Difficulty Rating of Sudoku Puzzles by a Computational Model

Radek Pelánek

FLAIRS 2011
Sudoku

"Easy? Medium? Hard?"
Introduction

Problem  Difficulty rating of Sudoku puzzles
Approach  Computational model of human problem solving
Goal  Primary simplicity, insight
      Secondary performance
Contributions

- abstract model with good performance
- new metric: dependency of steps
- **evaluation** on large datasets
Data

- fed-sudoku.eu:
  - 1088 puzzles, 100 solvers for each
  - time for each solver and puzzle
- sudoku.org.uk:
  - 731 puzzles, 1000 solvers for each
  - mean time for each puzzle
- czech-sudoku.com:
  - 15 puzzles
  - detailed log for each game
what does not work:

- number of givens
- metrics based on backtracking
- many other “plausible ideas”
Naive Metric: Number of Givens
A Computational Model

- general approach:
  - repeatedly determine a value of some cell using a “logic technique”
  - prefer simpler techniques
- difficulty rating:
  - according to difficulty of used techniques
### Logic Techniques

#### Hidden Single

<table>
<thead>
<tr>
<th>1</th>
<th>2,4</th>
<th>2,4</th>
<th>2,3,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,3</td>
<td>2,4</td>
<td>1</td>
<td>2,3,4</td>
</tr>
<tr>
<td>4</td>
<td>1,2</td>
<td>3</td>
<td>1,2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2,4</td>
<td>1,2,4</td>
</tr>
</tbody>
</table>

#### Naked Single
Example of a Model Run
A Computational Model

Common approach  Many logic techniques, many parameters

Our approach  Few techniques, few parameters
  • easier evaluation
  • better portability to other problems
  • better insight
Simple Sudoku Solver Model

- only two basic techniques:
  - hidden single
  - naked single
- techniques not applicable $\Rightarrow$ search
  - cell with smallest “refutation score” = how many steps are needed to refute incorrect values
  - approximation of more complex logic techniques
Model vs Humans

**Humans**

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

**Model**

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

---

**Model vs Humans**

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

---

**Model**

```
6 1 2 8
9 4 1 7
8 5 6 3
4 5 3 1
5 8 2 6
```
Model vs Humans
Difficulty Metrics

- basic approach:
  - combining difficulty of individual logic techniques
  - given by parameters / refutation score

- novel aspect:
  - evaluating dependency among steps
Example of a Model Run
Evaluating Step Dependency

![Graph showing the number of possibilities for the next step across different step levels (easy, medium, hard) over a range of steps. The x-axis represents the step number, and the y-axis represents the number of possibilities. The graph indicates trends for each level, with 'easy' having the least variability, 'medium' showing moderate variability, and 'hard' displaying the highest variability.]
### Results: Correlation Coefficients

<table>
<thead>
<tr>
<th>metric</th>
<th>fed-sudoku</th>
<th>sudoku.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of givens</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>Serate</td>
<td>0.70</td>
<td>0.86</td>
</tr>
<tr>
<td>Serate LM</td>
<td>0.78</td>
<td>0.86</td>
</tr>
<tr>
<td>Fowler’s</td>
<td>0.68</td>
<td>0.87</td>
</tr>
<tr>
<td>Refutation sum</td>
<td>0.68</td>
<td>0.83</td>
</tr>
<tr>
<td>Dependency</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>Combined (RD)</td>
<td>0.74</td>
<td>0.88</td>
</tr>
<tr>
<td>Combined (SFRD)</td>
<td>0.84</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Combined Metric: sudoku.org.uk
Summary

Sudoku difficulty rating, evaluation over extensive data

- many plausible approaches do not work:
  - “static” metrics
  - backtracking
- computational modeling of human behaviour works
  - even simple model is sufficient
  - combined metric – very good correlation