Sudoku Players' Forums

Rating rules / Puzzles. Ordering the rules

Goto page Previous 1, 2, 3 ... , 12, **13**, 14 Next

newtopic postreply **Sudoku Players' Forums Forum Index -> Advanced solving techniques**

View previous topic :: View next topic

Author | **Message**

coloin

Joined: 05 May 2005
Posts: 997
Location: Oxford

Posted: Mon Jun 29, 2009 5:16 am Post subject: quote

Yes, the solution grid from a puzzle will have no bearing on how hard the puzzle is.....although I think the sheer number of puzzles in a single grid means that a selection of puzzles from a single grid wont be much different fom a random selection from many different grids..

Is mike's biased puzzle collection [from the same solution grid] not biased by the generator - already shown to be biased in terms of its output ?

suex9s.exe generates [seed] puzzles from the same grid greater than size n [16]

Here is the clue distribution for ALL the minimal puzzles in the 40 clue subgrid which I chose.

Code:

```
.4......97....2.585...6..1...5....6....63...7.....192...21.....9.8.....3.679.....
25puzzle
24..18..97..3.2.585.3.6..1...5.94.6.1..63...7..4..1925..2187...9.8..6..3.679...82
25plus15puzz
```

Code:

```
c:\Suxx>clusta ran2040solsmax.txt
lines:713818 average clues:25.246793
   21  45
   22  1258
   23  22634
   24  136937
   25  272826
   26  206311
   27  65658
   28  7900
   29  248
   30  1
```

Possibly it is not too far from the real distribution from a complete 81-clue grid.

Compare with a mean puzzle siz of 24.39 for the first 16000 puzzles found [using suex9s].

Code:

```
puzz:16000 average clues:24.344562
puzz:15266 average clues:24.392637  [dups removed]
```

I agree that this has marginal relevance to this thread, and it would be interesting to fathom why larger clue puzzle tend to be harder.

Perhaps the best way to have a reference file for puzzle would be to have equal numbers of puzzles with each number of clues. Randomly generated 28+ clue puzzles have a significantly higher rating. Whether from the same or different solution grids.
C

Last edited by coloin on Mon Jun 29, 2009 6:30 am; edited 1 time in total

Back to top profile pm

Red Ed

Posted: Mon Jun 29, 2009 6:13 am Post subject: quote

Joined: 06 Jun 2005
Posts: 518

> **coloin wrote:**
> I cant see how to get a real sample of non-biased puzzles !

That's easy: it's just small and/or minimal puzzles that are hard to generate in an unbiased manner.

To sample at random from the space of all puzzles, simply repeat: generate a solution grid; delete each clue (independently) with probability .5; output the subgrid if it has a unique solution.

Back to top profile pm

coloin

Posted: Mon Jun 29, 2009 6:41 am Post subject: quote

Joined: 05 May 2005
Posts: 997
Location: Oxford

Yes.....i thought of that.....but it would be very slow...but I suppose possible.

The chances of getting a minimal 25 clue puzzle this way are very small.....

I think 50% of 41 clue templates of a solution grid have >1 sol. [the clues miss an unavoidable set]

Also the "real" clue frequency distribution wouldnt be produced this way either. [larger, harder minimal puzzles would be produced preferentially !]

Back to top profile pm

Red Ed

Posted: Mon Jun 29, 2009 8:00 am Post subject: quote

Joined: 06 Jun 2005
Posts: 518

> **coloin wrote:**
> Yes.....i thought of that.....but it would be very slow...but I suppose possible.

Not slow at all. I can't remember the percentage, but it's not far from half of all random subgrids that have a unique solution.

> **Quote:**
> The chances of getting a minimal 25 clue puzzle this way are very small.....

Yes. Waiting for 25-clue minimals by this method would be tiresome.

> **Quote:**
> I think 50% of 41 clue templates of a solution grid have >1 sol. [the clues miss an unavoidable set]

I'll take your word for it: I can't find the old thread now.

> **Quote:**
> Also the "real" clue frequency distribution wouldnt be produced this way either. [larger, harder minimal puzzles would be produced preferentially !]

No. The process I described samples uniformly flat randomly from the space of all proper puzzles.

Back to top profile pm

coloin

Posted: Mon Jun 29, 2009 9:30 am Post subject: quote

Joined: 05 May 2005
Posts: 997
Location: Oxford

Yes, all true, you would have almost all non-minimal proper puzzles.

But removing more unecessary clues [by suex9s] doesnt give "random" puzzles.

Would it be possible to achieve that ?

I suspect the the mean [minimal] puzzle size is more than 25.2 clues !

Back to top profile pm

Red Ed

Posted: Mon Jun 29, 2009 9:58 am Post subject: quote

Joined: 06 Jun 2005
Posts: 518

Are you asking if removal of unnecessary clues from uniform-random sub-puzzles gives uniform-random minimal puzzles? Almost certainly not; but quantifying the bias is beyond me. ☹

Back to top profile pm

denis_berthier

Posted: Thu Jul 02, 2009 9:46 pm Post subject: quote

Joined: 19 Jun 2007
Posts: 631
Location: Paris, France

> **denis_berthier wrote:**
>
> it should be easy to check if the U4-test is meaningful wrt to the complexity of puzzles (measured as their NRCZT or SER levels). We just have to compute the correlation coefficient between U4 and NRCZT (or SER). 10,000 puzzles should be enough.
> .....
> the correlation coefficients between the NRCZT or SER ratings of minimal puzzles and the numbers of instances of various patterns (Red Ed's patterns) in the corresponding solution grids are almost null.
> Conclusion: tests for occurrences of predefined patterns in complete grids are not relevant to the complexity of puzzles. This is understandable as predefined patterns in complete grids are washed out by the elimination phase of the puzzle generators.

There remained the minor point of checking whether this washing out is done early in the elimination phase of the generator. More precisely, is there any significant correlation between:
- the number of clues of the minimal puzzles obtained at the end of the elimination phase of suexg
- and the numbers of instances of various patterns in the complete grids used to start the elimination phase?
Answer: no. More precisely, for the three patterns from Red Ed's list already considered (using the first 100,000 puzzles in the sudogen0 collection), the correlation cefficients are:
#clues vs P-000907100000000000710000090-count = 0.031
#clues vs P-009000010000090008000081000-count = 0.024
#clues vs P-000000476000640000070000000-count = 0.000

Here is how I understand this result intuitively: a complete grid has 81 clues and almost all the minimal puzzles obtained by suexg (or all the other series of puzzles that have been proposed here) have between 20 and 30 clues. That is, between 51 and 61 clues are eliminated by the elimination phase. When 51 have been eliminated, almost all of the structure corresponding to a given pattern has been destroyed and eliminating 10 more clues can't destroy more.

Last edited by denis_berthier on Fri Jul 03, 2009 10:04 pm; edited 2 times in total

Back to top profile pm www

coloin

Posted: Fri Jul 03, 2009 2:08 am Post subject: quote

Joined: 05 May 2005
Posts: 997
Location: Oxford

Just a small contribution clarifying this

I understand that you are searching the grids/puzzles - and observing the disappearrance of these patterns.

except
they are not unavoidable sets
they have 6 clues

ie you cant call them U4 [only ONE type of U4 anyway]

If they were an unavoidable set.....of course you will always have a clue within the pattern, and you will never see a valid puzzle without a clue in the unavoidable set.

Perhaps the best way to get a random selection of puzzles is to just stick to 25 clue puzzles !

C

Back to top profile pm

denis_berthier

Posted: Fri Jul 03, 2009 2:20 am Post subject: quote

Joined: 19 Jun 2007
Posts: 631
Location: Paris, France

Coloin,
You must read the discussion in the "unbiased grid generation" thread. What I consider here is of course not the 3 specific patterns 000907100000000000710000090, 009000010000090008000081000 and 000000476000640000070000000, but all their isomorphs in all the complete grids.
These 3 patterns are taken from Red Ed's list of 3322. I've no doubt that the correlation results would be the same for all the 3322 patterns.

Last edited by denis_berthier on Fri Jul 03, 2009 8:49 am; edited 1 time in total

Back to top profile pm www

m_b_metcalf

Posted: Fri Jul 03, 2009 2:24 am Post subject: quote

Joined: 15 May 2006
Posts: 2144
Location: Berlin

coloin wrote:

> Perhaps the best way to get a random selection of puzzles is to just stick to 25 clue puzzles !

Isn't the best way to start with 17 clues with random positions and values (and constraints compatible) and then build upwards by adding one further clue at a time until a unique solution is reached, and then to remove any redundant clue(s)? Might take forever.

Regards,

Mike Metcalf

Back to top profile pm

denis_berthier Posted: Fri Jul 03, 2009 3:06 am Post subject: quote

Joined: 19 Jun 2007
Posts: 631
Location: Paris, France

m_b_metcalf wrote:

> **coloin wrote:**
>
> > Perhaps the best way to get a random selection of puzzles is to just stick to 25 clue puzzles !
>
> Isn't the best way to start with 17 clues with random positions and values (and constraints compatible) and then build upwards by adding one further clue at a time until a unique solution is reached, and then to remove any redundant clue(s)? Might take forever.

I think sticking to 25 (or to any predefined number of) clues is one of the worst ways of getting a random sample of puzzles: it is an arbitrary and artificial constraint.

Starting with 17 random clues and adding random clues one by one until a unique solution is reached may be an interesting way of modifying the first part of the suexg procedure.
I don't think it'd take much more time than suexg: maybe more time in the generation phase, but less time in the deletion phase.
Mike, can't you do something like this with your generator? If you can produce such a sample, I'd like to test it.

Back to top profile pm www

m_b_metcalf Posted: Fri Jul 03, 2009 3:12 am Post subject: quote

Joined: 15 May 2006
Posts: 2144
Location: Berlin

denis_berthier wrote:

> **Mike**, can't you do something like this with your generator? If you can produce such a sample, I'd like to test it.

I have a different old generator which I think I can more easily adapt to do this. But next week.

Regards,

Mike Metcalf

Back to top profile pm

Red Ed Posted: Fri Jul 03, 2009 6:35 am Post subject: quote

Joined: 06 Jun 2005
Posts: 518

coloin wrote:

> they are not unavoidable sets
> they have 6 clues
>
> ie you cant call them U4 [only ONE type of U4 anyway]

Just for the record, let back you up there, Coloin.

Specifically:

- None of those three patterns is an unavoidable set.
- Most of the 3322 patterns in my catalogue are *not* unavoidables.
- U4 refers to the essentially unique 4-clue unavoidable (which is part of the catalogue, as it happens).

Back to top profile pm

denis_berthier Posted: Fri Jul 03, 2009 7:05 am Post subject: quote

Joined: 19 Jun 2007
Posts: 631
Location: Paris, France

Red Ed,

OK, not unavoidables. So what are they? How would you name these 3322 patterns?

Back to top profile pm www

Red Ed Posted: Fri Jul 03, 2009 7:07 am Post subject: quote

Joined: 06 Jun 2005
Posts: 518

"Patterns".

Back to top profile pm

Display posts from previous: All Posts Oldest First Go

newtopic postreply **Sudoku Players' Forums Forum Index -> Advanced solving techniques**

All times are GMT - 8 Hours
Goto page Previous 1, 2, 3 ... , 12, **13**, 14 Next

Page 13 of 14

Jump to: Advanced solving techniques Go

You **cannot** post new topics in this forum
You **cannot** reply to topics in this forum
You **cannot** edit your posts in this forum
You **cannot** delete your posts in this forum
You **cannot** vote in polls in this forum