Sudoku Players' Forums

Rating rules / Puzzles. Ordering the rules

Goto page Previous 1, 2, 3 ... 11, 12, 13, 14 Next

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

View previous topic :: View next topic

Author	Message

denis_berthier

Posted: Sat Jun 27, 2009 6:00 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Mike,

Here are the results concerning your second series of puzzles (limited to the first 50,000):

1) Clues
mean number of clues = 24.40
standard deviation = 1.13
very close to the first series and to sudogen0_1M
This seems to be a stable value.

Code:

```
#Clues #Puzzles  E(SER) s(SER)
20       3        3.77   2.47 (E not meaningful)
21       138      3.25   2.01
22       1622     2.91   1.91
23       8619     3.03   2.02
24       16814    3.12   2.10
25       14953    3.23   2.18
26       6199     3.40   2.29
27       1458     3.63   2.37
28       182      3.90   2.42
29       11       3.19   1.98 (E not meaningful)
30       1        7.6   0.0 (E not meaningful)
all      50k      3.18   2.14
```

2) SER
mean SER = 3.18 (standard deviation = 2.14)
Much higher thant the first series
But still much lower than Sudogen0 : 3.77 (standard deviation = 2.42)
Only one puzzle at SER 9.1, no puzzle above.

The puzzles in the second series are globally harder than those in the first series, but still easier than those in sudogen0.

3) Correlation coefficient #clues vs SER = 0.06 : uncorrelated.
Moreover, no visible dependency between #clues and SER

Regards

Back to top profile pm www

PIsaacson

Posted: Sun Jun 28, 2009 3:44 am Post subject: quote

Denis,

Joined: 02 Jul 2008
Posts: 183
Location: Campbell, CA

I am in the process of a complete re-write of my nrczt+groups engine and testing the effectiveness of the new theory. Here's a comparison using my new ALS engine combined with the new nrczt engine using the 23 SER 9.2+ interesting puzzles for comparison:

Code:

```
Puzzle    ------------------------------------------------------------------------
--------    SER nrczt+ Clues
707
 4..3..6.8..5..1.3..16.4...7.7.....8......2....51..637...2.9.8.....1.........4..9
  9.2    5    25
2035
...7...637....4.9..6.95....8.....7...5.6.....9...2..16..3..8..224.....7.......9..
  9.2    4    24
88857
 3.........41......69578.3..51.64.78.4....3.5...7......9....6..2...9..1...5.....6.
  9.2    4    26
117617
91.......7..2...8....3..16..7..6.2..5.1........6....91.....5.......4.8.6.2.8.7..3
  9.2    4    24
124235
1......6..6..43...9....1..7..8..4.25.1...76......5..........8.2..4.9..5.8..3..7..
  9.3    4    24
143634
..5.3.4.2....54.9.1.........19.4..8.4.3..72....85....99..6.8...........7.27..5...
  9.2    4    26
175675
..85.....4......2..3..6.9.8....82...2.....3...7.9.1....9.8..6....52...7....7....1
  9.2    5    23
188892
...5...476...4..3..9....1....2...5..7.......8...6...1.9..46......6..1..28..9.24..
  9.2    4    24
200753
....6.3..4..2...1..1.4..8...86..9.3.....7.....4...816.7......8..9....7...34...2..
  9.2    4    24
298676
2...5...9..6...5.......7.1.4.5.....6.8.3.6.4....8..7....1...43..2...89..9...1..5.
  9.2    5    25
346155
7...63....4....76...3........9...45.23.4...7.47..8...1..72......1..492......3.51.
  9.2    4    27
420249
...37.4.....4..5.2.5.....872..9......6..24...78........4...17.5.1.......5.....8.9
  9.2    3    24
425702
..8..2....278...3.9....1....6..........3..95..5..1.4.61...592.4....8...5...1..7..
  9.2    6    25
431065
.3.....7.9....16...67.....1..9.467....61.5.4....3...1......8.......1...4..495716.
  9.2    5    27
618406
15.8.....2...4.19......1..3.4....78...6..2.1......8..9....2.9..3....4.5...71....6
  9.3    5    25
631828
1.57....4..6..53.....8...5...7.5...94.......6.2.6.9.......7.....6...3.92.3....41.
  9.2    5    25
641898
..74.2...48.....9..1..5......6.19.34.4..........7..8.9...2.5...3.....1.6.94.3..7.
  9.2    3    26
717866
...51....5.4.......21....9..7..6.42...3..8....9.7....83....2.1.7..8....9.....7.5.
  9.2    5    24
739105
..7..6.8....2...46....3.2..41......27......5...2.19..4..16.5...38.12...95........
  9.2    4    26
792244
1.........43..5.....938...1...92.5..9..6.8.23......9....4..6.75...7.....6...1..9.
  9.2    4    25
810153
..64......7..5....9....6..8.....5.8..2.9..3....83..7..5.4.37.1.........7...12.56.
  9.2    2    25
845088
.8..1...6..3...5..2....7.....2.3..........8.947.8...2.6.....9.73..2.4.......5....
  9.2    5    22
942062
.4......96....98..7.......2..7.13....8.....2.1...5.9......2.5..5..4...78..1.7...4
  9.2    6    24
```

The new combined engines are much more efficient at finding shorter paths as evidenced by the nrczt+ which indicates the maximum length chain with ALSs needed to find a solution. I've finally designed a breadth-first search that is capable of performing the zt promotions and capable of extemely long chains (> 64 levels) in record times.

I've also adopted a new POV regarding nrc relationships. For example, an nrc point (one of the 729 possibles) that causes an ALS to convert to a locked set is now considered nrc linked to all of the N of 729 candidates that are potentially eliminated by the new LS. These nrc relationships are weak-links in this case, but they are interesting in that they are "remote" and not restricted to a set of candidates that can directly "see" each other. I build a 729x729 adjacency matrix that is composed of all the possible nrc relationships that I can assemble using standard nrc links as well as ALSs, AURs, GSLs, X-coloring conjugate chains. This seems to be providing a vastly more complex tree structure, but one with vastly more possible eliminations. The new BFS search engine is capable of traversing the adjacency matrix in an extremely efficient manner and it is generating some really interesting chains and solutions. I'm in the process of adding code to generate *.sud files so that I can analyze the chains using Allan Barker's Xsudo program.

Cheers,
Paul

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Posted: Sun Jun 28, 2009 4:56 am Post subject: quote edit

Hi **Paul**,
Nice to hear of you.

> **PIsaacson wrote:**
> I am in the process of a complete re-write of my nrczt+groups engine and testing the effectiveness of the new theory

Great!

> **PIsaacson wrote:**
> The new combined engines are much more efficient at finding shorter paths as evidenced by the nrczt+ which indicates the maximum length chain with ALSs needed to find a solution.

Such comparisons of length will be meaningful only if ALSs inserts are given their proper length, consistent with their quasi-nrczt nature. As I suggested in a previous post, a Pair should add 2 to length, a Triplet 3 and a Quad 4. Otherwise, a chain could have different lengths whether you consider it as pure nrczt or nrczt with LS inserts.
Basically, your definition of length amounts to giving the same complexity to a Single and to a Quad.

> **PIsaacson wrote:**
> I've finally designed a breadth-first search that is capable of performing the zt promotions and capable of extemely long chains (> 64 levels) in record times.

Do you get whips or braids?
If you get braids, could you look at my general T&E procedure (T&E(FP), in the "lovely braids" thread) and explain what's basically different in your program?

> **PIsaacson wrote:**
> I've also adopted a new POV regarding nrc relationships. For example, an nrc point (one of the 729 possibles) that causes an ALS to convert to a locked set is now considered nrc linked to all of the N of 729 candidates that are potentially eliminated by the new LS.

Please be careful with the vocabulary and don't introduce confusion by modifying the most basic definitions: nrc-linked just means nrc-linked. If you define indirect links, you should give them another name, e.g. "indirect links" or "remote links", as you do later in this post.
nrc-links don't depend on the knowledge state, your remote links do. And they are obviously more complex than nrc-links.
It's not only a matter of definitions but also of consistency. I don't know exactly how you count the length of a chain. But if you count it by giving all such indirect links the same length 1 as direct links, as suggested by

your next paragraph, it clearly introduces a strong bias towards chains that seem shorter according to your modified definition but that are not really shorter if you adopt a consistent definition of chain length.

> **PIsaacson wrote:**
>
> These nrc relationships are weak-links in this case, but they are interesting in that they are "remote" and not restricted to a set of candidates that can directly "see" each other. I build a 729x729 adjacency matrix that is composed of all the possible nrc relationships that I can assemble using standard nrc links as well as ALSs, AURs, GSLs, X-coloring conjugate chains. This seems to be providing a vastly more complex tree structure, but one with vastly more possible eliminations. The new BFS search engine is capable of traversing the adjacency matrix in an extremely efficient manner and it is generating some really interesting chains and solutions.

As I've shown in the "lovely braids" thread, all the known puzzles can be solved by T&E(FP), where FP is a family of basic patterns, e.g. FP = LS + ...
Any program based on DFS or BFS is extremely fast (mine takes only a few milli-seconds in most cases, a few hundred milli-seconds in the worst cases). But there are two very different problems:
- finding a solution,
- finding a solution that satisfies some additional criteria, e.g. one with minimal length chains.
Problems of the second type are generally exponentially harder than problems of the first.
With your definition of length, your new algorithm solves the first problem, not the second.

Let me suggest an interesting exercise in complexity: try modifying your algorithm in order to introduce a consistent definition of length.

Finally, it is unclear if you can deal with A*LS that need several rlc's to be "zt-promoted" to LS, thus obtaining the full power of zt-whips(FP) instead of only whips with FP inserts. Your description suggets a negative answer. Do you have any example that can't be solved by standard nrczt-whips (or braids)?

This may seem a little too critical, but be sure I do appreciate your work.

Cheers.

Back to top profile pm www

denis_berthier

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Posted: Mon Jun 29, 2009 8:21 am Post subject: quote edit

> **denis_berthier wrote:**
>
> Thinking again of all this: 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.

I've just posted an answer to this question in the "unbiased grid generation" thread (http://www.sudoku.com/boards/viewtopic.php?p=78520#78520):
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. Red Ed reaches a similar conclusion.

For the purposes of this thread, this means that we have no reason to suspect any bias in sudogen0_1M, even if there was a slight one in the complete grid generation part of suexg.

Last edited by denis_berthier on Fri Jul 03, 2009 4:44 pm; edited 1 time in total

Back to top profile pm www

denis_berthier

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Posted: Mon Jun 29, 2009 10:11 am Post subject: quote edit

> **m_b_metcalf wrote:**
>
> For what it's worth, I generated an hour's worth of puzzles all from the same grid. The summary is:
>
> **Code:**

```
Number:       54074
Average:   24.25319

     20       4
     21     194
     22    2443
     23   10709
     24   19050
     25   14842
     26    5588
     27    1117
     28     120
     29       7
```

I noted a discrepancy betwen the value you give for the average number of clues and the one I get, 25, which is much above the sudogen0_1M mean: 24.38.
Then I noticed that there are 53005 (not 54074) puzzles in the file you sent me. Can you check your figures?
There's also one puzzle with 30 clues.

I'm running the computations with the first 50,000 in order to see if generating from a single complete grid makes a big difference wrt the complexities of the puzzles (but of course, it may depend on how this unique complete grid is chosen).

Regards

Back to top profile pm www

m_b_metcalf

Joined: 15 May 2006
Posts: 2143
Location: Berlin

Posted: Mon Jun 29, 2009 10:50 am Post subject: quote

denis_berthier wrote:

> Then I noticed that there are 53005 (not 54074) puzzles in the file you sent me. Can you check your figures?
> There's also one puzzle with 30 clues.

Denis,
The third file I sent you was the biased sample with 53005 puzzles (3 with 30 clues) and an average of 25. (The file was named puzzles_biased.) I will send you the same-grid file now.

Regards,

Mike Metcalf

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Posted: Mon Jun 29, 2009 10:59 am Post subject: quote edit

I've run the computations on the first 50,000 in this biased file.

1) Clues
mean number of clues = 25.00
standard deviation = 1.21

Code:

```
#Clues #Puzzles  E(SER) s(SER)
20       0        0.0    0.0 (E not meaningful)
21       35       2.41   1.36 (E not meaningful)
22       658      2.52   1.47
23       4198     2.56   1.61
24       12095    2.64   1.69
25       16544    2.70   1.77
26       11304    2.79   1.83
27       4179     2.82   1.84
```

```
28      882      2.85   1.85
29      102      2.70   1.78
30      3        2.0   0.0 (E not meaningful)
all     50k      2.71   1.13
```

2) SER
mean SER = 2.71 (standard deviation = 1.13)
Much lower than Sudogen0 : 3.77 (standard deviation = 2.42)
No puzzle above 9.0.

The puzzles in this third series are globally much easier than those in sudogen0.

3) Correlation coefficient #clues vs SER = 0.04 : uncorrelated. This seems to be a very stable result, true through all the sets of puzzles analysed here.

Regards

Back to top profile pm www

Red Ed

Joined: 06 Jun 2005
Posts: 513

Posted: Mon Jun 29, 2009 11:11 am Post subject: quote

> **denis_berthier wrote:**
>
> Conclusion: tests for unavoidable sets in complete grids are not relevant to the complexity of puzzles. This is understandable as unavoidable sets in complete grids are washed out by the elimination phase of the puzzle generators. Red Ed reaches a similar conclusion.

Yes (for certain types of bias), but ...

> **Quote:**
>
> For the purposes of this thread, this means that we have no reason to suspect any bias in sudogen0_1M, even if there was a slight one in the complete grid generation part of suexg.

... that's a completely different conclusion! You showed that certain types of bias in the solution grid generator do not affect puzzle complexity. Where have you shown that bias in the elimination phase of the puzzle generator is unimportant?

Back to top profile pm

Red Ed

Joined: 06 Jun 2005
Posts: 513

Posted: Mon Jun 29, 2009 11:19 am Post subject: quote

> **denis_berthier wrote:**
>
> **Code:**
>
> ```
> #Clues #Puzzles  E(SER) s(SER)
> 20       0         0.0    0.0 (E not meaningful)
> 21       35        2.41   1.36 (E not meaningful)
> 22       658       2.52   1.47
> 23       4198      2.56   1.61
> 24       12095     2.64   1.69
> 25       16544     2.70   1.77
> 26       11304     2.79   1.83
> 27       4179      2.82   1.84
> 28       882       2.85   1.85
> 29       102       2.70   1.78
> 30       3         2.0   0.0 (E not meaningful)
> all      50k       2.71   1.13
> ```
>
> ...
>
> Correlation coefficient #clues vs SER = 0.04 : uncorrelated. This seems to be a very stable result, true through all the sets of puzzles analysed here.

Well hang on, that's not the full story is it? There's quite clearly a trend in the table above for puzzles with more clues to have higher SER. It would be interesting to know why that is.

Back to top profile pm

denis_berthier

Posted: Mon Jun 29, 2009 11:32 am Post subject: quote edit

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Red Ed wrote:

> **denis_berthier wrote:**
>
> > Conclusion: tests for unavoidable sets in complete grids are not relevant to the complexity of puzzles. This is understandable as unavoidable sets in complete grids are washed out by the elimination phase of the puzzle generators. Red Ed reaches a similar conclusion.
>
> Yes (for certain types of bias), but ...

For the types of bias I stated. Or do you now mean that some tests with some of your patterns could be relevant?

Red Ed wrote:

> **Quote:**
>
> > For the purposes of this thread, this means that we have no reason to suspect any bias in sudogen0_1M, even if there was a slight one in the complete grid generation part of suexg.
>
> ... that's a completely different conclusion! You showed that certain types of bias in the solution grid generator do not affect puzzle complexity. Where have you shown that bias in the elimination phase of the puzzle generator is unimportant?

You're right to give such precisions.
But I didn't say that "bias in the elimination phase of the puzzle generator is unimportant". Of course, it is important, as shown by the first 2 lists of puzzles provided by m_b_metcalf, the first of which had an obvious bias. But I can't see any way a bias could be introduced by suexg elimination phase.

Last edited by denis_berthier on Fri Jul 03, 2009 4:45 pm; edited 2 times in total

Back to top profile pm www

denis_berthier

Posted: Mon Jun 29, 2009 11:34 am Post subject: quote edit

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Red Ed wrote:

> **denis_berthier wrote:**
>
> **Code:**
>
> ```
> #Clues #Puzzles  E(SER) s(SER)
> 20       0        0.0    0.0 (E not meaningful)
> 21       35       2.41   1.36 (E not meaningful)
> 22       658      2.52   1.47
> 23       4198     2.56   1.61
> 24       12095    2.64   1.69
> 25       16544    2.70   1.77
> 26       11304    2.79   1.83
> 27       4179     2.82   1.84
> 28       882      2.85   1.85
> 29       102      2.70   1.78
> 30       3        2.0   0.0 (E not meaningful)
> all      50k      2.71   1.13
> ```
>
> ...
>
> Correlation coefficient #clues vs SER = 0.04 : uncorrelated. This seems to be a very stable result, true through all the sets of puzzles analysed here.

Well hang on, that's not the full story is it? There's quite clearly a trend in the table above for puzzles with more clues to have higher SER. It would be interesting to know why that is.

Yes, but we have also seen cases in which the trend was different. And it is smaller than the standard deviations, therefore not really meaningful.
Anyway, this is not a good example because we know it is biased.

Back to top profile pm www

Red Ed

Posted: Mon Jun 29, 2009 11:48 am Post subject: quote

Joined: 06 Jun 2005
Posts: 513

Your comment about the "trend" being smaller than the standard deviations is misleading if you're referring to the upward trend in E(SER) compared to the std devs s(SER). For example, *if* the 25-clue puzzles from Mike's generation process have s(SER)=1.77 (as per the sample) *then* the mean, E(SER), of a random sample of 16544 25-clue puzzles will have s(E(SER))=1.77/sqrt(16544) ~= 0.014. So, relative to s(E(SER)), the trend in E(SER) *does* appear to be significant.

If you need further evidence, do a ranks test on E(SER), e.g. Mann-Whitney.

Back to top profile pm

denis_berthier

Posted: Mon Jun 29, 2009 12:23 pm Post subject: quote edit

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Red Ed wrote:

> Your comment about the "trend" being smaller than the standard deviations is misleading if you're referring to the upward trend in E(SER) compared to the std devs s(SER). For example, *if* the 25-clue puzzles from Mike's generation process have s(SER)=1.77 (as per the sample) *then* the mean, E(SER), of a random sample of 16544 25-clue puzzles will have s(E(SER))=1.77/sqrt(16544) ~= 0.014. So, relative to s(E(SER)), the trend in E(SER) *does* appear to be significant.

It'd be too easy to prove anything if one could do statistical computations the way you're doing them here: take still larger samples of 25-clue puzzles and their s(E(SER)) will be ~0. The error is that we are not dealing with E(E(SER)) as your computations implicitly assume but with E(SER).

I'm not denying the small trend of SER (or NRCZT) wrt #clues, even in sudogen0_1M, which is certainly a more interesting reference than the present biased example for such discussions.
I just want to relativise its importance, given the large standard deviations.

Now, the number of clues itself is only marginally relevant to this thread. It appeared here because coloin noticed such a small trend. The (possible) trend is not linear. The two important points relative to such a possible trend are:
- is it real? i.e. does it exist in the set of all the minimal puzzles? Or is it an artifact of the generator? Answer: unknown.
- if it was real and a generator of puzzles had some small bias wrt the number of clues, would it have any significant impact on the classification results? Answer: no. Because in all the cases we have examined here, the correlation between #clues and complexity (SER or NRCZT) is very small.

Back to top profile pm www

Red Ed

Posted: Mon Jun 29, 2009 12:40 pm Post subject: quote

Joined: 06 Jun 2005
Posts: 513

denis_berthier wrote:

> It'd be too easy to prove anything if one could do statistical computations the way you're doing them here: take still larger samples of 25-clue puzzles and their s(E(SER)) will be ~0. The error is that we are not dealing with E(E(SER)) as your computations implicitly assume but with E(SER).

No. We're talking about a trend in E(SER). To analyse that, you need to treat E(SER) as a random variable in its own right. If you prefer not to think about statistics of E(SER), then do a distribution-neutral test like Wilcoxon Rank Sum. You'll get a p-value that says the trend is non-random, like it or not.

However, I agree with the conclusions you reach in your "two important points relative to such a possible trend ..." paragraph. It would be interesting (to me) to know why the trend exists, though probably irrelevant to the main topic of measuring puzzle complexity.

Back to top profile pm

denis_berthier

Joined: 19 Jun 2007
Posts: 625
Location: Paris, France

Posted: Mon Jun 29, 2009 12:44 pm Post subject: quote edit

Here are the results for the last file propsoed by Mike: puzzles generated from a single grid (http://www.sudoku.com/boards/viewtopic.php?t=14592)

1) Clues
mean number of clues = 24.25
standard deviation = 1.12
This is a little less than sudogen0 (24.38) or the "unbiased" second list (24.40)

Code:

```
#Clues #Puzzles  E(SER) s(SER)
20       3        1.87   0.52 (E not meaningful)
21       185      2.46   1.46 (E not meaningful)
22       2245     2.54   1.57
23       9893     2.44   1.48
24       17599    2.34   1.38
25       13754    2.26   1.28
26       5172     2.19   1.15
27       1032     2.21   1.18
28       112      2.30   1.26
29       5        1.70   0.24 (E not meaningful)
30       0        0   0.0 (E not meaningful)
all      50k      2.33   1.36
```

2) SER
mean SER = 2.33 (standard deviation = 1.36)
Much lower than Sudogen0 : 3.77 (standard deviation = 2.42)
No puzzle above 9.0.

The puzzles in this third series are globally much easier than those in sudogen0.

3) Correlation coefficient #clues vs SER = - 0.04 : uncorrelated. This seems to be a very stable result, true through all the sets of puzzles analysed here.

Conclusion: I wouldn't recommend trying to generate random puzzles from a single complete grid.

Back to top profile pm www

Display posts from previous: All Posts Oldest First Go

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

All times are GMT
Goto page Previous 1, 2, 3 ... 11, 12, 13, 14 Next

Page 12 of 14

Stop watching this topic

Jump to: Advanced solving techniques Go