Analysis of the Behavior of People Solving Sudoku Puzzles

Reijer Grimbergen
School of Computer Science, Tokyo University of Technology, Hachioji, Japan
E-mail: grimbergen@cs.teu.ac.jp

Akihiro Nakamura
Department of Informatics, Yamagata University, Yonezawa, Japan.

Abstract

In this paper we present a method for analyzing human problem solving in the pencil puzzle *Sudoku*. A number of subjects were asked to solve a Sudoku puzzle and write an explanation concerning the decisions they made to solve the puzzle. From these written protocols a number of rules were extracted. These rules were then implemented in a computer program that used these rules to suggest possible squares to fill for each step in the original Sudoku puzzle. The suggestions of the computer program were compared with the decisions of the human solvers using two criteria: *output ratio* and *cover ratio*. The results suggest that human solvers prefer to fill rows and columns over subgrids unless the subgrids have very few unfilled squares.

Keywords: Cognitive science, Human problem solving, Pencil puzzles, Sudoku.

1 Introduction

Our research aims at implementing Marvin Minsky's *Society of Mind* theory [2] in a game-playing environment. The goal is to have a full implementation of this many-faceted theory for a two-player game like shogi. However, to know more about the cognitive behavior of game players, it might be a good idea to first analyze the behavior of humans solving puzzles. Puzzles do not have the complicating factor of interaction with other players and therefore might tell us more about basic human problem solving behavior.

In this research, we have used the popular pencil puzzle *Sudoku* to analyze human problem solving behavior. Pencil puzzles have been the topic of research before [1], but as far as we know have not been used for cognitive modeling of human problem solving behavior.

We will now explain our method for analyzing human problem solving behavior in Sudoku.

2 Sudoku

Sudoku is part of the family of *Pencil Puzzles*, which include crosswords, mazes, word searches, logic puzzles and so on. It is a very popular puzzle around the world and most will be familiar with its rules. In Figure 1 an example of a Sudoku puzzle is given. The puzzle is consists of a 9×9 grid with nine 3×3 subgrids. The aim is to fill each row, each column and each 3×3 subgrid with the numbers from 1 to 9. No row, column or subgrid should have a number more than once.

1		8	2		4	6		3
		4	9		3	5		
5	3			6			2	4
4			6		7			8
8	9		3		5		4	7
	2			9			6	
7	4		5		6		1	9
		3		4		8		
6		1	7		9	4		2

Figure 1: A Sudoku puzzle.

3 Sudoku Solving Behavior

To analyze the behavior of humans solving a Sudoku puzzle, we performed the following steps:

1. A number of subjects were asked to solve a Sudoku puzzle and explain the reason for the decisions they made.

2. The output of the subjects was used to extract

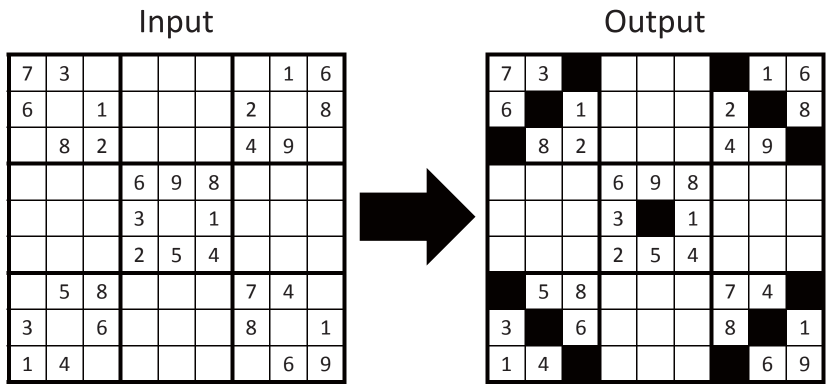

Figure 2: Example of the input and output of the computer program.

a number of rules that seemed to guide the problem-solving behavior.

3. These rules were implemented in a computer program.
4. The output of the computer program for the Sudoku puzzle was compared to the original output of the subjects.

We will now describe each of these steps in more detail.

3.1 Recording Human Behavior

Five subjects in their early twenties (students of our lab who had no knowledge about the objectives of the experiments), were asked to solve two Sudoku puzzles. For each step P (we will call this a Sudoku *position* from now on) in a Sudoku puzzle, a square S_P has to be selected on which to pencil in a number N_P ($0 < N_P < 10$). Each subject was asked to write down the reason for selecting S_P and N_P for each step P.

3.2 Rule Extraction

By analyzing the written protocols of the subjects, a number of rules about the problem solving behavior in Sudoku can be extracted. For example, a simple rule is “If there is only one empty square on a row or column, fill it with the missing number”. This rule is so specific that it doesn't say much about human problem solving methods, so we used rules that are more general. These rules will be explained below.

3.3 Computer Simulation

The rules were implemented in a simple computer program. The input of the program was Sudoku position P, and the output of the program were the set of squares $\{S_{P1}, \ldots, S_{Pn}\}$that were selected based on the rules (see Figure 2). The program is not suggesting actual numbers to pencil into the squares.

3.4 Simulation vs Human Behavior

To decide how well these rules explain human problem solving behavior in Sudoku puzzles, the output of the computer program was then compared to the original output of the subjects. This analysis was based on two criteria: *Output ratio* and *Cover ratio*.

3.4.1 Output Ratio

The output ratio for a single Sudoku position P is defined as follows.

$$O_P = \frac{C_P}{B_P} \times 100$$

Here C_P is the total number of squares given as candidates by the computer program and B_P is the total number of empty squares in the Sudoku puzzle position.

To get the output ratio for the whole puzzle, the average of the output ratio for each position is calculated as follows.

$$O = \frac{\sum_{P=1}^{n} C_P}{\sum_{P=1}^{n} B_P} \times 100$$

The output ratio is used as a measure of how many unnecessary candidates are produced by the rules. If a program produces all possible squares, this will clearly include the selection of the human subjects, but this cannot be called a good explanation of human behavior. The output ratio should be as small as possible. Ideally, the program should only suggest one candidate in each position, because this gives the same output behavior as a human solver selecting a single square for each Sudoku position.

3.4.2 Cover Ratio

The cover ratio is the number of positions where a square selected by a human solver was part of the set of candidate squares produced by the computer program. It is defined as follows.

$$C = \frac{R_c}{S} \times 100$$

Here R_c is the number of times the square selected by the human solver was among the candidates given by the computer implemented rules. S is the number of positions in the Sudoku puzzle from the initial position to the solution. This is the same as the number of empty squares in the initial Sudoku puzzle position.

The cover ratio should be as close to 100% as possible, because this means that all the squares selected by the human problem solver are suggested by the computer model.

4 Rule Set and Comparison

From the explanations of their choices when solving the first Sudoku puzzle (this is the puzzle in Figure 1), it seems that the subjects are using the following strategies:

- Give priority to rows and columns where there are three empty squares or less.
- Try to fill subgrids, rows and columns with few empty squares.
- Explore subgrids, rows and columns with only 2 or 3 empty squares. For example, when there is a row with 3 empty squares, for each empty square it is checked what the constraints are for the numbers based on the numbers that are already in the corresponding columns and subgrids.

These observations are transcribed into the following rules.

Rule 1 Generate all those squares for which the total number of empty squares on the corresponding subgrid, row or column is lower or equal than threshold T_1.

Rule 2 If the number of empty squares in a certain subgrid is lower or equal than threshold T_2, generate all empty squares in this subgrid.

Rule 3 If the number of empty squares for a row or column is lower or equal than threshold T_3, generate all empty squares on this row or column.

Note that the first rule is the most general one and that the other two rules are needed to explain a possible difference between subgrid detection and row/column detection. The thresholds T_1, T_2 and T_3 can be set to different values. From the explanations by the subjects it seems that 2 or 3 are the right values for these thresholds, but our experiments showed that this might not necessarily be the case.

4.1 Results for Sudoku Puzzle 1

We first compared the squares given of the computer model using the rules given in the previous section with the squares selected by the human subject for the first Sudoku puzzle. We compared the performance of the individual rules and also investigated combinations of different rules with difference threshold values in a hierarchy.

4.1.1 Individual Rule Results

First, we looked at the values for T_1, T_2 and T_3 that have the best output ratio and cover ratio when the rules are applied individually. We have given priority to cover ratio, which should be higher that 90%. The best result is therefore the threshold value with the lowest output ratio having a cover ratio higher than 90%. The results are given in Table 1.

Table 1: Best threshold values for puzzle 1.

Rule	*Threshold*	*Output ratio*	*Cover ratio*
1	9	54.6%	96.8%
2	4	54.0%	94.7%
3	3	52.8%	95.8%

From these results it can be concluded that when looking at individual rules, the third rule has the best combination of output ratio and cover ratio. Although the output ratio is quite high, this seems a first indication that human solvers prefer information about rows and columns over information about subgrids.

It is also interesting to note that the threshold for the first rule needs to be set to 9 to get the best result, which goes against the explanations given by the human solvers.

4.1.2 Ordered Rule Results

The output ratio can be improved by combining different rules with different threshold values into a hierarchy. When a certain rule in the hierarchy produces an empty square, no rules that are lower in the hierarchy are triggered. We tried many

7	3						1	6
6		1				2		8
	8	2				4	9	
			6	9	8			
			3		1			
			2	5	4			
	5	8				7	4	
3		6				8		1
1	4						6	9

Figure 3: The second Sudoku puzzle.

different hierarchies and the best result was the following ordering of rules:

1) Rule 1 with $T_1 = 2$
2) Rule 2 with $T_2 = 1$
3) Rule 3 with $T_3 = 3$

This combination of rules gave an output ratio of 40.0% and a cover ratio of 92.1%.

These results indicate that human solvers may first look for rows, columns or subgrids that have only one or two empty squares before trying rows, columns or subgrids with more empty squares.

4.2 Results for Sudoku Puzzle 2

Next, we made the same comparison for a different puzzle, which is given in Figure 3. This puzzle is a special case, because the initial numbers are concentrated in the subgrids. Therefore, this puzzle seemed a good test for the generality of the rules we used.

4.2.1 Individual Rule Results

The individual rule results for the second puzzle are given in Table 2.

Table 2: Best threshold values for puzzle 2.

Rule	*Threshold*	*Output ratio*	*Cover ratio*
1	12	66.8%	97.1%
2	8	46.7%	91.8%
3	5	99.9%	99.9%

From these results it can be concluded that even higher threshold values are necessary to get to the desired cover ratio. This goes against the observation that human problem solvers will pencil in the highest constrained squares first. Furthermore, it seems that rule 3 is not very effective for this puzzle, because it will have to return almost all squares to get the desired cover ratio.

4.2.2 Ordered Rule Results

Because the individual rule results showed that rule 3 was ineffective for this second puzzle, we only investigated rule ordering for rule 1 and rule 2. For this puzzle the best ordering was as follows:

1) Rule 2 with $T_1 = 1$
2) Rule 1 with $T_2 = 2$

Even though the threshold results are the same as for the first puzzle, the order of the rules is reversed. The reason for this is that the second puzzle has a lot of empty squares on rows and columns, so to solve the puzzle it is important to give priority to filling subgrids. However, this is only done if the number of empty squares in the subgrid is one, i.e. there is only one possible candidate left to pencil in.

5 Conclusions

In this research we have used the Sudoku pencil puzzle to analyze the behavior of human problem solvers. From the feedback received from five human subjects we created a number of rules that suggest candidate squares for entering numbers. We used output ratio and cover ratio to evaluate how these rules match the behavior of the human subjects, both for individual rules and for ordered rules. The results suggest that human Sudoku solvers prefer to look at rows and columns instead of subgrids unless the subgrid has very few empty squares.

However, the results of a different puzzle showed that the rules that we proposed (especially rule 3) may not be generally effective. Therefore, as future work we want to look at more complex sets of rules and include search that is used by human solvers to decide between multiple candidate numbers for the same square. Also, our computer simulation is only able to give candidate squares and this should be extended to give pairs of squares and actual numbers that should be penciled in.

References

[1] K. Korekawa, H. Shirai, K. Shibahara, and Y. Kotani. The Automatic Extraction of Resolutional Rules from Pencilpuzzle. In *The 13th Game Programming Workshop in Japan 2008*, pages 112–115, 2008. (In Japanese).

[2] M. Minsky. *The Society of Mind*. New York: Simon and Schuster, 1988. ISBN 0-671-65713-5.