonable amount of levels for the given scenario.

1.3 Quantal Cognitive Hierarchy

One of the main problems of Level-K is the rigid level structure where every agent of level i expects a homogeneous population of level i-1 agents.
Since a player does not necessarily fall under one of these archetypes Cognitive hierarchy extends the Level-K framework by enabling agents to reason about mixed populations.
A Level-3 agent might now base his decision for example on a population consisting of 30% Level-0 agents 50% Level-1 and 20% Level-2. This makes the model more flexible and in most cases more predictive than a standard Level-K hierarchy.

Quantal Cognitive Hierarchy is a special form of Cognitive Hierarchy consisting of two key components, *Quantal Best Response* and *Iterative Reasoning*.

Quantal Best Response
A very important aspect of human play is that it is in general imperfect. Small mistakes, like choosing the second best action rather than the best, occur comparably often, while big mistakes, like playing an action that always looses, are quite unlikely to be played.
Quantal Best Response tries to model this behaviour by making the agents respond stochastically to their incentives. This means that a higher utility action results in a higher probability to be chosen. By this logic a mistake becomes less likely the costlier it is.

$$QBR_i(s_{-i}, G, \lambda) = s_i(a_i) = \frac{exp[\lambda * u_i(a_i, s_{-i})]}{\sum_{a'_i \in A_i} exp[\lambda * u_i(a'_i, s_{-i})]} \tag{1.1}$$

The Quantal Best Response will always return a single mixed strategy s_i.
$u_i(a_i, s_{-i})$ is in this case the expected utility of agent i when playing action a_i against a mixed strategy s_{-i} in the game G. The variable λ, the *precision* represents the agents sensitivity to utility differences.
This is an important part of human decision making since people rarely compute exact utilities when playing a game. If two actions seem to be roughly equal concerning their utility they will usually be treated as such.

Iterative Reasoning
The Quantal Cognitive Hierarchy model requires higher-level agents to reason about the behaviour of lower-level agents and respond to them.
It is therefore very important to know the distribution of the different levels in the agent population to form an accurate response to their mixed behaviours.
For this paper a single-parameter Poisson distribution is chosen to describe the distribution of agent levels in the population.
This means we can predict the truncated distribution over actions for an agent i with level $0 \leq l \leq m$ with

$$\pi_{i,0:m} = \sum_{l=0}^{m} \frac{Poisson(l;\tau)}{\sum_{l'=0}^{m} Poisson(l';\tau)} \pi_{i,l} \tag{1.2}$$

τ is here the mean of the Poisson distribution.
With $\pi_{-i,0:m}$ being the truncated distribution over actions for agents other than i we can now write the predictions for agents of level m and of level 0 as

$$\pi_{i,m} = QBR_i(\pi_{-i,0:m-1;\lambda}) \tag{1.3}$$

$$\pi_{i,0} = |A_i|^{-1} \tag{1.4}$$

The predicted level-0 behaviour (1.4) is in this case just a uniform distribution over actions since this section just aims to outline the baseline model. Creating a more elaborate non-strategic behaviour is the topic of the next chapter.

2 Improved Level-0 Models

2.1 Non-strategic Behaviour

It is a common practice to equate non-strategic behaviour with uniform randomization when defining level-0 behaviour. This of course not very plausible when trying to model human behaviour. Level-0 agents do not reason over nor respond to the other agents strategies, but this definition does in no way imply that these agents should be strictly limited to random uniform action profiles.
Level-0 agents may still take account of payoffs of varying degrees. Maximising the own possible payoff or minimizing the possible loss does not require higher level reasoning and can therefore be considered non-strategic behaviour.
As a formal way to summarize this we will describe any behaviour as non-strategic that can be computed via a finite combination of *elementary agent models*.

We define these *elementary agent models* as follows:
An agent model for agent i is defined by a function $f_i(G)$ where G is the normal-form Game. This function maps to a vector of reals with the same dimension as the number of available actions.
If an agent model can be computed as $f_i(G) = h_i(\Phi(G))$ where $\Phi(G)_a = \varphi(u(a))$, with h being an arbitrary function, is *elementary*.
$\varphi(u(a))$ is in this case a linear combination of the players utilities with action profile a with weights defined by a vector $w \in \mathbb{R}^n$. $\varphi(u(a)) = w^T(u(a))$.

2.2 Level-0 Features

Now that we defined what non-strategic behaviour is more precisely we can define certain rules (features) that recommend actions to a certain degree.
These features can all be represented as *elementary agent models* and can therefore be directly computed from the normal-form game.
The features mentioned are all directly taken from the paper by James R. Wright and Kevin Leyton-Brown and do not cover all possible non-strategic features.
All the features have a real-valued and a binary representation.

Maxmin payoff - The best worst case
This feature recommends the safest decision, where the highest worst case payoff is achieved.

$$f^{maxmin}(a_i) = \begin{cases} 1 & a_i \in argmax_{a_i' \in A_i} min_{a_{-i}} u_i(a_i', a_{-i}), \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

The real valued version of this feature returns the worst-case payoff for an action:

$$f^{min[R]}(a_i) = \min_{a_{-i} \in A_{-i}} u_i(a_i, a_{-1}) \tag{2.2}$$

Maxmax payoff - The best best case
This feature recommends the action with the highest possible payoff

$$f^{maxmax}(a_i) = \begin{cases} 1 & a_i \in argmax_{a_i' \in A_i} max_{a_{-i}} u_i(a_i', a_{-i}), \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

The real valued version of this feature returns the best-case payoff for an action:

$$f^{max[R]}(a_i) = \max_{a_{-i} \in A_{-i}} u_i(a_i, a_{-1}) \tag{2.4}$$

Minimax regret - The minimal maximum regret
This feature recommends the action where the utility the agent could have gained by playing the best response to the other agents' actions is minimal.

If $r(a_i, a_{-i}) = \max_{a_i^* \in A_i} u_i(a_i^*, a_{-i}) - u_i(a_i, a_{-i})$
is the regret of agent i in action profile (a_i, a_{-i}), then

$$f^{mmr}(a_i) = \begin{cases} 1 & a_i \in argmin_{a_i' \in A_i} max_{a_{-i}} u_i(a_i', a_{-i}), \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

The real valued version of this feature returns the worst-case regret for playing an action:

$$f^{mmr[R]}(a_i) = inv[\max_{a_{-i} \in A_{-i}} r_i(a_i, a_{-1})] \tag{2.6}$$

Maxmax fairness - The "fairest" action

The difference between the maximum and minimum payoffs among the agents is used as a measurement for the unfairness in this feature.
$d(a) = \max_{i,j \in N} u_i(a) - u_j(a)$

$$f^{fair}(a_i) = \begin{cases} 1 & a_i \in argmin_{a'_i \in A_i} min_{a_{-i}} d(a'_i, a_{-i}), \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

The real valued version of this feature returns the maximum fairness that could result from playing a given action:

$$f^{fair[R]}(a_i) = inv[\min a_{-i} \in A_{-i} d(a_i, a_{-1})] \tag{2.8}$$

Max symmetric - The best response to oneself

This feature recommends the best response to the agents own action.

The symmetry is defined by the proposition
$Symm(u) \iff \forall i, j \in N, |A_i| = |A_j| \wedge \forall a_i, a_j \in A_i, u_i(a_i, a_j) = u_j(a_{j,a_i})$

$$f^{maxsymm}(a_i) = \begin{cases} 1 & Symm(u) \wedge a_i \in argmax_{a'_i \in A_i} u_i(a'_i, ..., a'_i), \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

For non-symmetric games it will evaluate to zero for all actions.
The real valued version of this feature returns the symmetric payoff of an action for symmetric games:

$$f^{symm[R]}(a_i) = \begin{cases} 0 & \neg Symm(u) \\ u_i(a_i, ..., a_i) & \text{otherwise} \end{cases} \tag{2.10}$$

Maxmax welfare - The best sum of utilities

This feature recommends the action where the sum of utilities for all agents is maximal

$$f^{welfare}(a_i) = \begin{cases} 1 & a_i \in argmax_{a'_i \in A_i} max_{a_{-i} \in A_{-i}} \sum j \in N u_j(a'_i, a_{-i}), \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

The real-valued version of this feature returns the maximum welfare that could result from playing a given action:

$$f^{welfare[R]}(a_i) = \max_{a_{-i} \in A_{-i}} \sum_{j \in N} u_j(a_i, a_{-i}) \tag{2.12}$$

Before we combine these features in order to get a decision, we have to evaluate their *Informativeness* in the given strategic situation that we want to evaluate.
With a special transformation we can zero out features that do not help the decision making process, due to recommending all actions to the same degree.

$$I(f)(a_i) = \begin{cases} f(a_i) & \exists\ a'_i, a''_i \in A_i : f(a'_i) \neq f)(a''_i) \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

This transformation helps producing much less noisy recommendations by ignoring uninformative features.
After this we normalize all the feature values to get non-negative values that sum to one, since otherwise high real valued feature values could potentially overwhelm other features.

To get a decision we now need to combine all these transformed feature values to produce a distribution of actions.
For this task we will use a *weighted linear level-0 specification* with F being a set of features and $w_f \in [0, 1]$ as a weight parameter.

$$\pi_{i,0}^{linear,F}(a_i) = \frac{w_0 + \sum_{f \in F} w_f f(a_i)}{\sum_{a'_i \in A_i}[w_0 + \sum_{f \in F} w_f f(a'_i)]} \tag{2.14}$$

2.3 Selecting a Model

An exhaustive evaluation of all subsets of binary features showed that certain feature combinations seem to vastly outperform others. Our goal in constructing a predictive model is to use as few features as possible while improving the performance of the model.

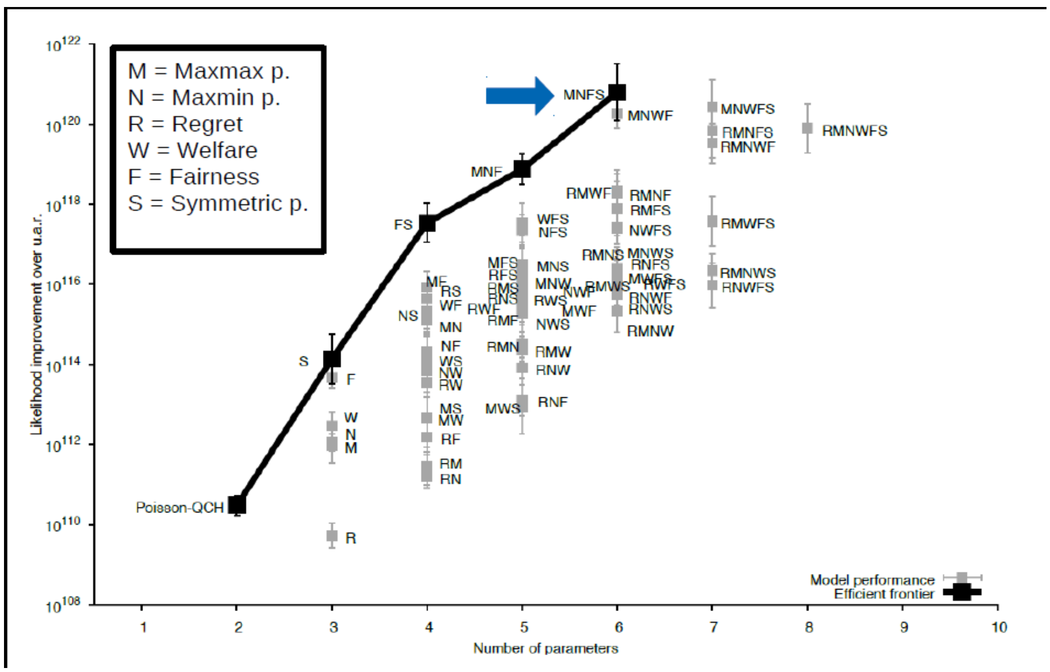

Figure 2.1: **The figure shows the differences in performance of the various combinations of binary features**

The best performing linear model, as seen in the figure combines the features **maxmax payoff, maxmin payoff, maxmax fairness, and max symmetric**. This model was called **linear4** due to being a linear combination of four features.

If we now extend the **linear4** specification with the real valued versions of the four binary features used, we get a specification called **linear8**. This model seems to, in most cases, outperform **linear4**.

A parameter analysis shows that this 8 feature model can safely be reduced to 7 features by eliminating the real valued fairness feature. This feature seemed to be not well identified and therefore not an important feature. The reduced specification is called **linear7**.

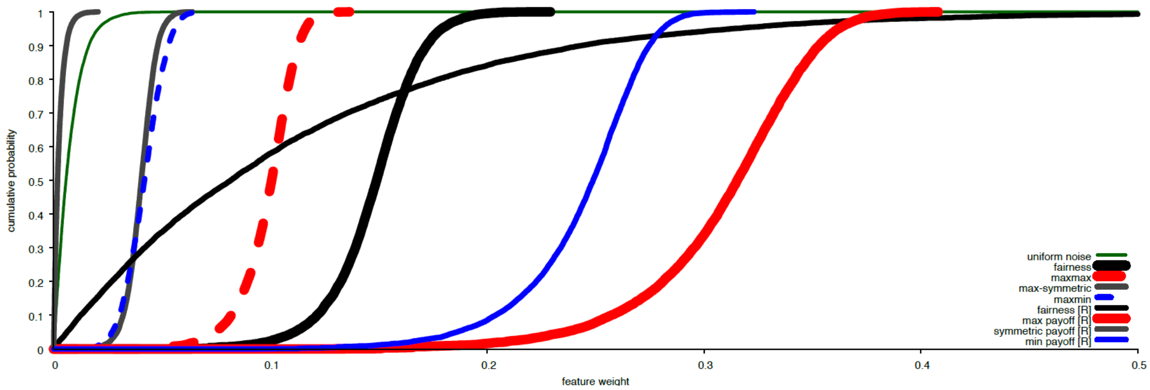

Figure 2.2: **Marginal cumulative posterior distributions over weight parameters of the linear8 specification showing the poorly identified fairness[R]**

Omitting the real valued fairness feature does by no means imply that fairness itself is not an important feature in human play. It merely shows that a real valued representation of it does not yield any improvement for the model. The binary version of the fairness feature on the other hand has a high weight and therefore is considered rather predictive.
As a comparison, the weights of the **linear4**:

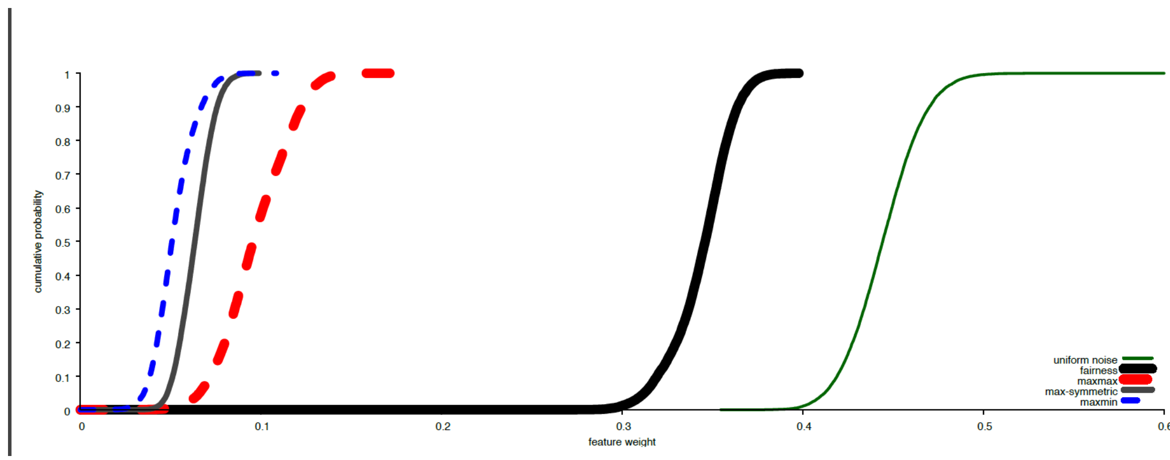

Figure 2.3: **Marginal cumulative posterior distributions over weight parameters of the linear specification**

Here the binary fairness feature is the highest weighted feature of all. The reason for that, as presumed by the researchers, is that fairness is highly predictive when it is present (not uninformative), but it is present in fewer games than most of the other features.

3 Conclusion

The results of James R. Wright and Kevin Leyton-Brown's research shows clearly the faultiness of the long-standing assumption that non-strategic agents exclusively exist in the minds of higher level agents.

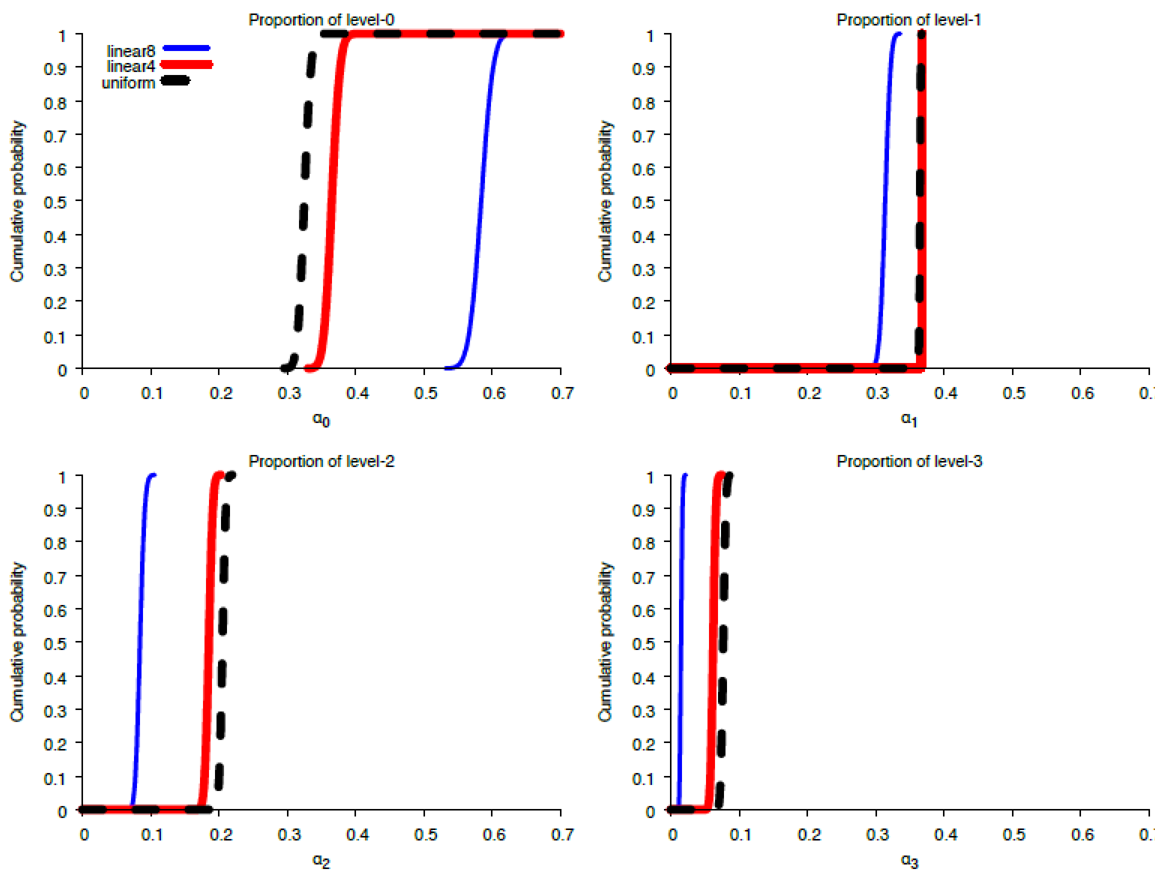

Figure 3.1: **Marginal cumulative posterior distributions of levels of reasoning for Poisson-QCH with linear8, linear4 and uniform specifications.**

The figure shows clearly that with a **linear8** specification the proposed amount of level-0 agents exceeds 50% of the agent population. This proves that non-strategic behaviour is an important aspect of the human decision making process, and that including sensible non-strategic behaviour in this kind of models greatly improves the performance in terms of predictiveness.

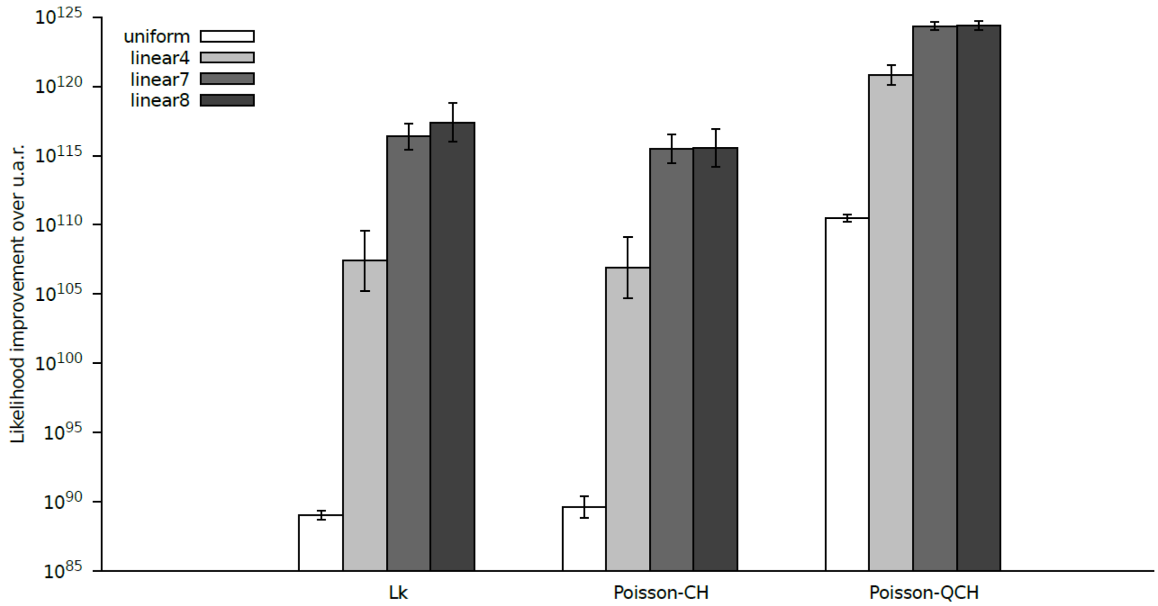

Figure 3.2: **Performance improvement with linear4, linear7 and linear8, compared to random uniform level-0 behaviour on level-K, Cognitive Hierarchy and Quantal Cognitive Hierarchy models**

The fact that this extended level-0 behaviour is solely derived from the normal form game itself makes it applicable to practically any domain of strategic decision making that can be represented in this form.

It is important to note that the dataset these models were trained on ,with 142 games and 13863 observations on human play, is comparatively small for a machine learning context in this field. The shown performance might very well differ from the results shown above for some applications.

Until now these models have only been trained and tested on two-player games. While the authors of the paper make the assumption that the models would probably show the same performance in multiplayer games since only the level-0 behaviour is changed it has not yet been proven and was left for future research.

References

1. Level-0 Models for Predicting Human Behavior in Games. J. Wright, K. Leyton-Brown. Journal of Artificial Intelligence Research (JAIR), volume 64, pp. 357383, February 2019.
`https://www.jair.org/index.php/jair/article/view/11361`

2. Level-0 Meta-Models for Predicting Human Behavior in Games. J. Wright, K. Leyton-Brown. ACM Conference on Economics and Computation (ACM-EC), 2014.
`http://www.cs.ubc.ca/~kevinlb/talk.php?u=2014-Level0.pdf`

Formalities

Michael Hartmann

- 3329547
- Angewandte Informatik (Bachelor)
- Seminar - Artificial Intelligence for Games