Knowledge Representation - Assignment 1 Report

Minh Ngo 10897402[1], Casper Thuis 10341943[2]

September 17, 2015

[1] University of Amsterdam
MSc. Artificial Intellegence
minh.ngole@student.uva.nl

[2] University of Amsterdam
MSc. Artificial Intellegence
casper.thuis@gmail.com

1 Introduction

SAT problems are NP-complete. They are not believed to be solved efficiently, but in spite of this fact SAT solvers using heuristics can solve many instances of SAT problem that occur in practice like Sudoku in the reasonable time. In the following project report the correlation between human difficulty and SAT solver's difficulty in the Sudoku task is investigated. The case investigation can lead to the better understanding of human reasoning by solving SAT problems or oppositely. SAT solver's can deal with Sudoku boards without any domain knowledge, with the rules of the game as an exception. Because of this lack of domain knowledge SAT solvers might consider different aspects of the Sudoku game in comparison to the human reasoning, and because of that cases where the Sudoku game is difficult for a human can have an opposite effect for SAT solvers or vice versa.

2 Hypothesis

Based on the "naive" simplified vision about the human difficulty, where a game becomes more difficult for human with the increase of the sparseness of the game board, the following assumption is made for the SAT solvers behaviour on different level of human difficulty in the Sudoku game: difficult (sparse) boards and easy (dense) boards are easy for SAT solvers, the difficulty of decision making reaches its peak with a middle level of human difficulty. This hypothesis can be intuitively argued by the fact that for the sparse boards there are more possible solutions that can be found, for the dense boards solutions can be quickly found by force moves.

3 Experimental setup

For experiments in this project a dataset of Sudoku boards has been created by parsing a website[1] that generates game boards. Sudoku boards, annotated by human difficulty ranking, is the main reason to prefer such approach for test cases generation. Next, 2000 board combinations for each of 9 level of difficulty [Table 1] have been sampled to form a dataset of 18000 unique Sudoku boards. Annotated information about game boards have been transformed later into the DIMACS[2] format, that is a common format for available state-of-art SAT solvers, by the script Sudoku-CNF-generator[3]. Finally, Sudoku boards in the DIMACS format have been used as an input of the SAT solver PicoSAT (Biere, 2008). PicoSAT is an extension of miniSAT (Een and Sorensson, 2003) and uses multiple heuristics to other techniques to speed up the process such as watched literals, occurence lists and restarts. In comparison to other SAT solvers, like GlueMiniSat (Nabeshima et al., 2014) or minisat2 (Een et al., 2010), PicoSAT indicates a wide variety of statistics that can simplify an analysis of experiment results.

[1] `http://www.menneske.no/sudoku/eng/index.html`
[2] `http://www.csc.kth.se/~jakobn/teaching/proofcplx11/minisat.php`
[3] `https://github.com/sergisiso/Sudoku-CNF-generator`

Difficulty	Integer notation	Techniques
Super easy	0	SiSo
Very easy	1	SC
Easy	2	SB
Medium	3	DS
Hard	4	DC
Harder	5	XW
Very hard	6	SF, JF, SB, NS, FI, FC
Super hard	7	SF, JF, SB, NS, FI, FC, SqB
Impossible	8	SF, JF, SB, NS, FI, FC , SqB, BB

Table 1: Difficulties and techniques

The following statistics for each experiment have been collected:

- **Visited count** is defined as the number of clauses visited during the Sudoku solving (Biere, 2008).
- **Fixed variables count** is defined as the number of variable with a fixed assignment for the solution to be satisfied[1].
- **Number of propagation** counts the process of updating the occurrence list of one literal (Biere, 2008).
- Run-Time

There are numerous difficulty measurements to rank or rate Sudoku (Pelánek, 2014). Some of these measurements contain the amount of given numbers, spread of the numbers or the techniques needed to solve the Sudoku. These methods are generally considered to be subjective since humans may differ in what they find difficult. Despite the subjective nature of raking or rating these Sudoku it is frequently used in newspapers or websites. The definition for difficulty of a Sudoku in this paper has a practical origin. The definition used in the data set is dependent on the techniques one needs to use to solve the puzzle. Table 1 indicates the difficulties and techniques accordingly. The techniques[2] are as follows: SiSo - Single Solution , SC - Single Cell, SB - Single Box, DS - Disjoint Subset, DC - Disjoint Chain, XW - X-Wing, SF - Swordfish, JF - Jellyfish, SqB - Squirmbag, NS - Nishio, BB - Bowman Bingo. These will not be explained because of the scope of the project report but are listed on the website[2].

4 Experimental results

Algorithm run-time can not be used for performance measurement because of the small size of the Sudoku problem. Therefore it was decided to use other measurements instead such as visited counts and number of fixed variables, that have been pointed out (Biere, 2008) that this is a good indication of the amount of work performed by the SAT solver. Mean and variance of visit counts for 2000 experiments of each human difficulty level have been computed [Fig 1]. The full distribution of values has been investigated in advance [Fig 2, 3] and a t-test has been preformed to determine whether the means are significantly different. The results of the t-test has a p-value lower than 5%. It shows that the distribution of visits count of the more difficult level is shifted from the distribution of the easier level.

[1] https://github.com/ContinuumIO/pycosat/issues/15
[2] http://www.menneske.no/sudoku/eng/reducingmethods.html

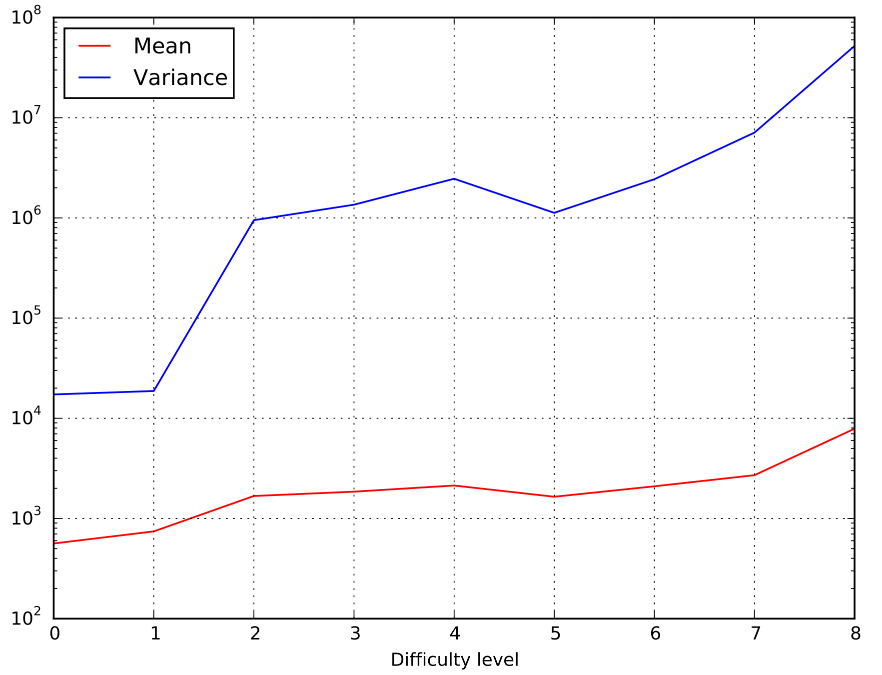

Figure 1: *Mean & Variance of visit counts. Measurements are in the logarithmic scale.*

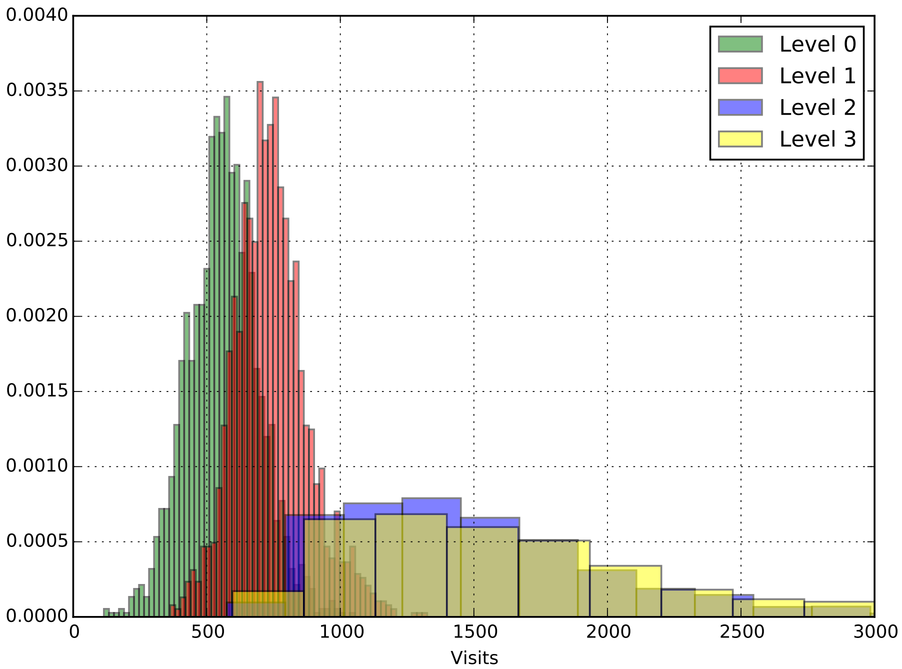

Figure 2: *Visits for the difficulty level 0 - 3*

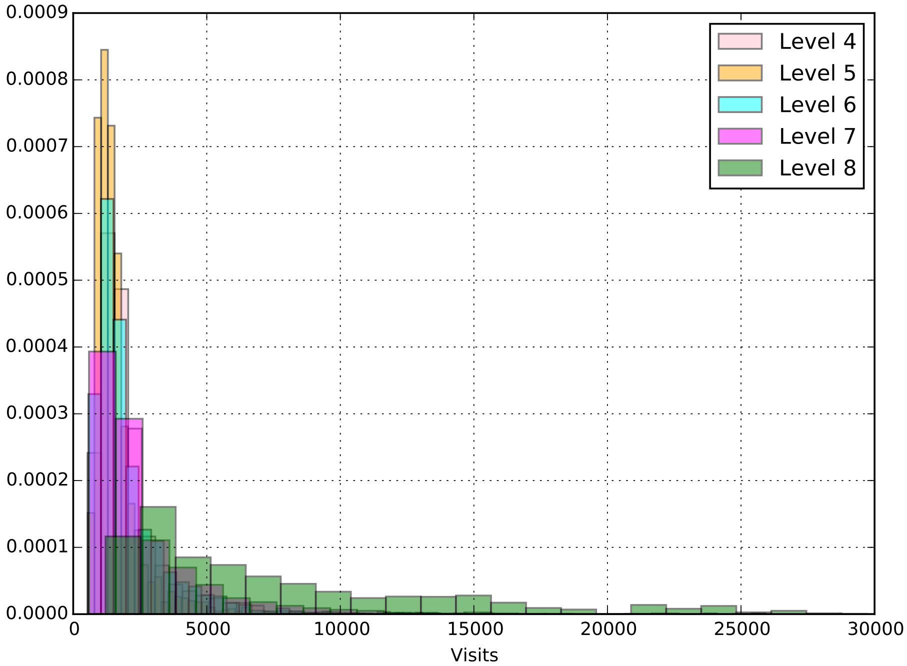

Figure 3: *Visits for the difficulty level 4 - 8*

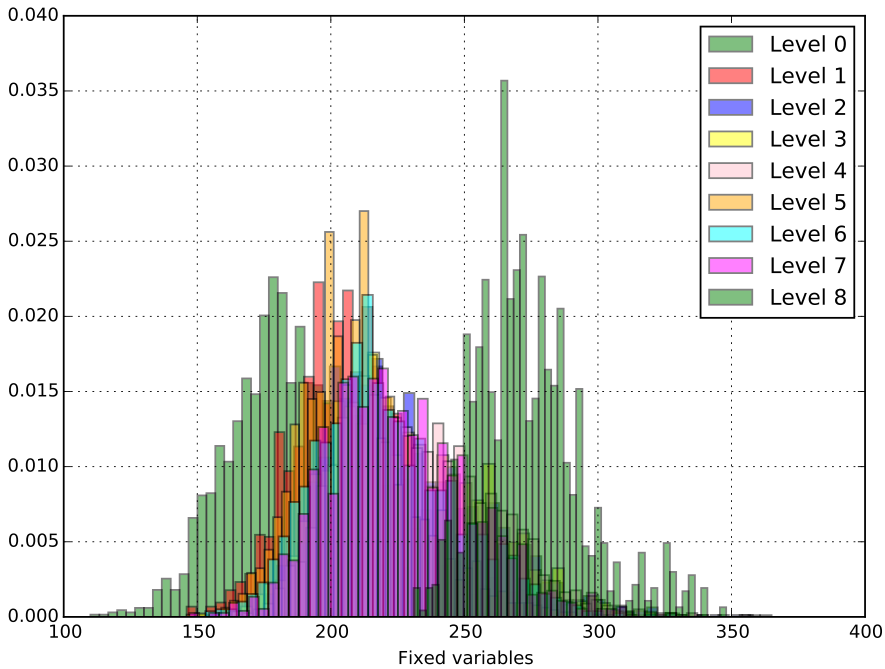

Figure 4: *Fixed variables*

5 Interpretation

The results in figure 1 show a clear trend that the amount of visits grow with the levels of difficulty. The trend might be seen as a logarithmic curve. This trend is also visible in the variance, with the increase variance per level of difficulty amount of visits in each levels differs more from the mean. To highlight the spread of the visits values distributions are shown in figure 2 and 3. The first levels have a small-spread distribution in comparison with the higher levels. Which might indicate that more higher levels might be harder to generate consistently. The figure shows that some levels such as, 0 to 3 and 4 to 8, have overlap. This means that some higher levels can be solved with less visits then lower levels, which indicates that the neighbouring levels are can be similar. This overlap might be explained because the random nature of generating the data. In general however the mean showed the trend that the performance decreases while the level increases.

The fixed variables is not a performance measure but can be interpreted as the amount of learning preformed in the solver. In figure 4 the fixed literals are shown. The increasing trend in fixed literals can probably be explained by the fact that more difficult levels might have a deeper search space and the solutions take longer to find. If the search takes longer the change of finding fixed literals increase or unit clauses increases.

6 Conclusion

In this project the investigation about relations between human difficulty and SAT solvers difficulty has been considered. The dataset of board games with annotated human difficulty has been collected to perform them later on the SAT solver PicoSAT. Based on the statistics provided by PicoSAT the insight has been made. Given obtained experiments results it can be concluded that the initial hypothesis, namely, difficult (sparse) boards and easy (dense) boards are easy for SAT solvers, the difficulty of decision making reaches its peak with a middle level of human difficulty, has been rejected.

There is a trend that indicates a correlation between human understanding about difficulty and SAT solver's performance. With an increase of a human difficulty of the problem a SAT solver difficulty also increases. That means that are might be similarities between human reasoning and SAT algorithms and heuristics. Indeed, many heuristics introduced in the SAT problem has come from the human intuition about decision making such as backtracking, force moves or restarts.

7 Future works

PicoSAT contains a plenty of heuristics for speeding up the decision making process. This might make the SAT solvers used in this project too efficient. The investigation based on the plain SAT solvers with the DavisPutnam algorithms without optimization (like cryptominisat[1]) can be considered in the future work to investigate heuristics contributions to the problem solving and their relations to the human reasoning. Researching the project hypothesis for less efficient SAT solvers might give different results.

As mentioned before human difficulty or just general difficulty of Sudoku are difficult to define. The definition used in this project is based on human techniques, however there are many other methods to rate or rank Sudoku. It might be that using other methods such as the Richter scale from Ercsey-Ravasz and Toroczkai (2012) or the method of Coelho (2007) to rank the Sudoku used here causes a different ranking. It would be worth investigating if these ranking methods have similar or different results.

[1] http://www.msoos.org/cryptominisat2/

References

A. Biere. Picosat essentials. *Journal on Satisfiability, Boolean Modeling and Computation*, 4:75–97, 2008. URL http://jsat.ewi.tudelft.nl/content/volume4/JSAT4_5_Biere.pdf.

J. S. Coelho. A scale to measure the difficulty of sudoku puzzles. *Revista de Ciências da Computação*, 2(2), 2007. URL http://lead.uab.pt/OJS/index.php/RCC/article/download/13/11.

N. Een and N. Sorensson. An extensible SAT-solver. 2003. URL http://lara.epfl.ch/w/_media/projects:minisat-anextensiblesatsolver.pdf.

N. Een, A. Mishchenko, and N. Amla. A single-instance incremental sat formulation of proof- and counterexample-based abstraction. 2010. URL http://arxiv.org/pdf/1008.2021v1.pdf.

M. Ercsey-Ravasz and Z. Toroczkai. The chaos within sudoku. *Scientific reports*, 2, 2012. URL http://arxiv.org/pdf/1208.0370v1.

H. Nabeshima, K. Iwanuma, and K. Inoue. Glueminisat 2.2.8. 2014. URL http://glueminisat.nabelab.org/home/download/glueminisat2.2.8.pdf?attredirects=0&d=1.

R. Pelánek. Difficulty rating of sudoku puzzles: An overview and evaluation. *arXiv preprint arXiv:1403.7373*, 2014.