Evaluating Go Game Records for Prediction of Player Attributes

Josef Moudřík
J.Moudrik@gmail.com
Faculty of Math and Physics, Charles University in Prague

Petr Baudiš
Independent researcher

http://gostyle.j2m.cz

Introduction

- There exist large collections of structured .sgf Go records
  - KGS Archives, GoGoD, gobase.org, gokifu.com, …
- So far, the computer analysis of the data is mainly limited to:
  - statistics of next move in fuseki and learning heuristics to improve Computer Go tree search
- Can we use the data differently?
- E.g. to answer questions like this one?

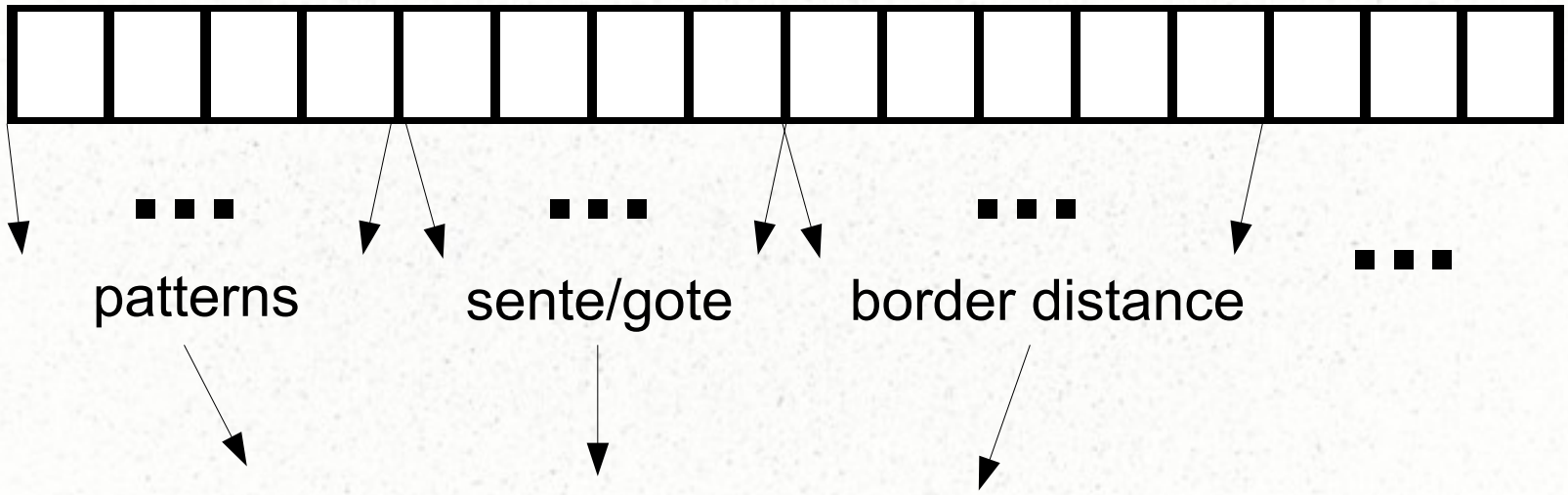
What constitutes the strength?

Our approach

- Given a **set of games G**, create an **evaluation** which
  - robustly describes the set of games for a player, by making statistics of certain events in the games
    - e.g. sente/gote use, most frequent patterns, high/low plays, ...

The Evaluation

- We process the games **G** one by one and summarize the different statistics (features) in a following vector:



F-1: Pattern Feature

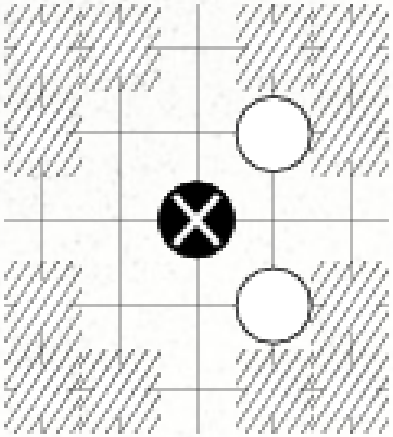
- Frequency of 400 (in the paper) most frequent spatial patterns
  - the patterns are centered, normalized to be black to play and invariant under rotation and symmetry

1) Find the 400 most frequent patterns

2) We count how many times does each of the 400 patterns occur in the set of games **G**

3) E.g. | 20 | 7 | 0 | 11 | ▪▪▪

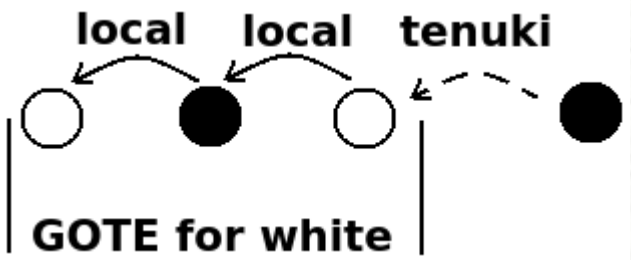
4) Normalize this, by dividing by #games in G

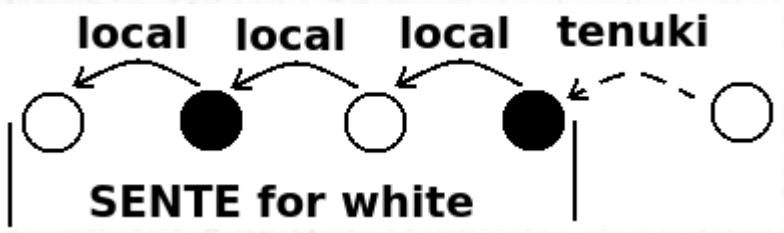


An example spatial pattern

F-2: ω-local Sente and Gote Sequences

- **Assumption:** Reply to a sente move is always local.
- **Definition:** A move is ω-local if it its distance to previous move is smaller or equal to a fixed number ω.
- Allows to approximate what is sente/gote:

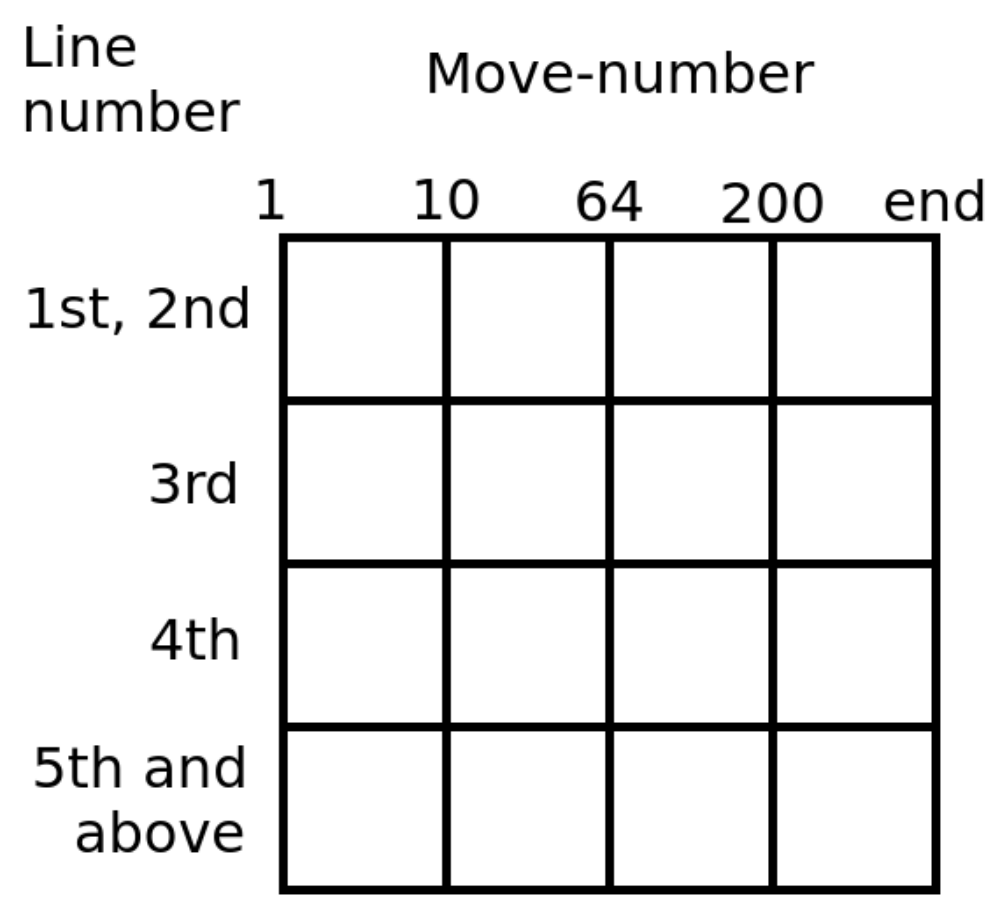




- The feature: Average number of sente and gote sequences per game and the difference between them.

F-3: Border Distance

- For each of the games in **G**, we count number of moves which fall into a particular bin.
- Numbers in the bin are averaged for all games in **G**.



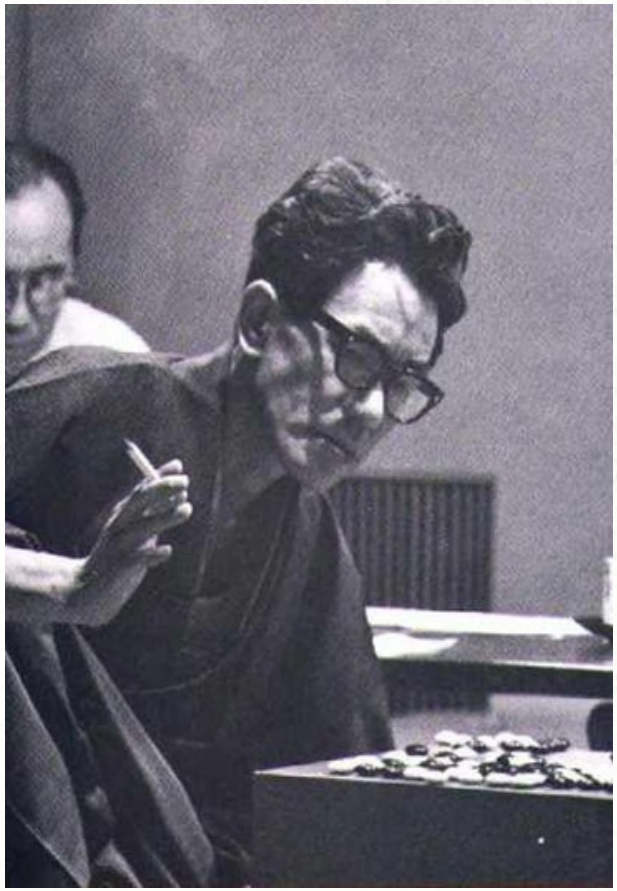
A Little Quiz

- Who is this?

Line number	Move-number 1	10	64	200	end
1st, 2nd	2%	18%	29%	43%	
3rd	60%	28%	14%	9%	
4th	35%	22%	13%	11%	
5th and above	3%	32%	43%	36%	

A Little Quiz

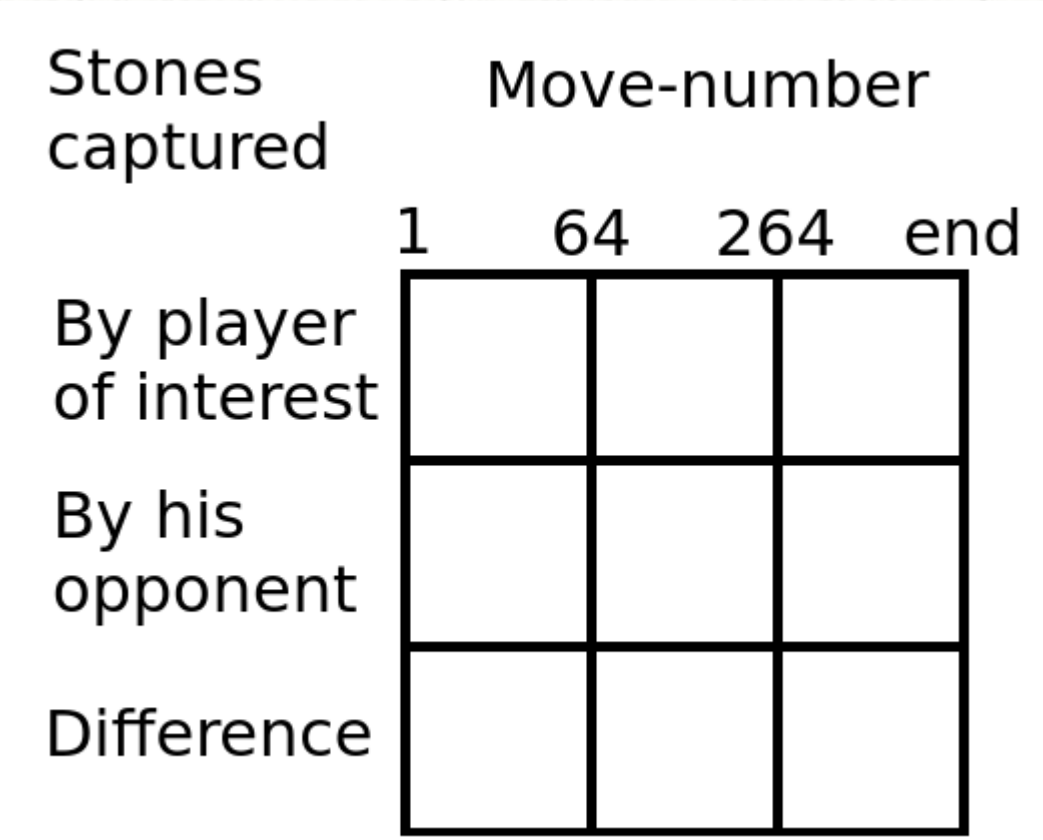
- Master Sakata!



Line number	Move-number 1	10	64	200 end
1st, 2nd	2%	18%	29%	43%
3rd	60%	28%	14%	9%
4th	35%	22%	13%	11%
5th and above	3%	32%	43%	36%

F-4: Captured Stones

- For each of the games in **G**, we count number of moves which fall into a particular bin.
- As before, numbers in the bin are averaged for all games in **G**.



F-5: Win/Loss Statistics

- From games **G**, we also count how many times did the player (on average):
  - win by counting
  - win by resignation
  - lost by counting
  - lost by resignation
- For games lost and win by counting we also count average size of the win/loss.

What to do with the evaluation?

- Strength prediction:
  - data from KGS Archives
  - for each rank from 20-kyu to 6-dan, we gathered 120 players at the particular rank
- Style prediction:
  - data from GoGoD and our questionnaire
  - for 25 professionals, we gathered 12 sets of 16 games

Style	1	10
Territoriality	Moyo	Territory
Orthodoxity	Classic	Novel
Aggressivity	Calm	Fighting
Thickness	Safe	Shinogi

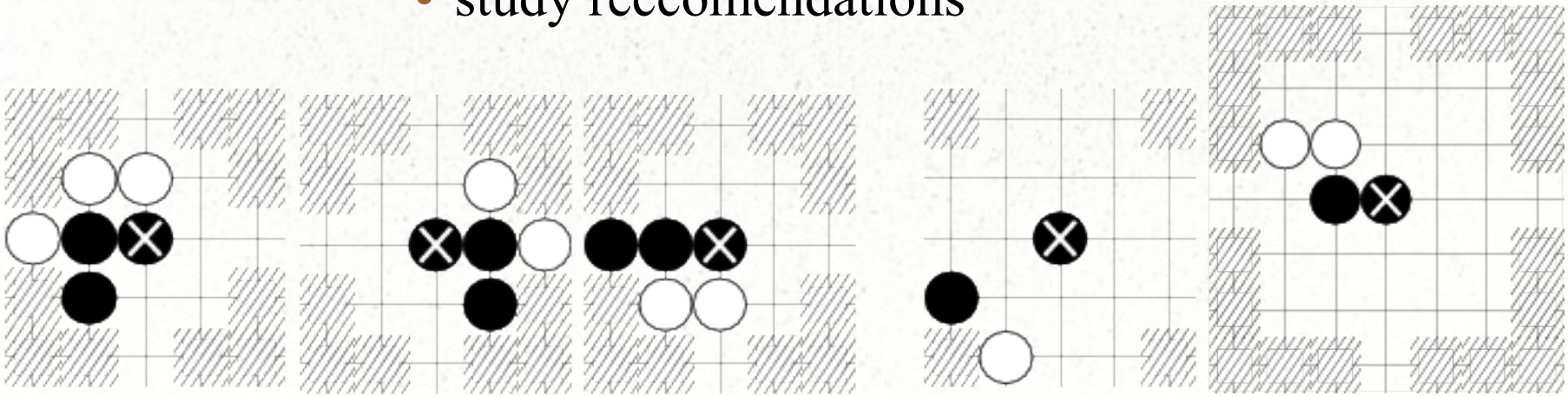
Prediction Methods and Results

- Machine learning method of Bagged Artificial Neural Network, see the paper for details
- Strength:
  - std. error of 2.712 rank
- Style:
  - std.error of ~ 1.55

Territoriality	Orthodoxity	Aggressivity	Thickness
1.527	1.734	1.548	1.572

Applications and Discussion

- Style prediction:
  - reccomend relevant pros to review
- Strength prediction:
  - help ranking systems to converge faster
  - correlations between strength and pattern frequency
    - study reccomendations



Realized at http://gostyle.j2m.cz

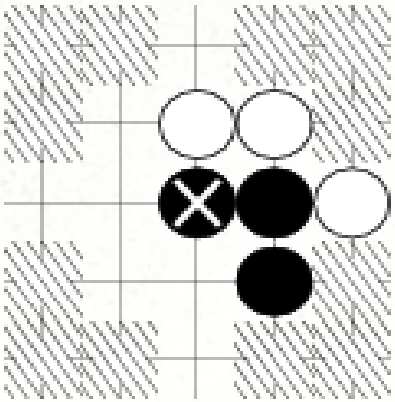
Thank you for your attention!

Please contact me if you have any remarks!

J.Moudrik@gmail.com

Visit our web-application at:

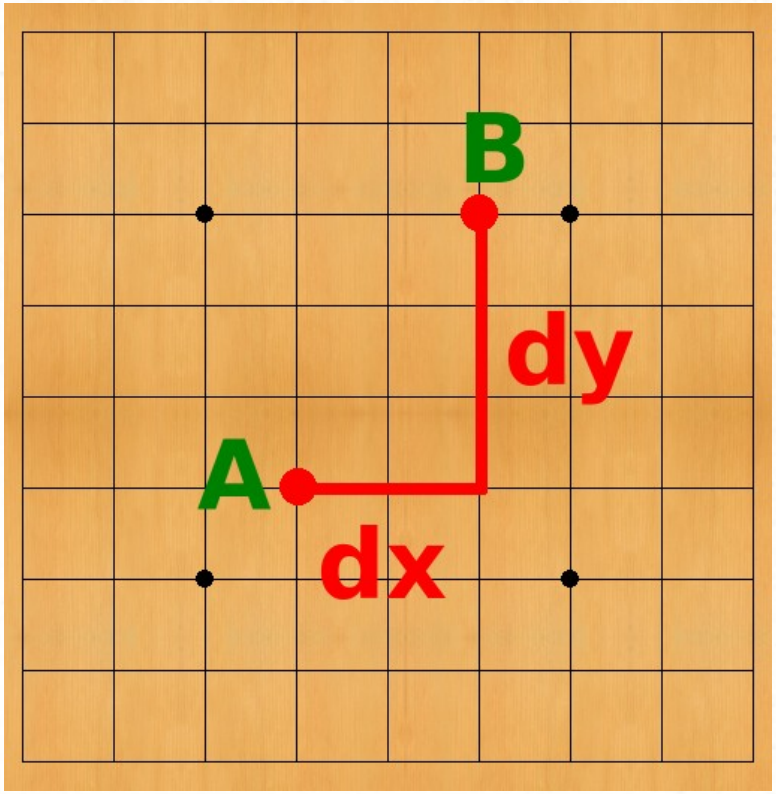
http://gostyle.j2m.cz



Gridcular Metric

- A function we use to measure distances on the goban:

$$distance(A, B) = dx + dy + max(dx, dy)$$

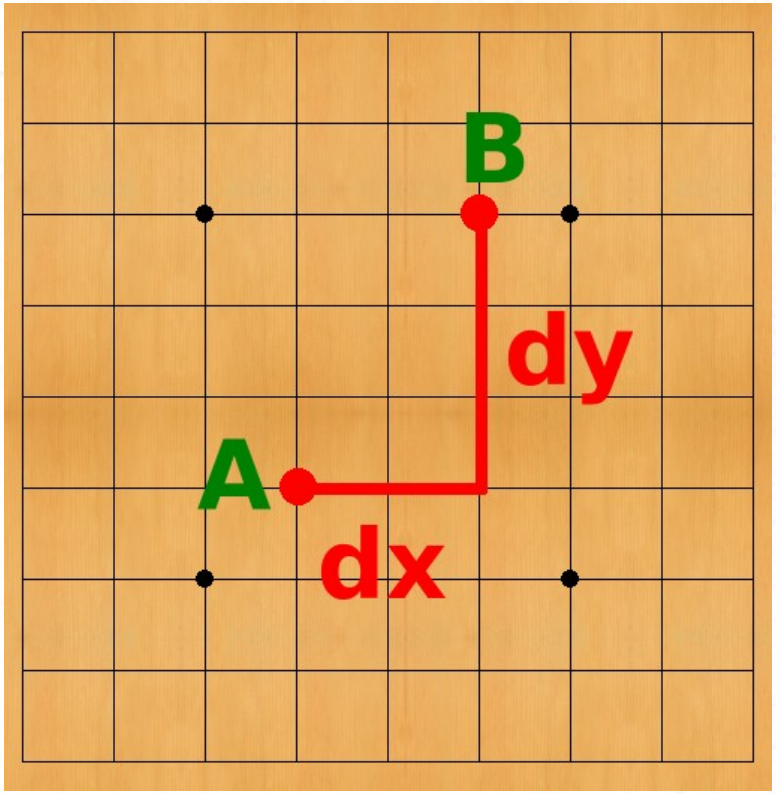


9	8	7	6	7	8	9
8	6	5	4	5	6	8
7	5	3	2	3	5	7
6	4	2	0	2	4	6
7	5	3	2	3	5	7
8	6	5	4	5	6	8
9	8	7	6	7	8	9

Gridcular Metric

- A function we use to measure distances on the goban:

$$\text{distance}(A, B) = dx + dy + \max(dx, dy) = 8$$



2 3 2 3

9	8	7	6	7	8	9
8	6	5	4	5	6	8
7	5	3	2	3	5	7
6	4	2	0	2	4	6
7	5	3	2	3	5	7
8	6	5	4	5	6	8
9	8	7	6	7	8	9