# 𑖖 (Qha)のPR文章

Ryoto Sawada, Yuki Ito, Toshihiro Shirakawa, Keigo Nitadori (Quorax 党 将棋部)

進捗を575でまとめると

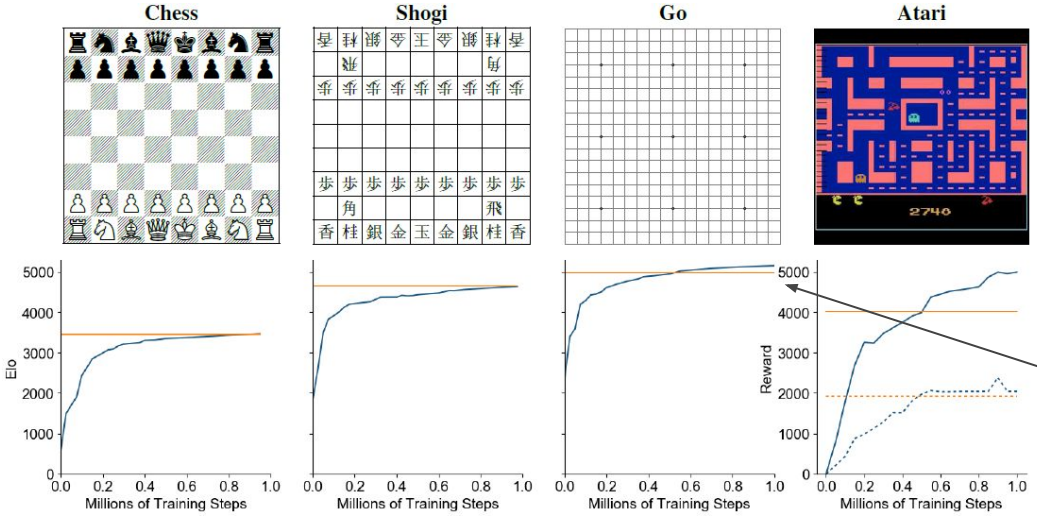
# MuZeroはAlphaZeroより弱かった

以下、MuZeroがどんなものであるか、どういう実験をしてどういう失敗をしたかについて解説します

大会で出すネタがねえとどうしよう

# そもそもMuZeroとは何者か

MuZeroはAlphaZeroの発展形のひとつ。探索時にゲームのシミュレータを必要としないことでより幅広い問題に適用することができる。また、原著論文によれば囲碁将棋などのゲームでもAlphaZeroと同等以上の性能を発揮している



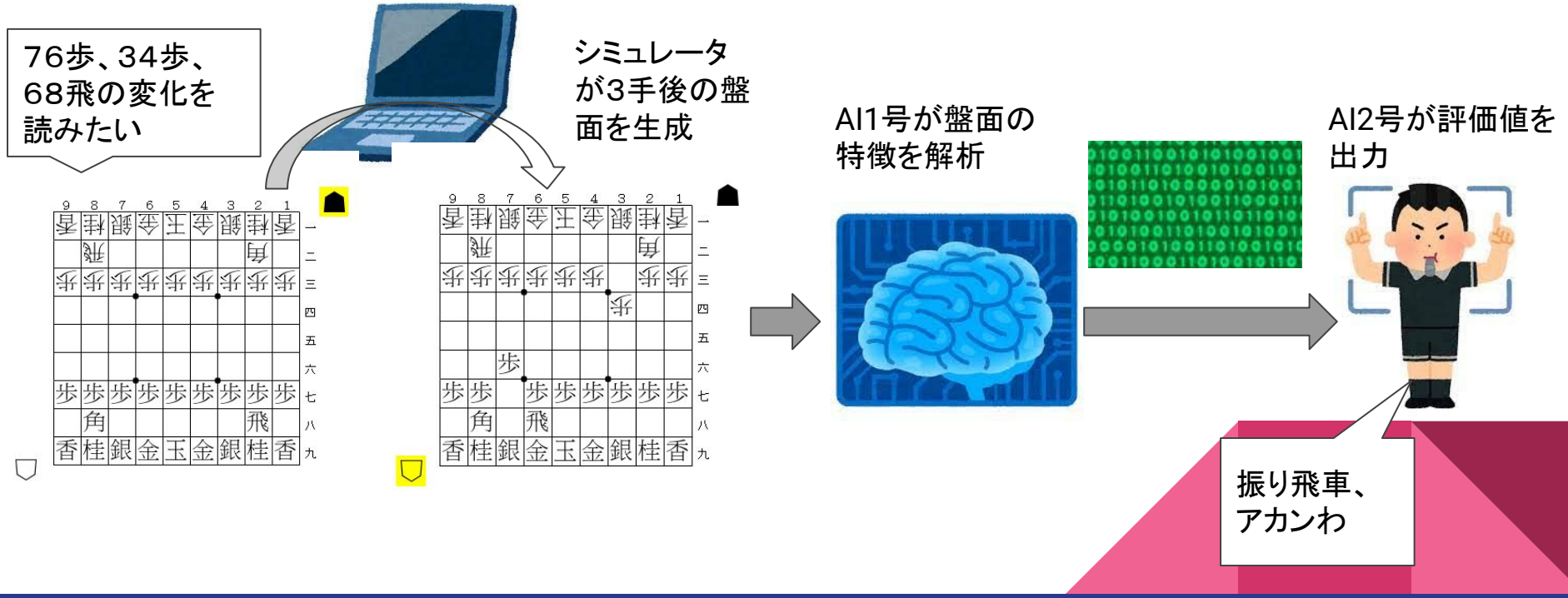

次ページからAlphaZeroとMuZeroの違いを解説

囲碁ではAlphaZeroよりも強いと言ってる

https://arxiv.org/pdf/1911.08265.pdf

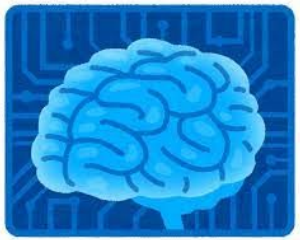
# AlphaZeroがどのようにして盤面を評価するか

AlphaZeroは局面を評価値に変換するAIとシミュレータ（図中のパソコン）からなる



# MuZeroがどのようにして盤面を評価するか

MuZeroは探索時にシミュレータを使わない



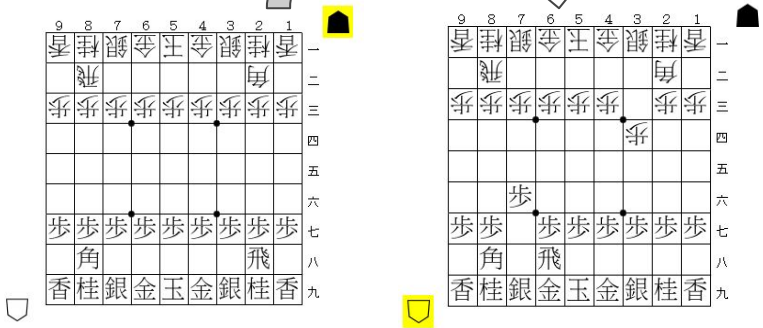
AI3号が指し手の情報から数手先の盤面の特徴を生成する

# AlphaZeroとMuZeroの違い

シミュレータが重いケース（テレビゲームなど）に対してMuZeroは有効

## AlphaZeroでも対応可能

シミュレータが3手後の盤面を生成（一瞬）



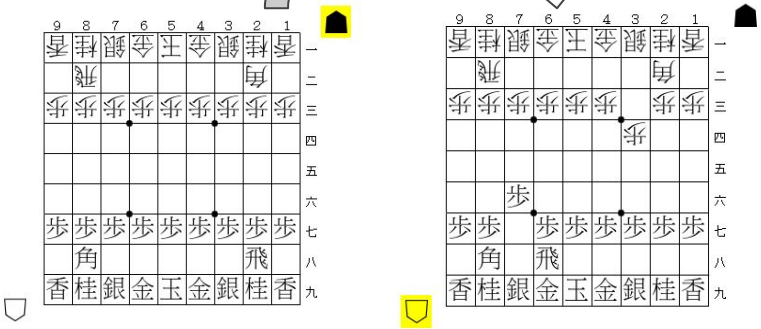
## AlphaZeroには厳しい

マ○オをジャンプさせた後の盤面を生成（高コスト）

# AlphaZeroとMuZeroの違い

囲碁、将棋などのシミュレータは極めて軽いため、MuZeroを使う必然性はない。
AlphaZeroとの優劣は数手先の局面をどれだけ正確に評価できるかで決まる

シミュレータが3手後の盤面を生成



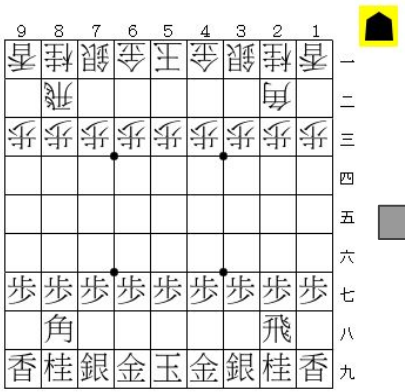
VS

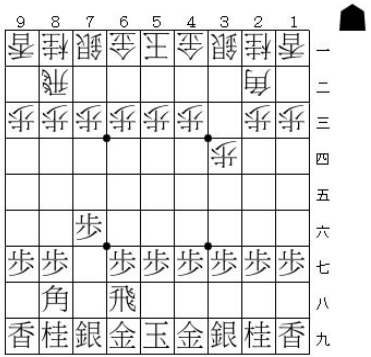
Q : 3手後の局面をどちらがより正確に評価できるか

# AlphaZero vs MuZero

AlphaZero派の主張：AIを使って3手後の局面を完全に復元できるとは限らない。シミュレータを使えば正確な結果が出るのだからシミュレータを使ったほうが良いに決まってる

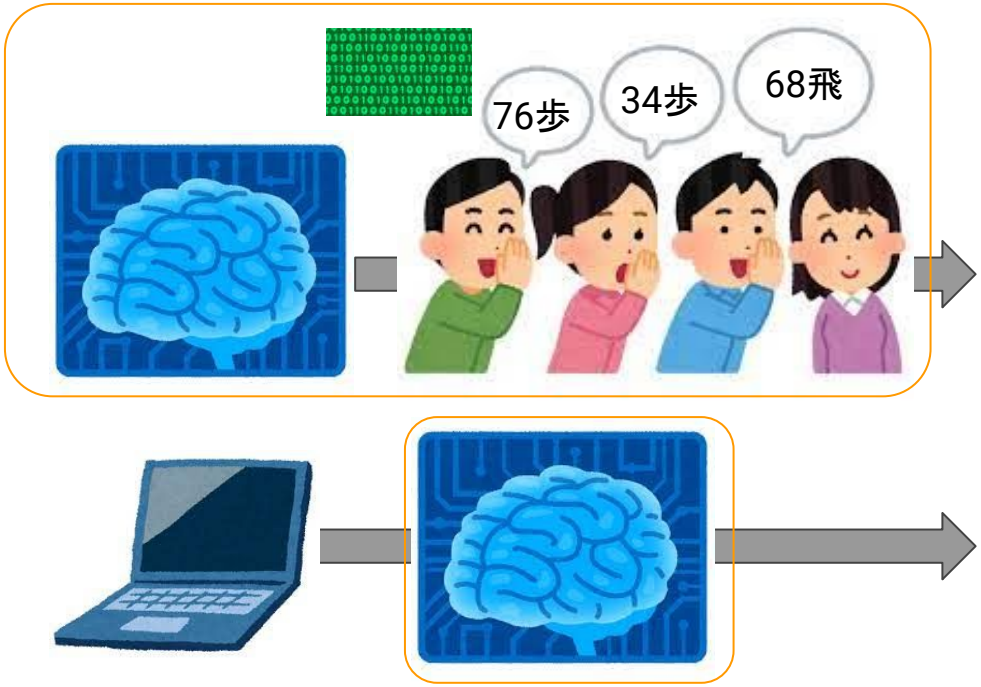
AIが将棋のルールを完全に理解できるの？ どこかでエラーがでたりしないの？



# AlphaZero vs MuZero

MuZero派の主張：AIのモデルは基本的に大きいほど精度が高い。盤面の伝言ゲームを経ることでより高度な情報を抽出することができる




評価値を計算するまでに経由するAIの数が多い → より沢山の計算を行っている → 精度が高い

This suggests that MuZero may be caching its computation in the search tree and using each additional application of the dynamics model to gain a deeper understanding of the position.
(原著論文より引用)

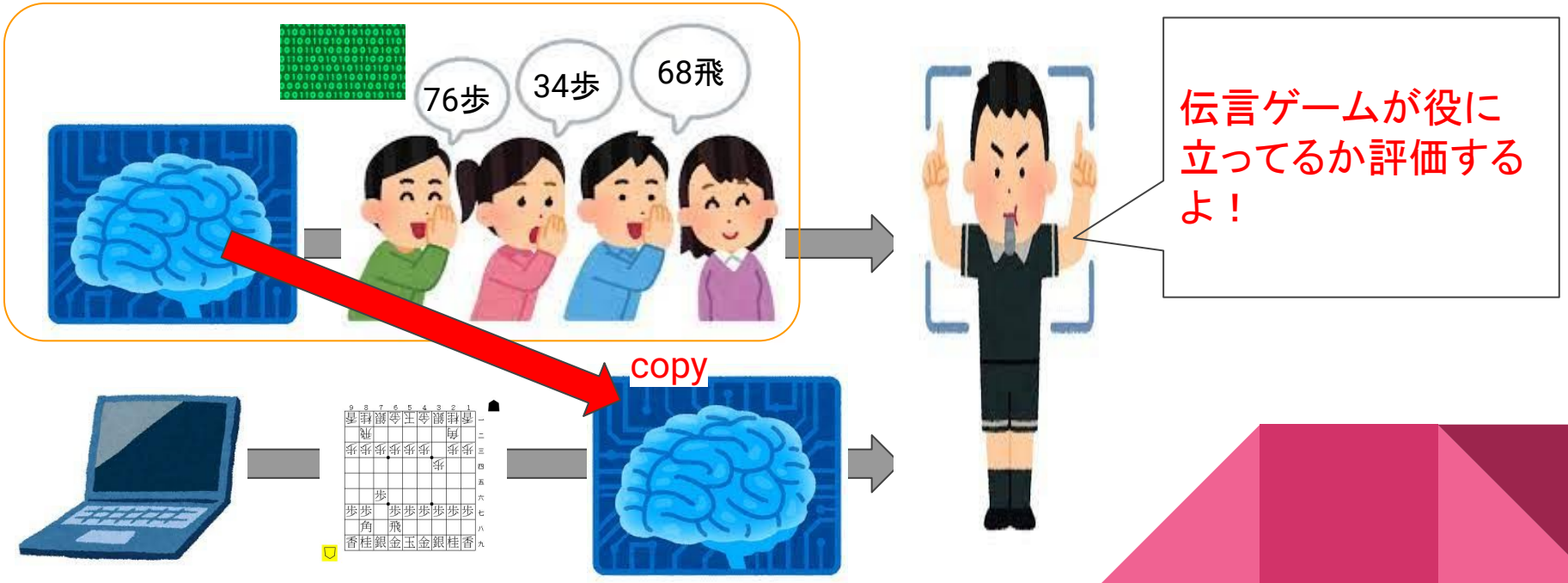
# 実験してみる

- python-dlshogi2を改造してMuZero形式に対応した → 全然勝てなかった
  - 1手1秒で20連敗したあたりで心が折れた

- 原著論文ではMuZero同士の対局では1手800 simulation(800nodesに相当？)、elmoなどとの対局では1手0.1秒で対局したらしい
  - スレッド数などの条件は不明。というか、1手0.1秒っておま、WCSC27のelmoってなどとツッコミどころは絶えない
  - ディープ系列同士の対局は同じ局面に偏りがちなので初期局面をどうするの問題もあまり触れられていないように見える

# 勝率以外の指標でも評価してみる

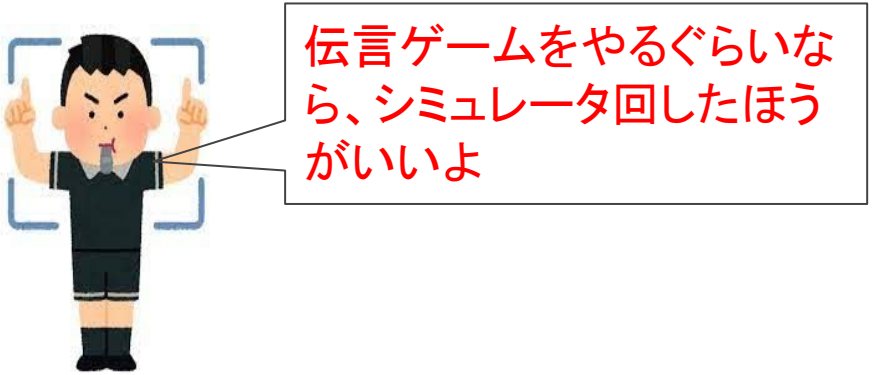
MuZeroが内包している盤面を特徴量に変換する部分を切り出し、AlphaZeroと同じ挙動をさせたうえで、各々のモデルの盤面評価制度を比較する



# 伝言ゲームの効果はあるのか

MuZero形式にすることで盤面評価精度が下がった

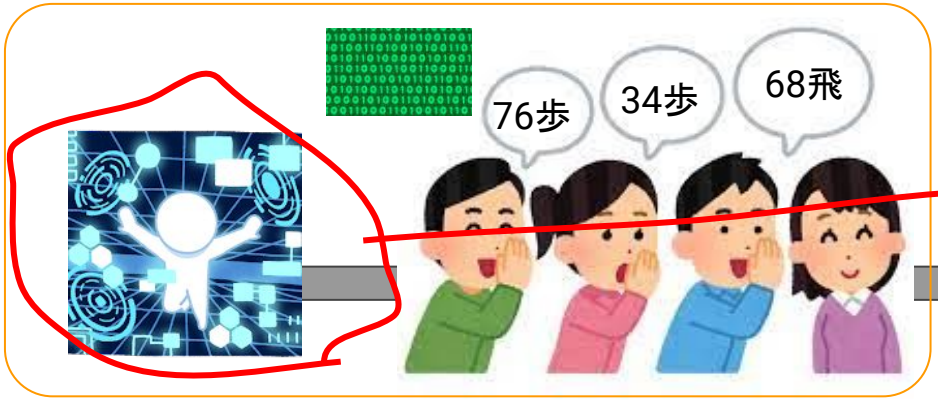
| | MuZeroの一致率（伝言ゲーム5回後の一致率） | MuZeroから切り出したAlphaZero部分＋シミュレータの一致率 |
|---|---|---|
| 公開モデルからのfinetune | 40.6% | 53.1% |
| ゼロからの学習 | 35.3% | 48.1% |




- train/validationはfloodgateの棋譜から生成したデータを利用（train:2000万、test:10万）
- MuZeroは「局面」ではなく「棋譜」のデータが必要で既存教師データを流用できない
- 教師の数としては正直頼りなくはある

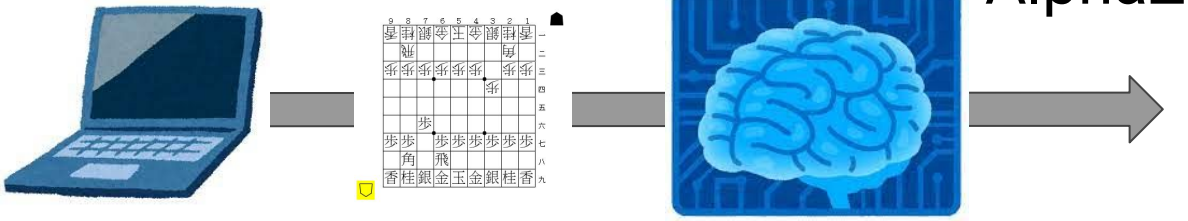
# 原著論文と実験結果の違いに関する考察

AlphaZeroとMuZeroとでは学習環境が厳密には一致していない。伝言ゲームが足を引っ張っていたとしても他の部分で強くなっていた可能性がある




MuZero

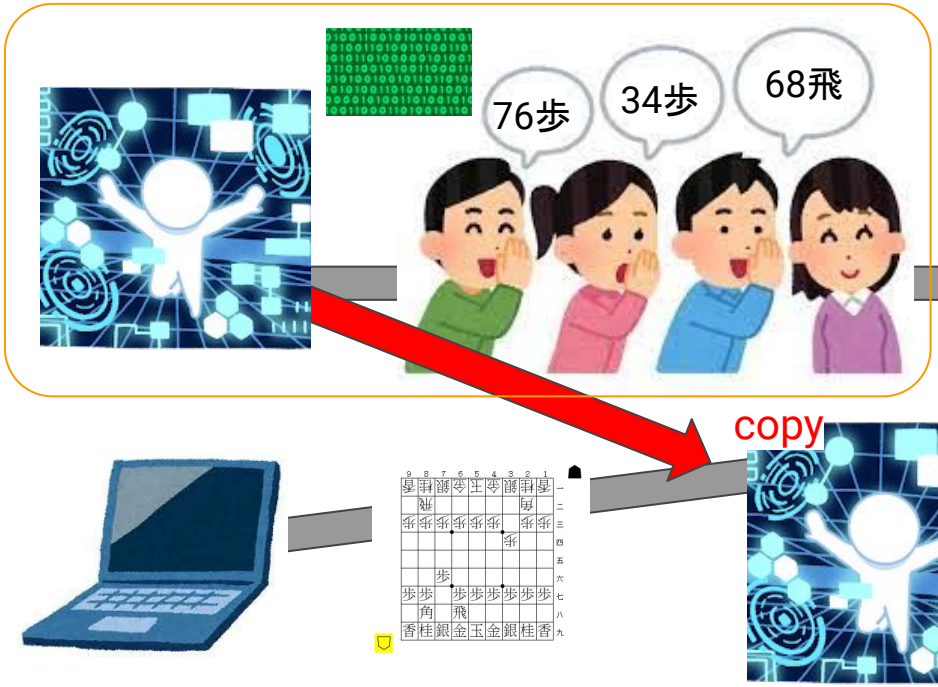
MuZeroAlphaZeroより強いと言っても、ここが強くなっただけかもしれない



AlphaZero

# （rsawadaの考える）原著論文の問題点

AI3号（原著論文でいうdynamics）が盤面評価の精度にどの程度影響を与えているかを評価するべきだった




MuZeroはAlphaZeroを内包してるのだから出来るはず。つか、学習ログ解析すれば生成できるよね、論文のレフェリーは居眠りでもしてたの？

# 原著論文と実験結果の違いに関する考察（論文を擁護）

- そもそもrsawadaの再現実装が失敗している
  - 学習の条件に敏感な手法であるとか、教師の数が少ないとか
  - とくに、伝言ゲームのネットワークが浅すぎた（伝言ゲームがシミュレータの質を超えるにはある程度のlayer数が必要）はありそう

- シミュレータ不要であることが売りであり、そもそも強さは売りにしてない
  - 原著論文では囲碁でAlphaZeroより強いことをあまり推してない（かも）
    - いや、abstractでもガッツリ触れてるな

- 大会が終わったらソースコードを公開するので遊んでみて欲しい
  - ~~コメント内に書き込んだDM社への悪口を消さないと~~

# 困ったぞ、出すものがない

この時期にコンテンツが
何もないのは数年ぶりだ

- 今から頑張って飛車を振る
  - ~~負けても振り飛車のせいにできる~~
- 実況を頑張る
  - 実況ツールの公開もやらないと（linuxはともかく、windowsで動かない）
- 別に負けてもいいやとMuZeroを放り込む
  - これ面白いか……？
  - 教師データの生成から学習ルーチン、探索部まで創っておいてお蔵入りするのも癪だけど
  - ~~もう全部、論文のレフェリーが悪い~~