Analysis on possible usage of Leela software as aid in PGETC games of Carlo Metta

Bojanic Milos,
Belgrade, Serbia

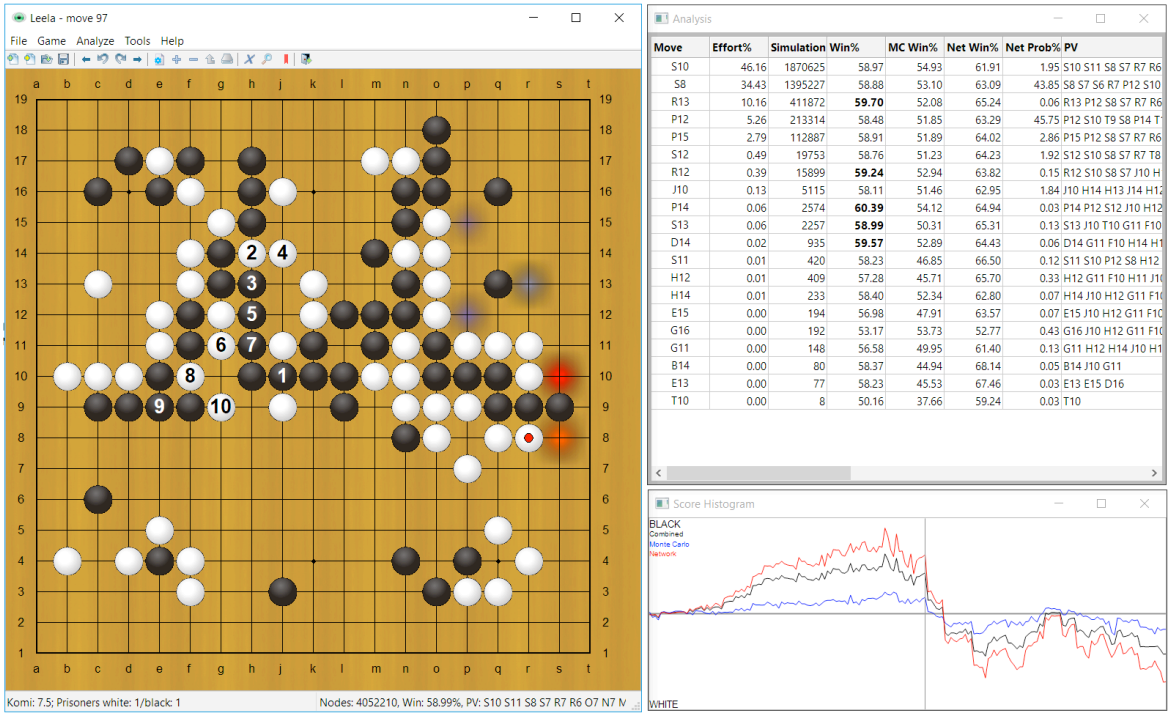
Introduction

For eight years, European team championship has been played on internet. This move made this championship more attended than ever. Since games were played online, it also opened opportunity that someone might use illegal help, such as joseki, puting position on board, or even help from a friend. Since those were national teams, all best players were in a team, so finding someone to help was very difficult. Using other means would not help your play much (at least in the A league). If we add to it that all top players knew each other for quite some time and in some cases had very friendly relationships, cheating was not so likely.
However, things have changed a lot since then. Now there are available programs for desktop computers which are stronger than european players. On the other hand, there is more players who play mainly online, without chance on meeting other players.

Leela

This program is one of the first programs available to wider audience that was stronger than european players. First version was introduced in April 2016. Version 0.10 was released in May 2017, and current version 0.11 in October of same year.
In game analysis, Leela can do several very useful stuff:

- it can estimate current winning percentage,
- show best moves for both players,
- analysis is quick, it can be done usually in less than a minute,
- Leela's weaknesses are Life and Death problems, and like other programs, it's ko play is different than human's.



Possible ways to use Leela as illegal help in internet games

Leela can be used for cheating on several ways:

- player can use win percentage data in order to estimate position.
- player can use software to check his own moves, or sequences, such as in life&death. Although in both cases player is not directly taking suggestions of moves from software, it is cheating.
- player can directly use move suggestions from software. It can be for entire game, or just for some critical moves, which can still can raise player's strength significantly.
- Player can chose to use help from software only in some games, in order not to arouse suspicion. If someone uses software aid only in games with players of his strength or weaker, it can go unnoticed, but it improves winning percentages significantly.
- Finally, human ingenuity knows no boundaries, so there could be even more ways to cheat.

Please note that several of those methods basically offer very little chance of being caught. Even directly using program's suggestions can go unnoticed in some cases, and difficult to prove.
Also, there are more and more programs now, and analysis must be made separately for each program, which complicates analysis further more.

Current analysis of games that are suspected of using Leela

So far, one game was brought to attention as suspected of using Leela software. In round 4, match Italy-Israel, game Carlo Metta – Reem Ben David, Israeli team noticed that most of Metta's moves was very similar to Leela's, and filed a complaint, on grounds that 98% of moves were similar to Leela's choices. League manager decided that software was used. However, after Italian appeal which claimed that it was not 98% but something less, and that in other games so high percentage of similar moves was found, PGETC Appeals commission reversed original decision.
After that, another statistical analysis was made by Ales Cieply.

Problems with statistical analysis

As it is already pointed out, when one player is playing stronger in online tournaments, it is indeed cause for suspicion, but it can not be taken as proof.
In one of early analysis, we used histogram of deviation of one player's moves to Leela's. It shows how much each move was better (green bar) or worse (red bar) than Leela's. However, especially in short fighting games of players of great difference in strength, such diagram can look similar to Leela's, since most moves are forced, and strong european players are not much weaker in fighting.

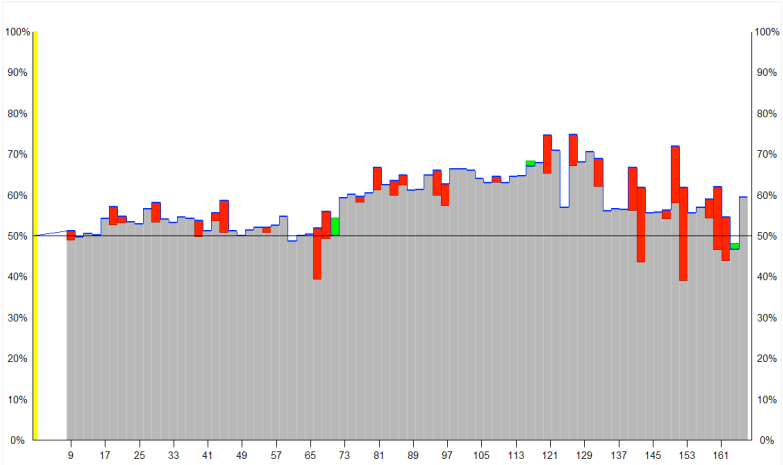


Diagram 1: Sample histogram of game between top european players (Jabarin-Junfu, round 2)
Lot of mistakes (in Leela's opinion) are visible, and player's chances of winning are changing often.

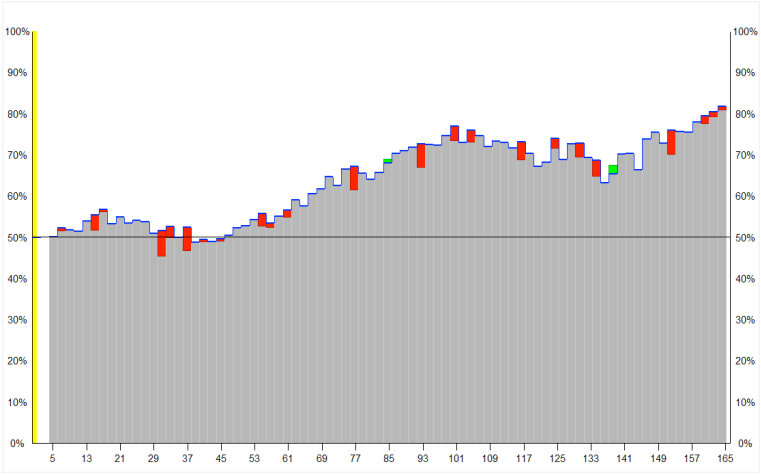


Diagram 2: Sample histogram of game Metta-Ben David.
Few not so big mistakes are visible on Metta's side, and after he took lead, there was no major reversals.

Please note that way that computers calculate who is in the lead differs much than to way human's think. Programs use calculations to find winning chances, and they can often play move is considered safer, not the best one, especially in the endgame, if they consider it safer. Even top human players do not have winning histogram similar to programs. Therefore, seeing diagram like this is definitely suspicious, but again, it is not a proof.

In statistical analysis that have been conducted so far, while focusing on sheer percentage, one key factor was skipped. One can play entire joseki, or complete opening, that can be very similar to program. It can be learned easily even by beginners.
In close fight, lot of moves can be forced, and if there could be few options per move, even several moves in a row could easily fit Leela's top suggestions. In endgame, there could be several large points, and playing one of program's suggestions is almost guaranteed.
However, hitting constantly Leela's top choices in above mentioned cases is rather suspicious.

New method of analysis

Because of presented problems, in this analysis I decided to go for different approach. Instead of statistically analyzing entire game, I focused only on key points in the game. For instance, what will one player do in the middle game if he is not forced? That is definitely move that player plays on his free will, and it can show his strength. Such positions are often used in go problems, and based on results in solving them, player's strength is often given.
In other part, I analyzed sequences of moves, and how many of them were similar to Leela's. It is expected that some moves were same, since they are either forced or easy to find. However, it would be highly unlikely that one player would play all top suggestions of Leela in several fights.

For purposes on analysis, I examined total of 4 games played in period of May 2017. - May 2018, two games played on internet and two live.

- League A qualifications 2016/17, Martin Ruzicka – Carlo Metta
  Middle game tenuki moves were: 70, 88, 100, 104
- League A 2017/18, Carlo Metta – Reem Ben David (both games played before Israeli report)
  Moves 51, 59, 65, 87, 97, 101,
- WAGC round 2, Kim – Carlo Metta (live game)
  Moves 34, 60, 64,
- WAGC round 6, Carlo Metta – Oscar Vazquez (live game)
  Moves 75, 77, 119

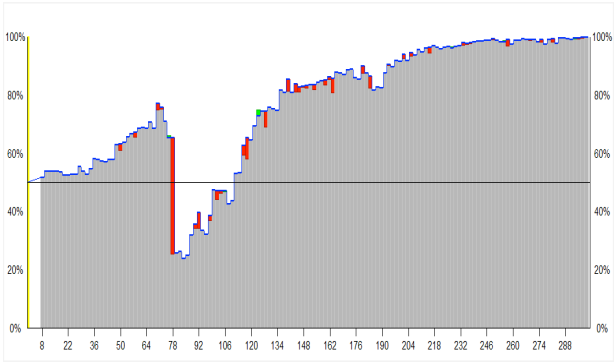
Methodology of analysis

Same method as potential cheater was used, game analysis was started and move suggestions were observed. Changes were observed until 50.000 variations, after which few changes were spotted. For each move, it was written when it appeared in suggestions. Please note that early in analysis things rapidly change.

Analysis of game 1

- League A qualifications 2016/17, Martin Ruzicka – Carlo Metta

move 70 – A (top Leela suggestion), appearing from 9k variations
Please note, also life&death mistake.
move 88 – A , from 2k
move 100 – A , from 2k
move 104 – A from 2-10k, after it was low level. Note that It was follow-up move to A choice, which was peep.



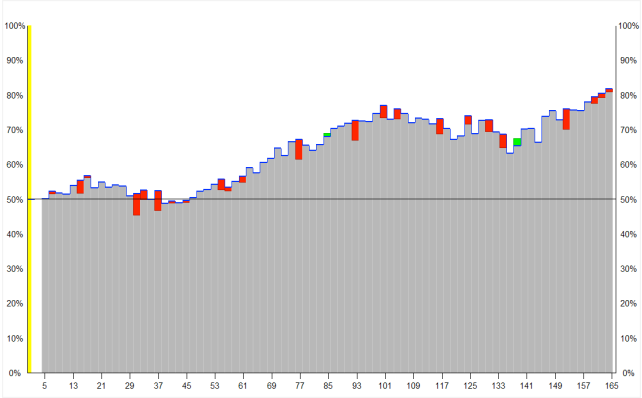
Please note that at the time actual version was Leela 0.10, which we used in analysis (results were similar to version 0.11).

Regarding fighting and middle game sequences of moves, not only they had small deviations to Leela's play which can be observed in histogram, but most moves (from 30 to 110) were Leela's top choice.

Analysis game 2

- League A 2017/18, Carlo Metta – Reem Ben David

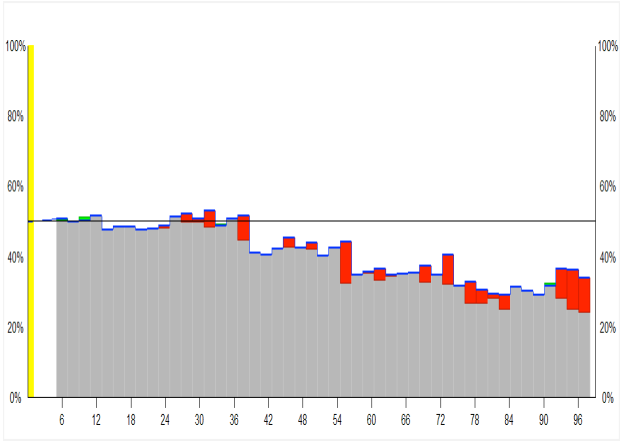
move 51 – A , from 2k variations
move 59 – A , from 40k
move 65 – A , from 2k
move 87 – A , from 5k
move 97 – low suggestion, although sente move. For this move, move 101 was A suggestion from 10k.
move 101 – A from 2k to 20k, and after that B. Sente combination of 97 made other moves better.



Again, most of the moves (from 30 to 105) in middle game and fighting were Leela's top choice.

Analysis game 3

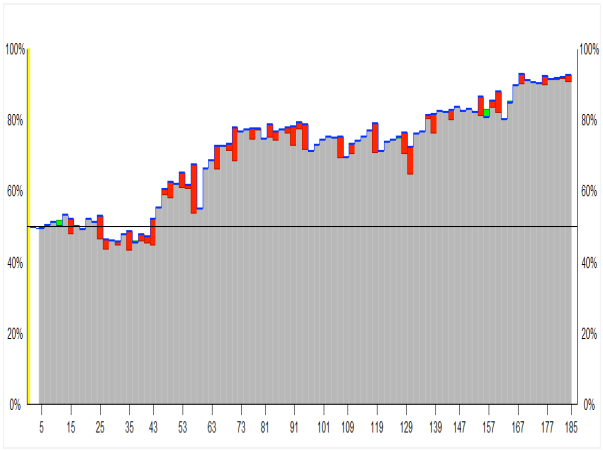
- WAGC round 2, Kim – Carlo Metta (live game)
  move 34 – not in Leela's suggestions
  move 60 – low level suggestion,
  mid level from 20k
  move 64 – not in Leela's suggestions,
  although major suggestions were nearby



Moves (from 30 to 99) in fighting and middle game varied a lot. Some of the moves were like Leela's top suggestions, but part of them were forced. Some moves were low-level suggestions, and some of them were not in the suggestions at all. There was no complete sequence that was without either low-level or not in the suggestions at all, which was not observed in first two games.

Analysis game 4

- WAGC round 6, Carlo Metta – Oscar Vazquez (live game)
  move 75 – mid level suggestion from 20k
  move 77 – low level suggestion from 2k
  move 119 – low level suggestion from
  2k to 30k, after that not in suggestions



Again, moves (from 30 to 135) in fighting and middle game varied a lot, and were similar to previous game.

Conclusion

Comparing two of the Metta's online games with two of his live games, we have found that there are significant differences in Metta's online games, compared to his live games.

In game vs Martin Ruzicka, not only most of the moves were similar to Leela's, but of 4 middle game moves, all 4 (one partially) were top Leela's suggestion – including one which was life and death mistake of Leela, resulting in strong attack on Metta's group.

In game vs Ben David, deviations histogram is again very similar to Leela's, with few and very small mistakes. Of 6 middle game tenukis, 5 are top Leela suggestions. Only one was low level suggestion, but it was sente combination, and after it Leela's suggestion was played.

In both internet games, middle game and fighting play sequences were almost completely made of top Leela's suggestions.
Both examined games are very similar to Leela's play.

In live game against Kim, we can see several mistakes in histogram. Some of the moves in the fighting sequences are low level not even in the list of suggestions, which was not the case in the internet games. Of the 3 middle game tenukis, one is not in the suggestions, second was low-mid level, and third one is not in the suggestions (although major suggestions were nearby).

In live game against Vasquez, again we can see lot of mistakes in histogram. Of three middle game tenukis, one was mid level suggestion, second was low level, and third was low level suggestion/not in suggestions.

In both live games, all middle game and fighting sequences had higher percentage of weak moves, some of which were not at all found as Leela's suggestions.
Both live games have similar quality, comparing to Leela's play.

Overall speaking, difference in Metta's play in those games is significant. His play in two internet games is so similar to Leela's, including it's mistake, that it is obvious that for the majority of important moves Metta used assistance from Leela. It is not just in sheer percentage – almost every important move was top Leela's choice which can be hardly explained on any other way.
In contrast, during live games, of 6 middle game moves, none was Leela's top choice, one was mid level suggestion, second low/mid level, and remaining 4 are either low level, or are not in suggestions at all. Also, in live games there was more differences to Leela play, and they were bigger. It is necessary to point out that largest mistake in all games, was actually also Leela's typical Life&Death mistake.
By observing level of play, not just statistics, and similartities, it is clear that in those two internet games help from Leela was used by Carlo Metta, which boosted his level of play significantly. His level of play in live games was much different than in two examined internet games.

Bojanic Milos
Belgrade, Serbia
June 12, 2018.