Sudoku Players' Forums

FAQ Search Memberlist Usergroups Register
Profile Log in to check your private messages Log in

THE REAL DISTRIBUTION OF MINIMAL PUZZLES

Goto page Previous 1, 2, 3 ... , 15, 16, 17 Next

newtopic postreply **Sudoku Players' Forums Forum Index -> General/puzzle**

View previous topic :: View next topic

Author **Message**

David P Bird

Posted: Sat Sep 19, 2009 2:39 am Post subject: quote

Joined: 16 Sep 2008
Posts: 136
Location: Middle England

As a pillow problem last night I considered practical means of speeding up the top down puzzle generating algorithm.

In the minimum clue set we know that 8 digits must appear at least once. Therefore at the start we could randomly select 8 cells to be preserved from deletion in the main algorithm reducing the first deletion choice down from 81 to 73 cells. These are the core cells which we assume must be present in the minimum clue set.

To assist in checking for redundant clues: the 2-digit unavoidable sets are easily identified and the cells each one covers can be listed.
We now have a tagging system for the cells in these lists
'S' indicates the cell is selected to be in the minimum set
'P' indicates the cell is present in the current clue set
'R' indicates the cell is required in the minimum set
'C' indicates the cell is already covered by a 'S' or 'R' cell

The 8 cells we originally selected for preservation are marked 'Selected' and the other cells in the sets containing them are tagged 'Covered'
All the other cells are then tagged as 'Present' in each set
At each new step which deletes cell X:
The tag for X is cleared in each of the 8 sets it belongs to.
If this reduces a set so that it only has a single tagged cell remaing that is shown as 'Present', that tag is changed to 'Required'
In the other sets that contain a new 'Required' cell, 'Present' tags are changed to 'Covered' for the remaining cells.

Now, if we weren't trying to operate a random process, as a cell is tagged as 'Required' it could be added to the core set, (and whenever all 8 instances of a cell in the unavoidable sets are tagged as 'Covered' it can only possibly be required to cover a multi-digit unavoidable set).

As it is, we can only use these tags to stop the grid generator and start again if at any stage a 'Required' cell has been randomly selected for deletion.

Although we can therefore abort from a mass of futile runs possibly much quicker by using these tags, they don't cover all eventualities and full redundancy checks would still be needed once we got down to 33 clues or so. If these show that one of the original core clues was in fact redundant, then the whole run would have to be aborted.

The question I find hard to answer is, would the pre-selection of the 8 starting cells help or hinder the overall rate at which successful runs would accumulate?

Last edited by David P Bird on Sat Sep 19, 2009 2:59 am; edited 1 time in total

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 2:55 am Post subject: quote

David P Bird wrote:

> In the minimum clue set we know that 8 digits must appear at least once. Therefore at the start we could randomly select 8 cells to be preserved from deletion in the main algorithm

If the 8 cells are selected so that each contains a different digit then you'll introduce bias.

Quote:

> The question I find hard to answer is, would the pre-selection of the 8 starting cells help or hinder the overall rate at which successful runs would accumulate?

Assuming that the 8 cells are selected flat randomly (allowing repeated digits), it should make little difference. Checking 2-digit unavoidable sets is so much quicker than running the solver than I think starting at 73 rather than 81 digits will have negligible effect.

But in general I agree that unavoidable-checking may give speed-ups.

Back to top profile pm

coloin

Joined: 05 May 2005
Posts: 1072
Location: Devon UK

Posted: Sat Sep 19, 2009 3:04 am Post subject: quote

I understand....... except if the solution grid doesnt really matter why cant we fix the grid ? [as long as its unbiased of course]

Then we can use suitable non-bias inducing optimizations.

Back to top profile pm

David P Bird

Joined: 16 Sep 2008
Posts: 136

Posted: Sat Sep 19, 2009 3:14 am Post subject: quote

Red Ed You've now got me onto bias again which I think I'll have to accept I'll never fully understand! However given an algorithm such as:

Randomly select a digit from 1 to 9 that can be absent from the minimum

Location: Middle England

> Randomly select a digit from 1 to 9 that can be absent from the minimum set.
> For each of the other digits, randomly pick a column in which the instance of that digit is preserved.

Couldn't the bias that it introduced be accounted for?

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 3:17 am Post subject: quote

coloin wrote:

> I understand....... except if the solution grid doesnt really matter

The solution grid does matter. We've always known this and it's even a point on which Denis and I agree! 😃

EDIT: you also need to consider that we want the output puzzles to be not only unbiased (drawn flat randomly from the population of all puzzles) but also uncorrelated (with each other). If you stick with a single grid then the latter property is obviously lost.

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 3:23 am Post subject: quote

David P Bird wrote:

> Couldn't the bias that it introduced be accounted for?

I know of no practical way to do so. Equally, though, I don't know how bad the bias would be -- perhaps not very.

Back to top profile pm

coloin

Joined: 05 May 2005
Posts: 1072
Location: Devon UK

Posted: Sat Sep 19, 2009 4:04 am Post subject: quote

Ah......but denis is saying that his "non-biased" puzzles from which we extrapolate the clue counts are probably/possibly not from unbiased grids. [we dont know yet]

I was actually meaning to suggest getting only ONE [unbiased] puzzle from each of your selected unbiased grids....hopefully getting it quicker than before.

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 4:20 am Post subject: quote

Coloin, there are two different aspects to solution grid bias that are not always well distinguished:

- Bias in the solution grid source being fed into the modified top-down generator. Mine's unbiased; Denis' is biased. A statistically significant effect percolates down to the number-of-clues distribution **but** the big surprise (apparently only to me) is that the effect is to almost all practical intents & purposes so small as to be negligible.

- Bias in the solution grids corresponding to output from the modified top-down generator. Those solution grids are just a subset of the ones mentioned in the first bullet (namely: the ones that happened to yield a minimal puzzle). They will be biased in the sense that those most likely to produce minimal puzzles are those most likely to appear. This is all well and good, nothing to worry about.

None of this is new; but hopefully I've cleared up some of the ambiguity for you.

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 804
Location: Paris, France

Posted: Sat Sep 19, 2009 6:30 am Post subject: quote

coloin wrote:

> Ah......but denis is saying that his "non-biased" puzzles from which we extrapolate the clue counts are probably/possibly not from unbiased grids. [we dont know yet]

You shouldn't confuse 2 things:
1) the source of the complete grids used to generate puzzles; for these, I said that a small bias in the complete grids is very likely to be washed away by the deletion phase (and I justified it in a previous post); I never said that we could use any biased sequence of complete grids.

2) the complete grids obtained as a result of solving the (not unbiased but controlled-bias) puzzles. If the puzzles are controlled-biased, there will be some bias in these grids. Exactly which kind of bias, I don't know. What's certain is that you can't use its existence to justify starting with biased grids in the generation phase.

[Edit] I had answered before reading Red Ed's answer. In essence, we say the same thing.

Back to top profile pm www

David P Bird

Joined: 16 Sep 2008
Posts: 136
Location: Middle England

Posted: Sat Sep 19, 2009 7:31 am Post subject: quote

OK so my first scheme won't really help much in the early stages and risks introducing yet another slight element of bias. So we can scrap pre-selecting any cells to be in the minimum set. So here is a re-worked version which now becomes Scheme II:

The cells covered by each 2-digit unavoidable set are listed in tables set up with accompanying tag fields.
Tag 'P' = the cell is 'Present' in the current clue set
Tag 'R' = the cell is 'Required' in the minimum set
Tag 'C' = the cell is already 'Covered' by a 'Required' cell
Tag 'N' = the cell is definitely 'Not Required' in the minimum set.

At the start all cells in are tagged as 'Present' in every list.
At each new step which deletes cell X:

The tag for X is cleared in each of the 8 sets it belongs to.
If this reduces a set so that it only has a single 'Present' tag remaining, that cell's 8 tags are changed to 'Required'.
In the other sets that contain the new 'Required' cell, any other 'Present' tags are changed to 'Covered'

If any cell elimination in the top down generator removes a cell tagged 'Required' the run can be aborted.

There is also an option to extend this scheme to include 6 cell 3-digit unavoidable sets found by simple pattern recognition checks too. Covering any bigger multi-digit sets would probably cost more time than it would save as the probability that all the cells of a 8+ cell set being eliminated without being detected by this scheme become rather small.

Now to cover the 'Not Required' tag:
When we are down to 33 cells or so or when there are no cells left which are simply shown as 'Present' (ie they are all either 'Required' or 'Covered') we are approaching a potential valid minimum set. Using only the cells marked 'Required' we can solve the puzzle as far as we can to check two things:
1) Are there any disjoint multi-digit unavoidable set(s) have escaped being covered so far? If so, we can add them to the set listing scheme marking their surviving cells either as 'Required' when there is just one, or 'Present' when there are more. As soon as this check shows that all uncovered sets are disjoint and so have already been included, it can be discontinued.
2) Are any of the cells tagged only as "Covered' already known to be redundant? If so they can be re-tagged 'Not Required'

We can now continue randomly excluding cells as before and if they are already tagged 'Not Required', no further checks are necessary. Finally if we reach the state where all the 'Not Required' cells have been eliminated before any of the 'Required' cells, we can claim our prize from the stall holder.

Clearly there is running time cost to implement this scheme, but it looks well spent in comparison to time wasted chasing potential clue sets far too long once it is clear they have no chance of success.

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 7:44 am Post subject: quote

David, your idea seems similar in spirit to my "alternative implementation improvement in step 3" in that it's (a) probably an improvement 😃 (b) at the cost of endless coding pain 🙁

I think it's rather nice that attacks in this vein exist using clue addition (maintaining solver state) *and* using clue deletion (tracking coverage of unavoidables). I don't fancy trying to code-up either method, though! Where's eleven when we need him ...?

If I program-up anything, it'll be the optimal path prober.

Back to top profile pm

eleven

Joined: 10 Feb 2008
Posts: 472

Posted: Sat Sep 19, 2009 8:06 am Post subject: quote

> **Red Ed wrote:**
>
> I don't fancy trying to code-up either method, though! Where's eleven when we need him ...?

🙂 Oh, i am not the one, who invests much time in sudoku coding. All i did here, was to adopt dukusos program to vary generation programs and find some special puzzles.

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 804
Location: Paris, France

Posted: Sat Sep 19, 2009 8:57 am Post subject: quote

Do we conclude that no optimisation work is underway?

Back to top profile pm www

Red Ed

Joined: 06 Jun 2005
Posts: 708

Posted: Sat Sep 19, 2009 1:00 pm Post subject: quote

It's taken time to get the path-probing strategy right. For best performance, the strategy needs to optimise the average time, not the average number of calls. Since calls to solve() are more expensive for lower numbers of clues, there is unfortunately an adverse effect on the total speed-up available. Hard to say without applying the strategy for real, but I should think we can get the code to run 4-5 times as fast when previously I was hopeful of 6x.

Applying the strategy for real is currently less attractive than the other options for this weekend.

Back to top profile pm

David P Bird

Joined: 16 Sep 2008
Posts: 136
Location: Middle England

Posted: Sun Sep 20, 2009 12:41 am Post subject: quote

Red Ed, Firstly I agree the pre-selecting 8 cells to be in the minimum set at the start is inferior as it carries a further penalty towards the end. This is because we have to keep checking if any of them are redundant.

Secondly in Scheme II it is possible to reach the stage where all cells are tagged 'Required' or 'Not Required' and we actually know the target minimum set we hope to find by chance. However, as we know the numbers of each, we can also accurately calculate what the chances are for actually hitting that target.

Would there be a way to use that information to add a result to the accumulated list without invalidating the clue distribution that's produced? If that is possible then the time savings would be even greater!

Now some seat of the pants stuff - say we are down to 30 remaining cells. If the

'Required' to 'Not Required' split is 20:10 the chances of reaching the target minimum set is considerably better than if the split is 27:3. Does this go any way to explaining why the generators favour the smaller minimum sets?

Back to top profile pm

Display posts from previous: All Posts ▲▼ Oldest First ▲▼ Go

newtopic postreply **Sudoku Players' Forums Forum Index -> General/puzzle**

All times are GMT - 8 Hours
Goto page Previous 1, 2, 3 ... , 15, 16, 17 Next

Page 16 of 17

Jump to: General/puzzle ▲▼ Go